REVITALIZING A NATION:

COMPETITION AND INNOVATION IN THE US TRANSPORTATON SYSTEM

Clifford Winston   Jia Yan and   Associates

September 2022

Table of Contents

Chapter 1. Introduction

Clifford Winston and Jia Yan

An efficient transportation system enables society to reap the benefits of a modern economy in several ways. First, firms can sell their products to consumers in many markets and produce at greater scale, which reduces their costs. Second, multiple firms can compete in a market, which prevents monopoly pricing and poor service. Third, consumers can choose among a variety of products produced locally, in other cities, and abroad. Fourth, employers and workers are better able to match their job requirements and skills. Fifth, firms and people are better able to locate near one another in cities and industrial clusters and become more productive through agglomeration economies. Sixth, households' leisure time is enhanced by affordable and extensive domestic and foreign travel opportunities. Finally, the expanded activities of individuals and firms increase the likelihood of innovations and technological advance that widely benefit the public and spur macroeconomic growth.

Because transportation provides significant and extensive benefits to a nation and because a transportation system is constantly evolving, policymakers can substantially increase a nation's welfare by implementing policies that improve transportation system performance. In this book, we argue that greater *competition* and *innovation* are the primary sources of transportation improvements and that policymakers can encourage greater competition and innovation more effectively by reforming many of their transportation policies.

Postwar Evolution of the US Transportation System

The evolution of the transportation system as the US emerged from World War II is a useful starting point for understanding how greater competition and innovation improved the system's performance. Government funded investments in transportation infrastructure that enabled the

major modes to operate more efficiently. Construction of the Interstate Highway System, which began in 1956, and the construction of new airports, which began in 1962 with the opening of Washington, DC Dulles Airport to accommodate the longer takeoff distances of jets, significantly contributed to the nation's welfare (Allen and Arkolakis (2014), Morrison and Winston (1985)). Investments in major metropolitan areas to build new suburban rail transit systems, funded by all levels of government, followed the huge federal investments in highway and airport infrastructure. The introduction of commercial jet service in 1958 was the major transportation innovation that occurred during the postwar period.

At first blush, it appeared that the US was on its way to building the best transportation system money could buy. However, inefficiencies created by pervasive public sector involvement in transportation, which Adam Smith (1776) had remarkably predicted long ago, had steadily worked their way into the system, including but not limited to:

— *Excess capacity* as the road and rail network were not designed in accordance with what commerce could pay but were overbuilt to serve low-density areas of the country with little freight and passenger traffic.

— *Growing deficits* as public infrastructure and services were not developed to be financed solely by users but also were funded out of general tax revenues.

— *Misallocation of transportation revenues* as repairs were often executed in a poor-manner and sometimes not executed at all leading to premature deterioration that was not easy to remedy because of the difficulty of raising new revenues, and revenues that were raised were not always used for their intended purpose but used to subsidize less efficient transportation facilities and services.

The classic books by Meyer, Peck, Stenason, and Zwick (1959) and Meyer, Kain, and Wohl (1965) identified in detail and quantified to some extent the costs of the most inefficient parts of the 1950s and 1960s US transportation system. Generally, the authors took issue with what microeconomists call the engineering approach to improving a transportation system, which focuses on increasing spending and investment, not on efficient policy to optimize the system's performance.[1]

<u>Improvements in the Transportation System</u>

Poor policy was the primary source of the system's inefficiencies. For example, the large investments in the highway system attracted more traffic and led to congestion, increased accidents and fatalities, and premature deterioration of the capital stock. The significant social costs of highway travel could not be reduced very much by increasing government spending on transportation.

At the same time, increasing government spending could not improve the efficiency of regulated intercity modes that were characterized by inflated costs, due in part to regulatory-induced excess capacity, and by inflated prices for travelers and shippers. Finally, subsidies for public urban transportation systems had to sharply increase to cover growing transit deficits while doing little to reverse transit's loss of market share.

Adam Smith is recognized for indicating the importance of competition to improve an economy, but he did not envision that increasing competition would be necessary to improve a transportation system that had been severely compromised by poor government policy. Yet, beginning in the mid-1970s, deregulation led to significant efficiency improvements in the intercity transportation system by increasing competition between and within the surface freight

---

[1] Winston (2021a) notes that beginning in the late 1980s, macroeconomists focused on increasing spending on transportation infrastructure to spur productivity growth but not on efficient infrastructure pricing and investment.

modes and between and within airlines and other passenger modes. Deregulation also spurred technological and operating innovations that reduced the modes' costs and prices and improved service quality (Morrison and Winston (1999)).[2]

Greater competition and innovation fostered by deregulation emerged clearly as the sources of significant improvements in intercity transportation. However, those sources of improvement were prevented from affecting the efficiency of the international and urban transportation systems because those systems were still highly regulated.

In the past few decades, new sources of competition and innovation have begun to improve the international and urban transportation systems to some extent. For example, beginning in 1992 with the Netherlands, the US has negotiated open skies agreements with some countries, which has stimulated competition by deregulating airline prices and entry on US international routes, resulting in lower fares and greater service frequency (Winston and Yan (2015)). Beginning in 2010 when Uber launched its first on-demand car service in San Francisco, rideshare companies have used advances in information technology to provide a new form of transportation service in metropolitan areas that has become a popular alternative to taxis and public transit.

<div align="center">Remaining Transportation Problems</div>

Unfortunately, the improvements in the US transportation system are often overshadowed by the system's formidable problems. The most important ones are that the system:

—Does a poor job of enhancing mobility for less affluent, disabled, and elderly people.

---

[2] Administrative transportation deregulation began with the 1976 Railroad Revitalization and Regulatory Reform Act, which established the basic outlines of regulatory reform in the railroad industry, and with the Civil Aeronautics Board's liberalization of fares by allowing, for example, "Supersaver" fares on some routes before fares were fully deregulated by the 1978 Airline Deregulation Act

—Continues to be plagued by externalities, including the nation's leading share of greenhouse gas emissions, highway, airport, and port congestion, and some 40,000 annual fatalities and millions of annual injuries from highway accidents.

—Does not maintain and expand the infrastructure capital stock efficiently in response to growing use of the transportation system by travelers and shippers.

—Attracts only a small share of travelers on public transit and Amtrak intercity passenger rail service in most markets, which run increasingly large deficits that are financed by taxpayers.

—Has failed to integrate efficiently its domestic and international parts, which increases prices and travel times and reduces reliability.

—Experiences supply chain disruptions on domestic and international routes, which cause shortages of products and raise consumer prices.

—Has become a significant potential source of infection from COVID-19.

New sources of competition and innovation can significantly help to address those transportation problems and others that may arise in the future. However, government policymakers must unleash those sources in two ways. First, it must enact constructive policy reforms that give private transportation entities greater freedom to compete and innovate and that facilitate public adoption of autonomous transportation modes. Second, it must increase the benefits that travelers and shippers obtain from using the transportation system by reforming current infrastructure policies to improve the system's physical condition, optimize traffic flows, and enhance safety.

Unfortunately, as noted, policymakers have not prioritized those reforms because they persistently rely on an engineering approach that focuses on spending large sums of taxpayers'

money to solve transportation problems and consistently falls short of producing significant improvements in the system. Congressional passage of the 2021 $1.2 trillion Infrastructure Investment and Jobs Act takes the same approach and is likely to produce the same disappointing outcome.

Transportation inefficiencies persist because policymakers, stakeholders, and the public often are fearful of the effects of competition and new technologies and generally prefer that government owns and operates transportation services and facilities, or regulates their behavior, or even prohibits their operation. For example, although automated elevators were introduced in the 1850s as a new transportation technology, they had a human operator for another century until passenger operated emergency equipment, such as a stop button, alert bell, and a telephone, helped to convince the public that they could use the technology safely (Lokshin and Newsom (2020)). As another example, gasoline-powered vehicles became a successful alternative to the horse only after steam-powered vehicles were effectively legislated out of business.

Government may have plausible reasons for owning and operating transportation facilities and for regulating private firms' prices and operations. However, government policy is static and very slow to react to the dynamic changes in competition and transportation technology that motivate reform. Indeed, the government should have implemented intercity transportation deregulation, which significantly benefited the public, decades earlier.

It also should be much further along with exploring how efficient infrastructure pricing and investment work together to enable transportation users to make the best use of the nation's limited infrastructure capacity and to keep the infrastructure in good repair. Unfortunately, the various government authorities that own much of the US transportation infrastructure generally have paid little attention to the inefficiencies of their pricing and investment policies. It may turn out that

new autonomous modes provide the incentive for government to reform its infrastructure policies because the failure to do so will compromise the performance of those new modes. Privatizing more of the US transportation infrastructure may be necessary to realize the substantial benefits from the new modes because private owners are more likely than the government to implement more efficient pricing and investment policies.

Plan of the Book

We organize the book around four parts. The first three identify the major current economic and social problems and challenges associated with the urban, intercity passenger, and freight transportation systems, respectively. We analyze various ways that new sources of competition and innovation could positively affect the modes that comprise those systems. We then synthesize our findings and conclusions in the final part of the book and explain how public policy can enable competition and innovation to realize their full potential to enhance today's as well as the future US transportation system.

Some chapters contain new empirical work that provides evidence on the actual and potential welfare improvements from greater competition and new innovations in each part of the US transportation system. Other chapters identify and discuss future sources of competition and innovation and their potential social benefits.

We argue that US policymakers should take several measures to generate more competition in the transportation system, which include:

—Negotiating open skies agreements with all countries.

—Allowing foreign airlines cabotage rights to spur global deregulation that would facilitate seamless global operations.

—Privatizing airports and air traffic control so both airports and airlines can operate more effectively and compete more intensely.

—Privatizing ports so they can operate more efficiently and compete more intensely.

—Conducting highway privatization experiments to explore the potential benefits and feasibility of private highways in the US.

—Withdrawing any obstacles that prevent ridesharing companies from competing in urban areas.

—Fully deregulating freight railroads and ocean shipping, including repeal of the Jones Act and eliminating rate conferences.

We also argue that government should facilitate and expedite the future adoption of transportation innovations that will use private and public infrastructure, including autonomous electric cars, trucks, railroads, ships, as well as air taxis and drones. Those innovations will further stimulate intra and intermodal competition and will provide enormous benefits to travelers, shippers, and the overall economy.

In the final analysis, because transportation plays such a vital role in the development, application, and success of any innovation, policymakers can promote a stronger culture of competition and innovation throughout the US economy by creating a stronger culture of competition and innovation in the US transportation system. And by overcoming the system's current weaknesses, policymakers can enable future transportation systems to be efficient, equitable, and innovative and to revitalize the nation.

PART I

URBAN TRANSPORTATION

The US urban transportation system is plagued by externalities caused by automobile travel, including congestion, fatal and serious accidents, and vehicle emissions. Policymakers' efforts to simultaneously tax and subsidize automobile travel and to heavily subsidize public transit to reduce automobile travel have not ameliorated those externalities. Efficient externality taxes on vehicle miles traveled (VMT) for automobiles are being considered in some parts of the country (Langer, Maheshri, and Winston (2017)), but VMT taxes have yet to be implemented anywhere in the United States.

In part I of the book, we provide new evidence on the difficulty facing policymakers in attempting to simultaneously reduce highway congestion and fatalities, and on additional inefficiencies of bus and rail transit operations that policymakers have created by not encouraging bus and rail to be effective competitors. We then examine how new sources of competition and innovation could reduce the cost of urban travel, improve service, and reduce automobile externalities by estimating the benefits of ridesharing and by providing an overview of the potential benefits of electric and autonomous vehicles.

Chapter 2. The Effect of Reductions in Vehicle Miles Traveled on

Highway Fatalities and Congestion with Heterogeneous Motorists

Vikram Maheshri and Clifford Winston[3]

1. Introduction

The interstate highway system is the lifeblood of urban and suburban travel. At the same time, the annual costs of highway congestion, injuries and fatalities to people involved in highway accidents, and vehicle damage from those accidents run in the trillions of dollars (Winston and Karpilow (2020)). In chapter 5, we discuss pollution from vehicle emissions, another important social cost of highway travel, in the context of electric vehicles.

Because congestion and accidents appear to be positively related to vehicle miles traveled (VMT), policies that decrease VMT could simultaneously improve both.[4] Such policies include congestion pricing that induces modal and intertemporal shifts by motorists and significant investments that build new public transit systems or improve existing systems. However, policies that have the potential to improve both safety and the speed of travel by reducing VMT could be undermined if they lead to adverse changes in driver behavior that offset those potential improvements.

We are not aware of previous research that provides empirical evidence of a specific policy in the United States that has simultaneously improved both safety and congestion.[5] But the Great

---

[3] We are grateful to Maryam Khan for research assistance.

[4] Of course, congestion and accidents interact with each other because accidents create delay and congestion, while congestion reduces motorists' speeds, which reduces the frequency and severity of accidents. Our primary interest in this chapter is to identify policies that by reducing VMT can reduce congestion and accidents, even accounting for their interaction.

[5] Tang and Van Ommeren (2021) find that the London Congestion Charge Zone reduced congestion and accidents. However, they find that it led to an increase in serious injuries and fatalities. In March 2020, Transport for London introduced a 20 miles per hour speed limit on all its roads within the central London congestion charging zone as part of its commitment to eliminate death and serious injury on London's roads. In the United States, the J-Tech Lane

Recession did have that effect. In Maheshri and Winston (2016), we found that the significant decline in VMT during the 2007-2009 Great Recession and its aftermath caused both a reduction in congestion and a decline in motorists' fatalities per VMT. We argued that those twin declines could be explained by the change in drivers' behavior as reflected in the mix of drivers who remained on the road, with a greater share of travel conducted by plausibly safer drivers (and hence a smaller share by plausibly riskier drivers).

The COVID-19 global pandemic provides a new opportunity to explore the effects of a significant decline in VMT on vehicle congestion and safety because it led to a sharp increase in the number of people who worked from home to reduce the spread of COVID-19, rising from 6% of workers in 2019 to a peak of 35% of workers in May 2020, according to the American Community Survey. Congestion also decreased because the initial decline in peak-period travel was not offset by latent travel demand; that is, travelers did not take advantage of reduced congestion during peak periods by driving more, as the pandemic chilled all forms of travel-related activity, including recreation.

Like the findings during the Great Recession, we would expect that the decline in VMT would reduce fatalities because motorists were less exposed to other motorists, thereby reducing the opportunities for multi-vehicle accidents to occur. However, the effect on safety of the concomitant reduction in congestion also depends on how drivers react to the change in highway travel conditions. For example, by enabling motorists to drive faster, the reduction in congestion could increase the likelihood and severity of accidents, other things equal.

Because the decline in driving during the Great Recession was disproportionately shouldered by riskier drivers, the reduction in congestion did not offset the benefits of less

---

Blade is marketed as reducing congestion and improving safety by safely removing debris from travel lanes while enabling workers to safely operate from the cab of the service vehicle. See https://jtechusa.com/laneblade

exposure to other motorists by allowing motorists to take greater risks while driving, such as drinking, using drugs, or speeding. But descriptive evidence shown in Figure 1 suggests that in contrast to their behavior during the Great Recession, when the fatality rate declined considerably, the remaining motorists on the road may have engaged in risky driving behavior during the pandemic, because the fatality rate began to spike in 2020 while the pandemic was underway. The problem has persisted given the first quarter of 2022 had the highest fatalities on US roads in more than two decades.

In this chapter, we report empirical evidence that suggests the decline in VMT that has occurred during the pandemic did not simultaneously improve congestion and automobile safety because it led to a change in the mix of drivers that appears to include a greater share of risky drivers who, among other things, appear to drive at excessive speed. Although total driving did not change considerably in the aggregate population of drivers, the proliferation of telecommuting and remote schooling shifted highway travel away from conventional peak periods. The intertemporal shift facilitated sharp reductions in congestion, but it also increased the risk to motorists' safety because some drove at higher speeds that reduced their reaction time to avoid an accident and increased the bodily impact of a collision.[6]

We explain why congestion pricing and investments in public transit could reduce VMT, but also could lead to more drivers engaging in risky behavior on certain roads at certain times of day. We conclude by pointing out that autonomous vehicles circumvent the travel speed–safety

---

[6] Gao and Levinson (2021) find that morning traffic flows in Minneapolis-St. Paul developed a pronounced double peak after COVID arrived, which they attribute to a change in the composition of the commuting workforce, with larger reductions in sectors that tend to commute during the middle of the peak. Workers in construction, natural resources, and mining tend to commute early, and cannot switch to telecommuting. Many office workers can telecommute, and they may be safer drivers than workers in construction, natural resources, and mining, who are predominantly male and possibly drive pickup trucks to their workplaces in the dark.

tradeoff, making their widespread adoption a particularly attractive way to simultaneously reduce congestion and automobile accidents.

2. Empirical Approach

Our goal is to estimate the effect of congestion on automobile fatalities. Although there is considerable research on the determinants of automobile fatalities in the economics and transportation engineering literature[7], estimating the effect of congestion on automobile fatalities is challenging because many of the features of the driving environment that affect congestion, including the geography and topography of roads, weather, local economic conditions, and socioeconomic characteristics of drivers, also are potential determinants of auto fatalities. We disentangle the confounding influences on congestion and fatalities by exploiting the unexpected shock to driving that accompanied the COVID-19 pandemic in the United States.

Congestion is a non-linear phenomenon that increases sharply for a given road capacity after vehicle traffic reaches a certain threshold.[8] Below this threshold, marginal changes in traffic contribute little or nothing to congestion. However, after the threshold is met, small increases (or reductions) in highway traffic may result in dramatic changes in congestion and travel delays. We leverage this feature of congestion in our analysis.

As driving plummeted throughout the nation during the early months of the pandemic, highways that previously were highly congested in cities, such as Chicago, New York, and Boston, became uncongested overnight. At the same time, highways in other cities that previously were not congested also experienced significant declines in traffic, but they did not become dramatically less congested. We can therefore use a difference-in-differences approach to compare the effect

---

[7] See, for example, Small and Verhoef (2007) and Retallack and Ostendorf (2019).
[8] See, for example, Lindsey (2012) for a discussion of the technology of highway congestion.

of pandemic-related driving shutdowns on auto fatalities in cities with highways that previously were highly congested with that effect in other cities that previously were not congested to estimate the overall effect of congestion on auto fatalities.

To conduct our analysis, we collect data from four distinct sources. First, we obtain county-level measures of monthly VMT from StreetLight Data, a private company that combines cell phone data, telematic data from vehicles, and actual readings from roadway sensors to estimate VMT at a spatially and temporally fine level. Second, we obtain baseline measures of auto fatalities from the Fatality Accident Reporting System (FARS) database maintained by the National Highway Traffic Safety Administration (NHTSA). Because those data were available only through the end of 2019, before the pandemic took hold in the United States, we contacted each of the 50 most-heavily traveled counties (as measured by 2019 VMT) to obtain monthly measures of auto fatalities from January 2020 through March 2021. Third, we obtain baseline measures of congestion at the Metropolitan Statistical Area (MSA) level from INRIX, a private company that specializes in estimating congestion in cities around the world. Finally, we obtain monthly data from reports in the *New York Times* on the number of cases and deaths due to COVID-19 in the 50 counties in our sample.

We use the monthly data for the 50 most travelled counties in 2019, as measured by VMT, from January 2020 to March 2021 to obtain 750 observations. We also include another 600 observations from monthly data during 2019. Thus, we have a period before the pandemic when there was not any notable change in VMT other than seasonality, a period during the pandemic when VMT declined substantially, and a period toward the end of the sample when VMT had returned to (or even exceeded) its pre-pandemic level in most counties. Finally, we exclude the 50

observations from January 2019 because we specify month of year fixed effects in the model, so January 2019 becomes the reference group.

We use our final sample of 1300 observations to estimate the effect of congestion on automobile fatality risk as specified in the following two-way fixed effects equation:

$$\log F_{it} = \beta_1 \times \log VMT_{it} + \beta_2 \times POST_t \times CONGEST_i + \beta_3 \times POST_t \times \log VMT_{it} + \lambda_i + \lambda_t$$
$$+ \ CONTROLS_{it} + \epsilon_{it},$$

where $\log F_{it}$ is the natural log of the number of auto fatalities reported in county i in month t, $\log VMT_{it}$ is the natural log of vehicle miles travelled in county i in month t, $POST_t$ is a post-COVID dummy variable that is equal to 1 if the month is after March 2020; and 0 otherwise[9], and $CONGEST_i$ is a dummy variable equal to 1 if county i experienced very high levels of congestion in 2019; and 0 otherwise.[10] The λ_i are county fixed effects, which, for example, capture characteristics of the road network, λ_t are month-of-year-fixed effects that account for time varying unobserved determinants of fatalities, including seasonality in driving and weather patterns, and ϵ_{it} is the error term. Finally, $CONTROLS_{it}$ is a vector of county and time varying controls that includes cumulative reported COVID-19 cases, and cumulative reported COVID-19 deaths.[11]

Following the literature on modeling fatality risk, we specify all continuous variables in logarithms. Thus, those factors enter multiplicatively for purposes of interpretation. For instance,

---

[9] Because the dramatic reductions in driving due to COVID-19 commenced in the middle of March 2020, March 11th can be considered as the unofficial start of the Covid-19 panic in the US given it coincided with the date that the National Basketball Association suspended its entire 2020 season and when Tom Hanks revealed his positive test for the coronavirus. We also used an alternative definition of $POST_t$ equal to 1 beginning in April 2020; 0 otherwise, and we also replicated our analysis omitting data from March 2020 entirely. Those changes did not have any effects on our findings.

[10] A city that is very highly congested is one of the four cities, Boston, Chicago, New York, and Philadelphia, where drivers lost, on average, over 120 hours due to traffic conditions in 2019. Measures of congestion come from INRIX. We note later that we use higher and lower levels of hours lost to congestion in 2019 for sensitivity purposes.

[11] As noted, data on COVID-19 caseloads and deaths were collected from the *New York Times*.

VMT enters multiplicatively when describing fatality risk, as in "fatalities per billion VMT;" hence, we can interpret the coefficient β_2 as the differential change in fatality risk in congested counties versus uncongested counties following the introduction of COVID-19 shutdowns.

Table 1 reports the means and standard deviations of the variables for the entire sample, divided into pre and post pandemic. Absolute monthly fatalities have increased, although total VMT has decreased 8.8%. The decrease includes a sizable reduction in peak-period VMT that resulted in zero annual hours lost from congestion after the pandemic began in March 2020 compared with 68 annual hours lost from congestion pre-pandemic. Finally, we report the monthly COVID-19 cases and deaths during the pandemic.

3. Empirical Findings

Following standard procedures for estimating models with dependent variables that are (small) non-negative integers, such as automobile fatalities, we estimate the parameters in equation (1) by maximum likelihood using a Poisson regression model.[12] By exponentiating the parameters in equation (1), we obtain Incidence Risk Ratios (IRRs), which we present in table 2 along with stars to provide easy interpretation of statistical significance. An IRR greater than 1 corresponds to a positive relationship to vehicle fatalities, and an IRR less than 1 corresponds to a negative relationship to vehicle fatalities. For instance, an IRR of 1.10 indicates that a unit increase in the variable corresponds to a 10% increase in fatalities, and an IRR of 0.9 indicates that a unit increase in the variable corresponds to a 10% decrease in fatalities.

---

[12] Although many empirical analyses in the transportation literature estimate fatality equations using a negative binomial model, this is inadvisable, because, as pointed out by Wooldridge (1999), the negative binomial estimator is a non-robust estimator of conditional mean parameters, and this weakness is exacerbated by the use of fixed effects. Negative binomial regressions may be appropriate if the objective is simply to maximize the fit of the model and there is over-dispersion in the dependent variable. However, a Poisson regression is preferable in our case because our dependent variable is under-dispersed, with a mean of 13.3 and standard deviation of 11.04. Roback and Legler (2021) provide a textbook discussion of Poisson regression models.

Because our measure of congestion is at the MSA level and some MSAs span multiple counties, we first construct the $CONGEST_i$ variable by considering all the counties that are included in an MSA and coding them as congested. For example, Kings County and Queens County are both considered congested because the New York MSA is congested. The IRRs capturing the different effects of congestion presented in table 2 are based on this approach. Because we may be overstating congestion in certain metropolitan areas, we performed a robustness check where we coded only the largest county in a congested MSA as congested; however, our findings were virtually unchanged.

In column (1), we estimate a baseline regression that does not include information on different congestion levels in the MSAs. We find, unsurprisingly, that VMT is a strong positive predictor of fatalities, and this association declined only slightly, 5%, post shutdown in March 2020. This finding indicates that driving did not get substantially safer on a per-mile basis after the shutdowns, which is suggestive of our conjecture that although congestion increases exposure to other vehicles in the traffic flow, as it did pre-shutdown, it can reduce fatalities by causing motorists to drive more slowly and cautiously to avoid a collision. By significantly reducing congestion, COVID-19-related shutdowns and reduced economic activity may have increased auto fatalities under certain driving conditions.

In column (2), we test more directly our conjecture by allowing the effects of the shutdown to vary by cities' congestion using a flexible specification of congestion: highly congested cities compared with other cities (the omitted category).[13] In column (3), we include the additional control variables noted previously, such as cumulative reported COVID-19 cases and cumulative reported COVID-19 deaths, which are related to the contemporaneous COVID-19 situation. In all

---

[13] We tried lower, 100 hours, and higher, 125 hours, alternative annual thresholds of our base case for highly congested cities of 120 hours lost due to traffic congestion. We obtained similar results to those presented here.

specifications, we find that more driving results in more fatalities, all else constant. Quantitatively, a 10% increase in VMT corresponds to a 4-5% increase in fatalities.[14]

Although the first-order effect of the shutdowns was to dramatically reduce VMT and fatalities, that does not imply that all cities experienced a decrease in fatalities. As shown in columns (2) and (3) by the estimated effect of 1.12 of shutdowns in highly congested cities, the decrease was 12% smaller in counties that were highly congested in 2019 relative to those that were not highly congested in 2019 and the estimate is statistically significant. That is, as highly congested counties became less congested, the effect of the shutdown on fatalities was smaller than in counties that previously were less congested. This finding is consistent with the non-linear nature of congestion and its effect on driving speed: moderate levels of congestion affect driving speed similarly, but beyond a certain threshold, the impact of congestion on driving speed is significantly greater. Accordingly, reductions in congestion would be expected to affect fatalities only when congestion was previously a binding constraint on drivers' ability to engage in risky behavior.[15]

4. Discussion

The different effects of congestion on fatalities for cities that experienced different levels of congestion before the pandemic began suggests that fatalities may have actually increased in

---

[14] Vickrey (1968) distinguishes between an accident externality elasticity that measures the effect of the quantity of driving on fatalities as opposed to the effect of the quality of driving. We estimate that the elasticity of fatalities with respect to VMT is inelastic, but this estimate is not a pure quantity elasticity because we are not holding the quality of driving constant.

[15] Our finding and explanation for it is consistent with Tang and Van Ommeren (2021), who find that the London congestion charge led to more accidents resulting in serious injuries and fatalities and argued that by removing traffic along busy roads, the congestion charge reduced bottlenecks and increased travelling speed, resulting in more serious injuries.

previously highly congested cities during the pandemic. We confirm this possibility with the following calculation and then estimate the cost of greater fatalities in those cities.

Based on data from StreetLight, VMT in the highly congested cities of Boston, Chicago, New York, and Philadelphia was observed to decrease 7.4%, on average, after the shutdown in March 2020. Thus, based on the estimated IRRs from our model in column (2), the decrease in congestion should have resulted in a 2.15% *reduction* in fatalities ((1.43-1.0) × 5%). The four highly congested MSAs in our sample averaged 8.22 fatalities per month in the pre-COVID-19 period and increased to an average of 9.89 fatalities per month in the post-COVID-19 period. However, based solely on the 7.4% decrease in VMT in those four MSAs, we would have expected fatalities to have fallen to 8.04 per month in the post-COVID-19 period.

The additional 1.85 fatalities per month that we predict in the four highly congested MSAs amounts to roughly 177 more fatalities over a 12-month period.[16] Assuming a value of a statistical life of $11.6 million, based on US Department of Transportation 2020 guidelines[17], the annual cost attributable to the decline in congestion in those four MSAs alone amounts to roughly $2.05 billion.

This estimate, however, is an extreme lower bound of the annual cost of the decline in congestion because it does not include the large cost of non-fatal but serious injuries and vehicle damage. In addition, our simple calculation does not account for the fact that MSAs other than those deemed highly congested may also have had increases in fatality risk due to relaxing of the congestion constraint on some specific roadways, which may pose greater risks than other

---

[16] In our sample, 8 of the counties lie in one of the four highly congested MSAs; thus, we obtain approximately 177 fatalities as the product of 1.85×12×8.

[17] https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis

roadways because of their design, condition, and lack of police traffic enforcement. A more disaggregated analysis would be necessary to explore that possibility.

Recall that in Maheshri and Winston (2016), we found that during the Great Recession, both fatalities per VMT and congestion declined. Why did this relationship change during the pandemic? A natural explanation is that although a change in the mix of safe and risky motorists led to improved safety during the Great Recession because relatively fewer risky drivers were on the road, the response to the pandemic appears to have led to relatively more risky drivers on the road.

An extreme example is that the highly illegal Cannonball Run, a cross-country race from the Red Ball Garage in Manhattan to the Portofino Inn in Redondo Beach, California, recorded several record-breaking times in 2020 as drivers took advantage of empty roads that had been created by the coronavirus lockdown. Tefft et al. (2022) obtained evidence based on a national survey where motorists self-reported risky driving behaviors that the relative increase in driving exposure among segments of the driving population with relatively riskier safety profiles contributed to the increase in the rate of traffic fatalities during the COVID-19 pandemic.

Of course, our analysis is at the local level. Given that we cannot identify risky motorists in metropolitan areas during the pandemic based on their specific individual characteristics, we consider whether alternative measures of risky driving behavior that contributed to fatal accidents before and during the pandemic have changed. Table 3 presents descriptive evidence that there was little change in alcohol consumption being a factor in fatal crashes for the selected counties where we could obtain recent data, Cook, IL, Dallas, TX, Harris, TX, and Los Angeles/Riverside, CA, and that drug use was even less of a factor. The factors that contribute to fatalities vary in magnitude across counties in ways that we cannot explain, such as, why drugs are much more

often a factor in fatalities in Cook County than in Harris County; however, our interest is in the difference in a factor's contribution to a fatal accident for a given county before and during the pandemic.[18]

In contrast, table 3 shows that speeding has become a more prevalent factor in fatal accidents for all the selected counties, which is consistent with congestion improving safety before the pandemic by reducing drivers' speeds. Since the pandemic, speeds have increased in cities that were highly congested before the pandemic. Although we do not have data on changes in excessive speeding (exceeding 100 miles per hour) for those cities during the pandemic, anecdotal evidence for other cities exists, which suggests such changes in speeding also occurred for the cities that were highly congested. For example, Highway Patrol officers in California have issued nearly 28,500 traffic tickets for speeds over 100mph in 2020, almost double the 2019 total (Baumgaertner and Mitchell (2021); traffic tickets for speeds over 100 miles per hour have increased 30% in Houston, Texas; and along Maryland Route 210, just south of Washington DC, county speed cameras in 2021 caught thousands of vehicles well above the 55 miles per hour speed limit, with the highest recorded speed at 149 miles per hour.[19] Importantly, when a motorist is driving above 50 miles per hour, even modest increases in speed are sufficient to markedly increase a driver's risk of death in an accident (Malyshkina and Mannering (2008)).[20]

---

[18] State laws and public campaigns have increased seatbelt use to 90% in 2020. However, some states have reported that people who were involved in a fatal accident during the pandemic and were found to not be wearing a seatbelt modestly increased. The age of drivers is generally a factor in the occurrence of driving accidents, with younger and older drivers more likely to be involved in an accident than drivers of other ages. However, the short time span of our analysis and COVID's different effects on the vehicle-miles-traveled of motorists of different ages makes it difficult for us to draw any conclusions about drivers' ages being a factor in fatalities before and after the pandemic started.

[19] https://abc13.com/houston-speeding-tickets-100-mph-speeders-drivers-fast/6216651/ and https://www.cnn.com/2022/02/08/politics/record-traffic-fatalities-highway-safety/index.html
We are not aware of any evidence that police enforcement has changed during the pandemic, which could account for the higher incidence of speeding tickets for speeds over 100 miles per hour.

[20] More recently see: https://www.iihs.org/news/detail/new-crash-tests-show-modest-speed-increases-can-have-deadly-consequences

As congestion has declined in previously highly congested cities, more motorists have tended to drive at speeds above the speed limit, which appears to have led to greater fatalities in those cities than before the pandemic.[21] Psychologists and other researchers who study risky driving behavior suggest that driving at greater speeds may reflect people releasing their anxieties from being cooped up or locked down during the pandemic or from working in stressful jobs where they deliver goods or provide essential services in workplaces that expose them to greater risks of becoming infected by the virus. Other risky driving that leads to accidents may occur when people are already stretched to the breaking point, and someone else's inconsiderate driving triggers an angry response that leads to a multi- or single-vehicle collision.

As more people return to work, the increase in VMT will have a first order effect on fatalities. For example, according to our estimates, a 10-15% increase in VMT would lead to a 7-8% increase in fatalities, which is broadly consistent with recent aggregate national fatality statistics.[22] We speculate that some of the bad driving behaviors on uncongested roads that were facilitated by shutdowns will persist, so fatalities are likely to increase further although road congestion has increased. We also speculate that bad driving behaviors developed more in metropolitan areas that were highly congested pre-Covid. The extent of telecommuting and other structural changes in the labor market in the long run also may affect the volume and composition of highway traffic.

---

[21] Hughes, Kaffine, and Kaffine (2022) find that during the pandemic, the frequency of severe accidents increased as accidents fell in California, which suggests that motorists were traveling at higher speeds that increase accident severity. It is possible that drivers' higher speeds above the current speed limits are "appropriate" free-flow speeds based on formulas used in the Highway Capacity Manual. Nonetheless, if a motorist gets into an accident, the higher speed will generally increase the severity of the accident.

[22] According to the most recent NHTSA estimates, fatalities in the first quarter of 2022 increased by 7% relative to the first quarter of 2021. Meanwhile, VMT almost completely returned to pre-shutdown levels (national VMT in February 2022 was equal to 99.3% of national VMT in February 2020). We estimate that the rebound in VMT of 10-15% from its low point during shutdowns (US Federal Highway Administration) would increase fatalities by 4.8-7.2%, which is consistent with NHTSA's estimates. We obtain our estimates by multiplying the increases in VMT by 0.48, which is obtained by subtracting one from the IRR of VMT estimated in column (1) of Table 2.

5. Implications of the Analysis

We have found that the large reduction in VMT caused by the COVID-19 global pandemic has had the opposite effect on highway fatalities in cities that were highly congested before the pandemic compared with the effect of the large reduction in VMT caused by the Great Recession. Because the pandemic has reduced congestion but increased fatalities in those cities it has raised questions about whether a public policy that reduces VMT would necessarily improve highway safety and congestion. We have argued that the key to achieving those improvements is for a public policy not to adversely affect the mix of safe and risky drivers, either because individual drivers change their behavior in response to fewer vehicles on the road or because there is a greater share of risky drivers in the mix.

Unfortunately, it is possible that both congestion pricing and investments in public transit—the policies that are most often advocated to reduce VMT—could increase risky driving behavior on certain roads and at certain times of day. The introduction of congestion pricing could have multiple effects on highway travel conditions and on the mix of drivers, which make its overall effect on automobile fatalities unclear. On the one hand, the reduction in congestion during peak travel periods of the day could lead to higher speeds, especially by risky drivers, which could increase the likelihood of a fatal accident on tolled roads. On the other hand, there may be a greater share of safer drivers along with fewer vehicles on the road during peak travel periods because a high congestion toll may cause risky drivers with lower incomes and lower values of time to shift to off-peak travel periods or to use less traveled but longer alternative routes. However, that shift by risky drivers could increase fatalities if those drivers travel at excessive speeds because there are fewer vehicles on the alternative routes, and they may be less patrolled by the police (Lave and Lave (1999)). Finally, Winston and Yan (2021) find that congestion pricing reduces automobile

fatalities by causing motorists to shift from larger vehicles, which they prefer to drive in congested conditions, to smaller vehicles, which reduces the likelihood of a fatality should a collision occur between those smaller vehicles.

Public transit has experienced a steady decline in its share of travelers to the point where its share was less than the share of telecommuters before the pandemic, and its share has continued to decline during the pandemic.[23] Thus, it would take significant and well-targeted investments in public transit to attract motorists, especially during peak periods. Public transit is thought to be an inferior good; however, suburban rail transit systems have primarily catered to travelers with above-average incomes. So, the construction of new suburban rail systems or substantial improvements in existing systems may attract commuters with above-average incomes and increase the share of less affluent, risky drivers on the road. Investments in bus transit systems that serve lower-income neighborhoods and that could enable more households to have convenient access to more jobs in a metropolitan area could reduce the share of less affluent, risky drivers on the road.[24]

Given the uncertainties about the effects on automobile safety of public policies that could reduce VMT, autonomous vehicles potentially represent the only reliable way to achieve both safer and faster travel. As we discuss in greater detail in chapter 6, autonomous vehicles could virtually

---

[23] For example, congestion in New York City outside of Manhattan has recently increased because some commuters became so frustrated by the reduction in transit service, and possibly by fear of a greater likelihood of catching COVID-19, that they bought cars to commute to work.

[24] Our assessment of the effect of congestion pricing and investments in public transit on automobile safety and the safety of the overall transportation system is not altered if we consider the relative safety performance of automobiles and public transit, which is typically based on a measure of fatalities per passenger mile (see, for example, https://injuryfacts.nsc.org/home-and-community/safety-topics/deaths-by-transportation-mode/). That is, congestion prices that cause risky drivers to shift, for example, to public transit will still improve the overall safety of the transportation system when we account for transit's relative safety performance. Similarly, a transit investment that induces safe drivers to shift to public transit and enables risky drivers to drive faster will still decrease the overall safety of the transportation system when we account for transit's relative safety performance.

eliminate automobile accidents by replacing risky drivers with technology that does not exceed the speed limit, use alcohol or drugs, read and send texts while driving, and the like.[25]

At the same time, autonomous vehicles also could significantly reduce congestion by eliminating incident delays, which account for roughly one-third of all delays. In addition, as we discuss in more detail in chapter 6, autonomous vehicles would effectively increase road capacity in several ways, including: (1) enabling vehicles to travel closer together at higher speeds; (2) allowing road authorities to build or redesign roads with more but narrower lanes; (3) enabling road authorities to expand capacity by eliminating the breakdown lane, given that vehicles would effectively be "inspected" before being able to access the freeway; (4) enabling travelers to perform tasks in the vehicle that give them the flexibility to avoid peak travel; and (5) facilitating the adoption of congestion pricing, which would spread demand for highway capacity more evenly throughout the day.

In sum, a technological innovation could free policymakers from setting policies that potentially tradeoff between congestion and safety and could benefit the public by simultaneously reducing the enormous social costs of congestion and accidents.

---

[25] Comparisons of the safety performance of various transportation modes raise the fundamental question of what it is they are measuring? Is it differences in the modes' technology, types of trips, operators, or something else? Automobiles' safety performance is clearly compromised by dangerous operators—that is, risky drivers. Autonomous vehicles have the potential to improve automobiles' absolute and relative safety performance because risky drivers would no longer operate the vehicle.

Figure 1. Deaths per Billion Vehicle Miles Travelled, 2005–21
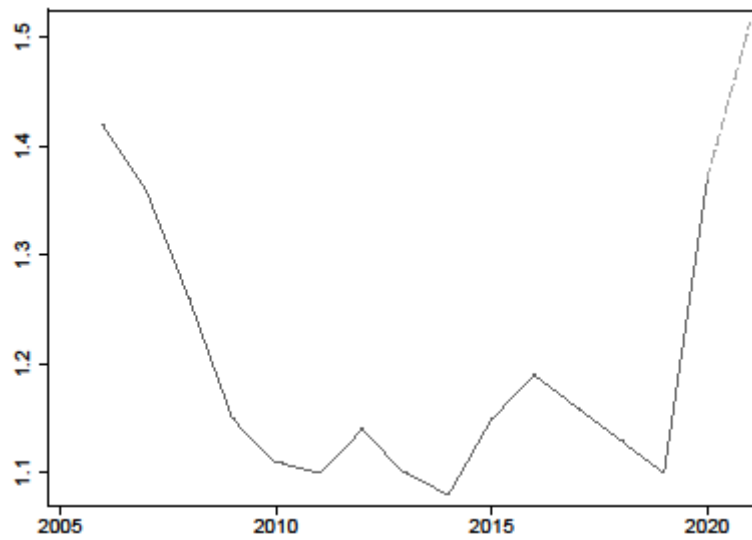
Sources: Fatality Analysis Reporting System (FARS), National Highway Traffic Safety Administration (NHTSA).

Notes: Data for 2021 are projected using the NHTSA's preliminary estimate of a 10.5% increase in deaths per billion VMT from 2020Q1 to 2021Q1.

Table 1. Summary Statistics

Variable	Pre-Pandemic	Post-Pandemic
Monthly fatalities	12.94	13.77
	(10.76)	(11.38)
Monthly vehicle miles travelled (billions)	1.76	1.61
	(1.19)	(1.31)
Cumulative COVID-19 cases (millions)		0.36
		(0.30)
Cumulative COVID-19 deaths (thousands)		12.0
		(14.4)
Number of Observations	730	620

Notes: Pre-pandemic includes data from January 2019 until the middle of March 2020, and post-pandemic includes data from the middle of March 2020 through March 2021.

Table 2. Maximum Likelihood Estimates of the Effects of Congestion on Auto Fatalities as Mediated through Shutdowns

Variable	(1)	(2)	(3)
Post Shutdown × Highly Congested		1.12***	1.12**
log(VMT)	1.48***	1.43***	1.35***
Post Shutdown × log(VMT)	0.95*	0.95***	0.99
Month Fixed Effects	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes
COVID-19 Controls	No	No	Yes
Pseudo R^2	0.53	0.53	0.53
Number of Observations	1,300	1,300	1,300

Notes: Maximum likelihood estimates are from a Poisson regression where number of fatalities is the dependent variable. We report incidence rate ratios (IRR). An IRR > 1 implies a positive marginal effect and an IRR < 1 implies a negative marginal effect. Statistical significance is indicated at the ***1%, **5%, and *10% levels.

Table 3. Sources of Risky Behavior in Fatal Crashes Before and After Shutdowns in Select Counties

	Alcohol		Drugs		Speeding	
	Pre	Post	Pre	Post	Pre	Post
Cook County (Chicago)	67%	59%	89%	86%	36%	56%
Harris County (Houston)	42%	40%	28%	22%	28%	32%
Dallas County	45%	45%	70%	69%	41%	53%
Los Angeles and Riverside Counties	23%	23%	58%	32%	40%	53%

Notes: "Pre" corresponds to January 1, 2019–December 31, 2019. "Post" corresponds to March 1, 2020–April 1, 2021. Pre-shutdown data were obtained from FARS. Post-shutdown data were obtained from direct inquires to officials in Harris, Dallas, Cook, and Los Angeles/Riverside counties.

Chapter 3. Is Public Bus Transit a Competitor
or a Subordinate to Public Rail Transit?

Austin J. Drukker and Clifford Winston

1. Introduction

Public ownership and operation of urban transit has been a long-running experiment that followed in the wake of the financial contraction of private transit companies. The growth of household income and automobile ownership in the 1950s caused private transit systems to incur large financial losses and confronted city governments with three choices: (1) provide financial assistance and regulatory relief to the private systems and give them a chance to make adjustments that might enable them to once again become financially viable competitors; (2) allow private transit companies to fail if they could not compete against the automobile without assistance; or (3) convert transit companies to subsidized public enterprises, as a few cities had already done.

By 1964, city governments overwhelmingly chose the last option, especially because the Urban Mass Transportation Act of 1964 promised them federal grants. Unfortunately, subsidizing public transit service did not improve its competitive position, and its market share of travelers in most urban areas has continued to decline to the point where it was exceeded by the share of telecommuters before the pandemic. Currently, fewer than 5% of commuters use public transit. At the same time, transit's subsidies have continued to grow and now exceed $20 billion annually.

Because of its declining ridership and growing subsidies, transit's social desirability was in serious question before the pandemic (Winston (2013)).[26] The adverse trends in ridership and

---

[26] Anderson (2014) studies the Los Angeles metropolitan area and argues that the net benefits of public transit systems may be much larger than believed because they reduce congestion. However, considering the findings in chapter 2, it is important to assess whether safety is compromised when congestion is reduced. In addition, transit's benefits from reducing congestion could be achieved at a lower cost by implementing congestion pricing.

subsidies have accelerated since the pandemic making transit's social desirability highly doubtful. Since the COVID-19 pandemic began, transit ridership is down more than 25% and it is not expected to return to pre-pandemic levels for at least a decade because some people who used to commute on public transit will continue to work from home (Quiroz-Gutierrez (2021)). Transit also has continued to lose passengers because as its ridership has declined, it has reduced service frequency and experienced higher crime rates (Bosman et al. (2022)) causing some passengers to shift to alternative modes, such as auto, and possibly not returning to transit even when the COVID-19 pandemic is over and when transit's service and safety improve.

In response to the sharp drop in ridership and revenue, public transit has received additional taxpayer-funded subsidies in the form of nearly $70 billion of supplemental funding provided during the pandemic. Few limits have been set on the amount of time transit agencies have to spend the supplemental funding or on when they have to return any of it.

Generally, public transit's poor performance can be explained by demographic changes and operating disadvantages that have resulted in average load factors (the percentage of seats filled by paying passengers throughout the day) of less than 30% for most systems (US Federal Transit Administration (2020)), which is not close to a break-even load factor. Transit cannot keep pace as households' high levels of automobile ownership enable them to move to new suburban communities and the exurbs (Baum-Snow and Kahn (2005)).[27] Transit's operations are characterized by the "streetcar problem," when during morning peak periods it commits seat capacity provided by large buses or trains composed of several passenger cars that is filled with passengers who access transit from their residences and board enroute to a central destination.

---

[27] Expansion of bus and rail are primarily limited by the cost, political difficulties, and the time needed to add new routes to increase service frequency (in the case of bus) or to add new track to expand coverage (in the case of rail).

Almost the same capacity is offered throughout the day, but it is not filled again until evening peak periods when passengers return from the central business district to outlying residential areas.

Transit's load factor has fallen further when bus systems have expanded their networks and new rail systems have been built where there is low potential demand for public transit. Buses in some cities, such as Los Angeles, were once able to fill nearly 50% of their seats over the course of a day, but today, bus transit's average load factor throughout the day is less than 15% and rail transit's load factor is less than 30% (US Federal Transit Administration (2020)). We are not aware of a precise estimate of the break-even load factor for public transit, but break-even load factors for airlines are roughly 70%.[28]

An important feature of urban transportation is that many metropolitan areas are served by both bus and rail transit systems. When cities took over bus and rail operations, they could have developed their operations such that multimodal transit service generated competition that would improve transit efficiency and increase its overall market share while benefiting urban travelers. Or they could have developed their operations such that they exacerbated transit inefficiencies by raising costs and increasing deficits that must be funded by taxpayers.[29]

We are not aware of any evidence that suggests policymakers have sought to increase urban transit competition. We hypothesize that policymakers in cities with rail systems, which tend to be more heavily subsidized than bus systems, prefer that their rail system does not compete with their bus system. That is, they have designed their bus system operations to play a subordinate and occasionally complementary role to their rail system. Such a system would increase the costs

---

[28] It could be argued that transit's financial problems, especially rail transit's, are attributable to inefficient road prices that do not account for automobile externalities and reduce their costs to travelers. However, Winston and Shirley (1998) find that transit's share would decline if the urban transportation modes charged efficient marginal-cost prices and if transit provided optimal service frequency.

[29] Ley (2022) discusses transit service in Queens where the bus system is used to serve areas that are not served by the rail system. The travel speed on one of the system's routes is 6.5 miles per hour.

of the bus system by reducing its network efficiency and average load factors, without necessarily improving the financial condition of the rail system.

In this chapter, we document the effect of rail competition on bus costs by analyzing a sample of bus systems that serve 94 US metropolitan areas and a sample of light and heavy rail systems that serve 28 of those metropolitan areas. We find that the presence of a rail system significantly increases bus costs, with newer rail systems exacerbating this effect relative to older systems, and that this effect accounts for a notable share of transit's overall deficits. We discuss the possible mechanisms at work and suggest that they strengthen the case for the private sector to play a greater role in urban transportation to improve its efficiency.

2. Conceptual Approach

Multimodal transit systems in US metropolitan areas appear to have a competitive interface because they serve many of the same origins and destinations. For example, when bus and rail provide downtown trips, bus routes often run directly above subway tunnels. Generally, rail fares are higher than bus fares while rail offers faster trip times, especially as the trip distance increases. In theory, transit competition could lead to efficiency and reduce fares, but this does not occur in practice because transit fares are usually established through elaborate political negotiations, including contributions by local jurisdictions based on property or other taxes, and are insensitive to competitive forces that reduce costs.

Costs are therefore the relevant performance measure of bus transit that could be affected by the presence of rail service in a metropolitan area, and vice-versa. In addition, the same agency operates bus and rail transit service in most cities that are served by both modes, so their cost structures are related, and they suffer from the same inflated management overhead.

We hypothesize that transit agencies do not want to attract any unnecessary attention by running excessive deficits, so they avoid having bus and rail compete by giving rail competitive advantages to increase its revenues and decrease its (larger) budget deficits and drain on the public purse. Operationally, transit agencies might reconfigure their bus systems suboptimally to "force feed" their rail systems, inflating bus operating costs by exacerbating peaking, empty backhauls, deadheading, and the like. We test our hypothesis by estimating how the presence of rail operations affects the cost of bus operations.

We are not aware of previous work that has analyzed the interdependence of bus and rail transit costs; thus, we begin with a model that takes the following generic form:[30]

$$\text{bus cost} = f(\text{rail, output, capacity, input prices, service quality, other influences}). \quad (1)$$

The presence of rail is captured by a dummy variable, which we assume is exogenous because civic boosters and city officials play a vital role in bringing a rail system to a city through their considerable efforts to generate political support and obtain outside (usually federal) funding, as rail systems are thought to enhance a city's reputation and prestige (Winston and Maheshri (2007)). Such efforts can take decades; for example, it took some thirty years for Baltimore's light rail system to evolve from a proposal to serving passengers. In addition, rail's profit motives evaporated when local governments took ownership of those systems and tightly regulated fares, which included subsidies, and did not adjust their service offerings in response to changing demographics. Even the private rail systems that existed previously were subject to pricing regulations that compromised profitability (Gómez-Ibáñez and Meyer (1984)).

---

[30] Viton (1981) and Winston and Shirley (1998) developed equilibrium models of bus and rail transit competing against each other and automobile modes.

The standard measure of a bus system's output is passenger-miles traveled, defined as the cumulative sum of the distances ridden by each passenger. Capacity is measured by a system's route-miles; input prices include fuel, wages, and capital costs for vehicles; and service quality is typically measured by travel speed and reliability.[31] Regarding other cost influences, bus transit companies are under less pressure to reduce costs if they receive additional revenues from dedicated taxes and they may incur additional costs by operating in difficult weather conditions or on unfavorable terrain.

Although equation (1) is a standard specification of transportation costs, the primary goal of our analysis is to estimate the effect that the presence of a rail system has on bus operating costs, so it is important for us to exclude variables that could capture the effect of the rail system. For example, it is inappropriate to include passenger-miles in the specification because the presence of an alternative mode of travel, namely, rail, will decrease bus ridership. We therefore cannot hope to identify the true effect of rail presence on bus costs if we include bus passenger-miles as an explanatory variable. In place of passenger-miles, an important omitted variable, we include *current* population, which is highly correlated with passenger-miles but is arguably uncorrelated with the presence of rail, which appears to be explained by political factors and long-standing boosterism. Note that cities with large populations (e.g., Los Angeles) did not have a rail transit system until recently and older cities with smaller populations (e.g., Boston) have had a rail system for more than a century. We explore how our findings are affected by the choice of including current population or passenger-miles in the specification.

We also should not include regional fixed effects because those would be perfectly correlated with the rail dummy, which does not vary for individual transit systems over our period

---

[31] Buses use petroleum and diesel fuel, but their prices are highly correlated.

of study.[32] Route-miles and average trip distance also are potentially problematic as explanatory variables because rail could serve destinations that bus might otherwise serve and because systems with rail tend to have shorter average trip distances for bus, especially where buses provide more feeder service and fewer long-distance trips.

Given the preceding considerations, we specify our base case model of bus costs for transit system i at time t in a log-linear form as a function of the rail dummy and purely exogenous influences as:

$$\ln(\text{cost}_{it}) = \beta\, \text{rail}_i + \alpha_1 \ln(\text{population}_{it}) + \alpha_2 \ln(\text{fuel price}_{it}) + \alpha_3 \ln(\text{dedicated taxes}_{it})$$
$$+ \alpha_4\, \text{days freezing}_{it} + \gamma_t + \varepsilon_{it} \tag{2}$$

We measure bus costs as operating expenses because it is not clear why the presence of rail also would affect bus capital costs; we found through sensitivity tests that the results are not changed when we specify total costs, which include operating and capital costs.[33] The explanatory variable of interest is a time-invariant rail dummy indicating the presence of a rail system at location i. We classify a rail system to be either heavy rail (e.g., Washington's Metrorail system) or light rail (e.g., Dallas's DART Light Rail). As noted, population is used as an explanatory variable in place of passenger-miles. Fuel prices, specifically, the average price per gallon of gasoline, is an exogenous influence on bus operating costs. Wages and capital costs do not vary sufficiently to permit their inclusion in the specification and their exclusion did not affect our main findings. Dedicated taxes levied by state or local governments generate revenues that can be used to offset excessive operating costs. We contend that those are largely governed by exogenous political forces. Such revenues reduce the incentive for transit systems to operate efficiently, so

---

[32] There are two exceptions: Phoenix's Valley Metro Rail, which began operations in December 2008; and Virginia Beach's Tide light rail, which began operations in August 2011.
[33] Capital costs include the annualized value of the bus fleet and structures.

we expect dedicated taxes to have a positive effect on operating costs, holding population constant. Days freezing refers to the number of days in the calendar year for which temperatures are below 32°F, an influence that increases bus maintenance costs and is clearly exogenous. We also include year fixed effects, γ_t, to control for macroeconomic factors that affect all transit systems in a year, and ε_{it} is the error term.[34]

3. Sample and Data Sources

To construct our estimation sample, we considered the 100 largest metropolitan areas by population. After making some appropriate adjustments and accounting for limitations of some of the areas because of missing data, we assembled a balanced panel of 94 transit systems with annual data from 2002–2018 that are included in the US Department of Transportation's (DOT) National Transit Database.[35]

Table 1 summarizes the transit systems in our sample, which include 28 bus-and-rail systems and 66 bus-only systems. This sample enables us to identify the effect of a rail system on bus costs in both a geographic dimension (that is, some systems do, and some do not have rail service at the same point in time) and a temporal dimension (that is, systems that added rail service did so at different points in time). We classify the systems as old if the bus-and-rail system or bus-only system was operating during or before 1960 and new if the bus-and-rail system or bus-only system began operating after 1960. We conducted sensitivity tests and found that our conclusions do not change much if we use 1965 or 1970 as a cutoff year to distinguish between new and old systems.

---

[34] We were unable to obtain data on exogenous measures of service quality.
[35] We dropped Hartford CT, New Haven CT, McAllen TX, and Poughkeepsie NY due to missing data, and we dropped Ogden UT and Provo UT because they belong to the same transit system as Salt Lake City UT.

We distinguish between those systems because we hypothesize that metropolitan areas attempt to increase ridership of their new, expensive rail systems by giving them priority over bus in terms of favorable routes and type of service. Metropolitan areas may be less likely to prioritize old rail systems, so the effect of those systems on bus costs may be less than the effect of new rail systems on bus costs. Rail was introduced to US urban transit beginning with the Boston subway in 1897; thus, it is not surprising that a narrow majority of bus-and-rail systems are old, but a clear majority of bus-only systems are relatively new.[36]

Table 2 presents summary statistics for the variables used in our estimations. Most of the data we use come from the DOT's National Transit Database. Population is at the level of the metropolitan statistical area and is from the US Census Bureau.[37] Weather data are from the National Oceanic and Atmospheric Administration and data on fuel prices are from the Energy Information Administration. All dollar values are deflated using the consumer price index and are expressed in real 2019 dollars.

The old transit systems carry more passengers over a larger network, operate in larger cities, and as expected, have higher total operating costs than the new systems. However, the systems have comparable operating costs per passenger-mile, suggesting that the old systems are unable to exploit possible economies of size and route density to significantly reduce operating costs.

---

[36] As a point of qualification, some cities that only have buses today once had electric streetcars, which were replaced by buses once buses became less costly to operate. The earliest electric streetcar systems predate subways by a few decades.

[37] We collapse the transit systems of Chapel Hill NC and Durham NC into one system because they belong to the same metropolitan statistical area.

4. Estimation Results

We present ordinary least squares estimates of our base specification in equation (2) in column 1 of table 3. The central finding is that the presence of rail *increases* bus costs, and the effect is large and statistically significant: rail's presence increases bus costs by roughly 71%.[38] We discuss below the specific mechanism by which the presence of rail increases bus costs after we finalize our discussion of the estimates. The remaining variables in the model have their expected signs, as population (a proxy for demand), gasoline prices, dedicated taxes, and days below freezing also bear a positive relationship to bus costs. Overall, the model fits the data well, with an R^2 of .83, even though we have not included passenger-miles.

The second column of the table begins our exploration into how the presence of rail affects bus costs; that is, what variables might the rail dummy be capturing? Column 2 of table 3 shows that when we replace population with passenger-miles, it has a statistically significant effect on costs and noticeably reduces the effect of the rail dummy, which suggests that rail affects bus costs by causing changes in bus operations in ways that reduce its patronage.

Of course, passenger-miles may be endogenous because it is correlated with unobserved influences on bus costs. Thus, we instrument passenger-miles with population, which is correlated with passenger-miles but is not causally related to costs once we control for passenger-miles in the specification.[39] As we show in column 3 of table 3, the coefficient on the rail dummy variable falls even further when we estimate the model in this way.[40] In sum, it appears that the presence of rail increases bus costs primarily by affecting its operations in ways that decrease its traffic.

---

[38] Recall that the percentage impact of a dummy variable on the dependent variable in a log-linear regression is equal to $100 \times (\exp(b - \mathrm{SE}(b)^2/2) - 1)$, where b is the estimated coefficient on the dummy variable and $SE(b)$ is the estimated standard error of b.

[39] The first-stage regression using population as an instrument for passenger-miles had an F statistic of 33.73.

[40] Note that when we instrument passenger-miles with population, the coefficient for passenger-miles increases and the standard error slightly decreases.

We checked whether the findings were sensitive to the size of the transit systems by re-estimating equation (2) for the top 60 metropolitan statistical areas based on population, composed of 28 bus-and-rail systems and 32 bus-only systems. The coefficient estimates for the effect of rail on bus operating costs with and without passenger-miles are very similar to the estimates obtained from the full sample of 94 transit systems. We also estimated equation (2) using a sample of transit systems with system-wide passenger-miles above the median, which resulted in 27 bus-and-rail systems and 20 bus-only systems. Using that sample did not result in any material changes to our findings. Finally, the results did not change much when we excluded New York's transit system from the sample, which accounts for about 40% of transit passengers nationwide.

We have assumed that the presence of rail has a homogeneous effect on all bus systems' operating costs, regardless of any important differences between the systems. A key distinguishing feature between transit systems that may influence the effect of rail's presence on bus operating costs is the age of the system providing rail and bus service. For example, it might be expected that older systems, where rail and bus "grew up together," would be designed in a harmonious way such that rail has a smaller effect on bus costs compared with newer systems, where rail service was added to an existing system and bus operations had to be adjusted to accommodate rail.

In table 4, we present estimation results for the base specification without passenger-miles for old and new systems, and for the specifications with passenger-miles with and without passenger-miles instrumented by population. We set a cutoff year of 1960 to distinguish old from new systems and we performed some sensitivity tests around this cutoff point. Comparing columns (1) and (4), we find, as hypothesized, that the coefficient for the rail dummy is larger for the new systems than for the old systems when we do not control for passenger-miles. This finding is consistent with the hypothesis that city officials and planners wanted to encourage rail ridership

in the new systems even to the detriment of bus, because it was much more expensive to construct those systems in the more recent period than in the past.[41]

However, when we control for passenger-miles, the rail coefficient for the old systems becomes larger than the rail coefficient for the new systems, suggesting much of the effect of rail on new systems' bus operating costs is captured by passenger-miles. Because new rail systems are those where bus routes were reconfigured to curtail their line-haul service to downtown destinations and feed the rail system by providing more suburban service, higher passenger-miles on rail are partly achieved by using feeder bus service to carry more passenger-miles, which is likely to be more costly than using bus to carry passengers for line-haul service. It is conceivable that the agencies' reduction in bus line-haul service and increase in feeder service could reduce total bus operating costs overall. But we did not find that to be the case, perhaps, as we discuss below, because of the relatively higher cost of feeder service.

Note the small effect of rail in column (6). In contrast, rail's effect on bus operating costs in old systems, as shown in column (3), is larger and apparently includes additional influences that collectively increase bus operating costs roughly 33%. We conducted sensitivity tests using alternative cutoffs for new systems as operating during or after 1965 and 1970 and found no substantive change in the results; namely, the presence of rail has a greater effect on bus operating costs for new systems than for old systems but controlling for passenger-miles explains most of the effect for new systems, although additional influences still affect operating costs for old systems. We discuss those possible influences below.

---

[41] Rosenthal (2017) shows that the construction costs of New York's Second Avenue Subway amount to a record $2.6 billion per mile. The cost of Honolulu's light rail system is likely to significantly exceed $0.5 billion per mile by the time it is completed.

Based on the parameter estimates in table 4, we estimate that the presence of rail in old transit systems raises annual bus operating costs by about $2.1 billion due to both demand and pure cost effects, and that the presence of rail in new transit systems raises annual bus operating costs by about $2.7 billion, due primarily to demand effects. The combined $4.8 billion increase in operating costs for new and old systems accounts for a notable share of public transit's total operating and capital subsidies, which exceed $25 billion (Winston and Karpilow (2020)).

5. Discussion

What explains the different effects that new and old rail systems have on bus costs? We hypothesize that cities that have built expensive, relatively new rail systems, especially ones that serve suburbs, do not want to be criticized for running excessive rail transit deficits, which might jeopardize additional rail subsidies for system expansion or modernization. Thus, the newer systems attract some riders who may have used bus by removing buses from line-haul service and relegating them to feeding the rail system.[42]

This change in operations increases bus costs because feeder service is more costly to operate than line-haul service. Feeder buses are less utilized because they spend dwell time waiting at rail stations and they also operate less efficiently by stopping more often, traveling at slower speeds, and so on. Bus system costs also may be higher because fewer resources are allocated to keep buses in good condition. Finally, the change in operations converts for many passengers

---

[42] We were unable to find data that indicated the share of buses that were previously used for linehaul service and relegated to feeder service after a rail system was built. But anecdotal evidence supports the hypothesis that certain bus routes are converted to feeder service when a new rail system is built. For example, the US Department of Transportation's (1995) report on bus–rail integration in St. Louis in the early 1990s notes that "extensive advertising and promotions" were used to inform bus ridership that "some bus routes would be slightly modified to become a feeder network for the light rail system," and St. Louis employed a "Rail Ambassadors" program to encourage rail ridership.

what was once a direct bus trip into a more time-consuming bus trip plus a rail trip, which reduces bus demand and load factors.

Bus ridership, as measured by passenger-miles, aggravates these cost-increasing effects because much of it is concentrated during peak hours, increasing dwell times at bus stops and rail stations, and so on. Thus, when we controlled for passenger-miles in the specification, we found that it captured much of the original effect of the rail dummy variable. That is, the rail dummy was capturing demand-related effects that increase costs.[43]

However, we also found in old systems that the rail dummy is capturing some additional effects on bus costs even when we controlled for passenger-miles. One explanation is that there may be omitted variables that raise operating costs but do not affect demand, which the rail dummy was capturing. For example, suppose the presence of a rail system forces a bus system to hire additional workers who are used to help coordinate busses with rail operations. This would increase bus operating costs but not affect demand, so the rail dummy would be capturing pure cost-related influences on bus costs. The effect also might arise from higher wages for rail operators and mechanics—and perhaps managerial overhead as well—that spills over to bus operations such that bus drivers and mechanics have higher wages in cities that also have rail systems, especially older ones, than in cities with only bus service.

An alternative explanation is that the inclusion of passenger-miles may create endogeneity bias, which biases the rail dummy variable, so it appears to capture pure cost-related effects. For example, although fares are regulated, costs may affect passenger-miles by reducing service

---

[43] It could be argued that passenger-miles reflects the equilibrium of demand and supply. Transportation supply is typically measured by seat-miles. Seat-miles multiplied by average load factor yields passenger-miles. As discussed, transit's low load factors exist because its operations do not enable it to keep pace with the nation's demographic and socioeconomic changes. Because rail's operations are even less flexible than bus's operations, a system that favors rail will be more disadvantaged by demographic and socioeconomic changes that affect demand.

quality. Thus, there may be reverse causality between those variables. Bus passenger-miles are negatively correlated with the rail dummy, so the bias to bus passenger-miles could bias the estimate of the rail dummy. However, when we tested for this possibility by instrumenting passenger-miles with exogenous population to reduce the endogeneity bias, we found that the pure cost-related effect captured by the rail dummy still existed for old systems.

6. A Greater Role for the Private Sector in Urban Transportation?

Public urban bus and rail transit have accumulated vast inefficiencies and have developed a symbiotic relationship that has increased bus costs and is likely to ensure that transit's financial problems and declining market share persist. In addition, bus service is relegated to feeding rail service when it could benefit the public in certain situations by substituting for it. For example, both bus and rail transit significantly curtailed service during the COVID-19 pandemic and greatly inconvenienced some essential and other workers who normally relied on transit to travel to their workplaces (Verma (2020)). Those workers would have been less inconvenienced if new bus service were initiated on line-haul routes served by rail transit. By completely substituting for rail line-haul service and offering greater frequency than rail was offering during the pandemic, the new bus service would have reduced travelers' waiting time at a possible cost of increasing their travel time and would have reduced taxpayers' overall transit subsidies because rail is more costly than bus transit to operate.

The failure to use bus more effectively in metropolitan areas with bus and rail transit service harms lower income travelers because bus riders are poorer than rail riders and less able to afford an alternative mode, such as auto, and have fewer opportunities to work at home. Bus has recovered much of the ridership that it lost with the onset of COVID in such cities as Washington,

DC, and Boston, while rail is struggling to recover even half of the ridership that it lost in those cities (Gelinas (2022b)). At the same time, rail has cut its service in those cities and bus routes and frequencies have not been adjusted in response to demographic changes and travelers' dependence on bus service, especially harming essential workers.

In theory, bus systems are more flexible the rail systems and their routes could be more easily reconfigured to adapt them to modern traffic patterns, improve the speed and reliability of service, and attract more passengers. However, proposals to redesign a bus network often are shelved because they are met with strong opposition from riders who claim they will be inconvenienced because stops will be eliminated. Thus, current average bus speeds in the densest metropolitan areas in the country are less than 10 miles per hour, while average bus speeds in the New York metropolitan area are about 8 miles per hour (Ley (2022)).

In addition to being hamstrung to improve service quality, bus and rail transit systems seem incapable of taking the most basic actions to reduce their deficits and taxpayer-funded subsidies. To take some examples, the Washington, DC, Metropolitan Transit Authority has lost millions of dollars in revenue because roughly 1 in 3 Metrobus rides goes unpaid as drivers do not attempt to prevent fare evasions (George (2022)). According to the Metro bus operating training manual, drivers are discouraged from bringing attention to someone who skips a fare and should not stop a fare evader or delay a trip if a passenger avoids paying. In New York City, nearly 30% of bus riders and 8% of subway users do not pay their fare, costing the city more than $300 million a year, according to new data released by the state-operated Metropolitan Transportation Authority (Rosenberg (2022)). And fare evasion has contributed to putting the Seattle and Tacoma Sound Transit System on a financially unsustainable trajectory and facing insolvency (Schulz (2022)). Schulz points out that the lack of enforcement is to blame for widespread fare evasion and that

New York, Seattle, and several other cities with transit systems have reduced or completely withdrawn enforcement efforts.

An alternative approach to improving the efficiency of public transit is to privatize it and not force bus to be subordinate to rail. Empirical evidence indicates that privatization of all US bus systems would produce large cost savings and eliminate bus subsidies (Jerch, Kahn, and Li (2017), Winston and Shirley (1998)).

Private van and minibus service, which have achieved success throughout the world because of their low costs and responsive service, could improve urban transportation if they were allowed to serve more US urban areas. Currently, such service exists in a handful of major US cities, but the operators tend to keep a low profile because, although some companies are legal, others have not obtained a license to operate. Nonetheless, they exist because they provide a highly valued, low-cost service, with fares that are usually only a few dollars and with a route network made by and for working class people. Correal (2018) provides examples of private vans' service areas and fares in the New York City-New Jersey metropolitan area, while Reiss (2014) provides a graphic overview of the service provided by "dollar vans" in New York City, which shows their routes are generally not served by subways and buses. Said (2019) reports that Via, a private transit operator, will run on-demand vans in Cupertino, California, which are eagerly awaited because the public bus option run by Santa Clara Valley Transportation Authority is too slow. Indeed, some companies in the San Francisco Bay Area, such as Google, are dissatisfied with the public transportation system and provide private transit service to their employees.

The developers of a new app, Dollaride, are hoping to generate more business for private vans in New York City (and possibly elsewhere) and to challenge public bus transit in the same way that Uber has taken on the taxi industry (de Freytas-Tamura (2019)). The app allows users to

see where licensed vans are operating within a one-mile radius, as well as where they are headed, helping commuters decide whether taking a van might be faster than waiting for a bus. The app allows van drivers to spot passengers more easily and plan for demand and possibly future routes. The app also may encourage unlicensed vans to register with the city.

Private urban rail systems exist in Hong Kong and Japan, taking advantage of high traffic density and multiproduct operations that include non-transportation services. It is an open question whether any US urban rail system could be as financially successful as those Asian systems. It is more doubtful that urban transit public-private partnerships would significantly improve US public transit systems.[44]

As we discuss in more detail in chapter 6, greater competition *and* innovation is critical for improving urban transportation. The private sector's role in improving urban transportation will surely increase when autonomous vehicles are widely adopted. Households will effectively have their own low-cost transit system that provides good service to any destination in a metropolitan area, which will make it difficult for public bus and rail transit to survive financially, even if they are automated (Winston and Karpilow (2020)).

Instead of seeking political support to obtain additional subsidies, bus and rail transit systems should focus on how, if at all, they could possibly complement autonomous vehicles to justify their continued operations. For example, on-demand autonomous transit is being explored in Arlington, Texas (Rideshare, Automation, and Payment Integration Demonstration (RAPID)) with the hope of becoming a viable part of an autonomous transit future.

---

[44] Austen (2022) reports that Canada has had mixed experiences with public-private urban transit partnerships. Rail transit projects in Ottawa, Toronto, and Montreal have been marred by cost overruns and delays, while projects in Vancouver and Kitchener/Waterloo are regarded as successful.

Privatized bus and rail systems are likely to have greater flexibility and financial incentives than public transit systems to adapt to a world of autonomous urban transportation. Importantly, their financial survival would be subject to the market test of providing services that respond effectively to changes in households' socioeconomic and demographic characteristics and that increase social welfare. If private autonomous transit companies cannot compete effectively with autonomous cars, it would be unwise for government to repeat the mistake it made with public transit by taking ownership of those companies.

Table 1. List of Transit Systems Considered

Bus-and-Rail Systems		Bus-Only Systems			
City	Age	City	Age	City	Age
New York	Old	Riverside	New	Albany	Old
Los Angeles	New	Detroit	Old	Knoxville	Old
Chicago	Old	Tampa	New	Baton Rouge	New
Dallas	New	Orlando	New	Oxnard	New
Houston	New	San Antonio	New	El Paso	New
Washington	New	Las Vegas	New	Allentown	New
Miami	New	Austin	New	Columbia	New
Philadelphia	Old	Cincinnati	Old	Sarasota	New
Atlanta	New	Kansas City	New	Dayton	New
Phoenix	New*	Columbus	Old	Charleston	New
Boston	Old	Indianapolis	Old	Greensboro	New
San Francisco	New	Nashville	Old	Fort Myers	New
Seattle	New	Providence	New	Stockton	New
Minneapolis	New	Milwaukee	Old	Boise City	New
San Diego	New	Jacksonville	New	Colorado Springs	New
Denver	New	Oklahoma City	New	Little Rock	New
St. Louis	New	Raleigh	Old	Lakeland	New
Baltimore	New	Memphis	New	Akron	New
Charlotte	New	Richmond	Old	Des Moines	New
Portland	New	Louisville	Old	Springfield, MA	New
Sacramento	New	Birmingham	New	Winston-Salem	New
Pittsburgh	New	Grand Rapids	New	Deltona	New
Cleveland	Old	Rochester	Old	Madison	New
San Jose	New	Tucson	Old	Syracuse	New
Virginia Beach	New*	Fresno	Old	Durham / Chapel Hill**	New
New Orleans	Old	Tulsa	New	Toledo	New
Salt Lake City	New	Honolulu	Old	Wichita	Old
Buffalo	New	Omaha	New	Augusta, GA	New
		Worcester	New	Palm Bay	New
		Bridgeport	New	Jackson, MS	New
		Greenville	New	Harrisburg	New
		Albuquerque	Old	Spokane	New
		Bakersfield	Old	Chattanooga	Old

Notes: Old systems are bus-only systems that were in operation before 1960 or bus-and-rail systems where rail was in operation before 1960. Rail refers to light rail or heavy rail. The New Orleans streetcar system is classified as light rail. * Indicates that operations began mid-sample. ** Indicates that the transit systems were combined to coincide with the Metropolitan Statistical Area definition.

Table 2. Summary Statistics

	Overall		Old Systems		New Systems	
	Mean	SD	Mean	SD	Mean	SD
Operating Cost	146,339,535	310,880,544	243,229,816	536,493,923	113,173,560	164,546,022
Passenger-Miles	130,261,960	246,240,983	190,040,717	363,314,781	109,799,425	186,155,769
Population	2,113,140	2,676,142	2,637,806	3,960,055	1,933,254	2,030,857
Income	41,623	10,399	41,739	8,731	41,583	10,916
Dedicated Taxes	44,889,469	200,676,480	84,490,654	360,536,636	31,333,807	94,463,620
Route Miles	533	425	500	315	544	456
Route Density	186	243	281	406	153	136
Fuel Price	3.551	0.789	3.581	0.800	3.540	0.785
Days Freezing	65	49	74	41	62	52

Notes: Old systems are bus-only systems that were in operation before 1960 or bus-and-rail systems where rail was in operation before 1960. All dollar values are expressed as 2019 dollars. Route density is thousands of people per route per year. Fuel price is dollars per gallon.

Table 3. Estimation Results

Log Operating Expenses	(1)	(2)	(3)
Rail	0.555***	0.399***	0.204***
	(0.194)	(0.075)	(0.072)
Log Population	1.199***		
	(0.105)		
Log Passenger-Miles		0.835***	0.936***
		(0.032)	(0.031)
Log Gasoline Price	3.378***	0.223	–0.146
	(1.246)	(0.289)	(0.400)
Log Dedicated Taxes for Operations	0.014**	0.002	–0.002
	(0.007)	(0.002)	(0.003)
Days below Freezing	0.004***	0.001***	0.001***
	(0.001)	(0.000)	(0.000)
Estimator	OLS	OLS	IV
R^2	0.833	0.958	0.952
No. of Observations	1,596	1,596	1,596

Notes: All specifications include year fixed effects. The specification in column (3) uses log population as an instrument for log passenger-miles. Robust standard errors clustered at the transit-system level are shown in parentheses. Statistical significance is indicated at the ***1%, **5%, and *10% levels.

Table 4. Estimation Results for Old Systems versus New Systems

Log Operating Expenses	Old Systems			New Systems		
	(1)	(2)	(3)	(4)	(5)	(6)
Rail	0.585***	0.482***	0.295**	0.651***	0.388***	0.191**
	(0.183)	(0.115)	(0.134)	(0.239)	(0.092)	(0.082)
Log Population	1.179***			1.185***		
	(0.104)			(0.133)		
Log Passenger-Miles		0.852***	0.943***		0.818***	0.916***
		(0.058)	(0.060)		(0.038)	(0.035)
Log Gasoline Price	5.719***	–0.352	–1.228	2.026*	0.521	0.350
	(1.640)	(0.624)	(0.815)	(1.180)	(0.331)	(0.362)
Log Dedicated Taxes for Operations	–0.001	0.003	0.001	0.025***	0.001	–0.004
	(0.009)	(0.004)	(0.004)	(0.008)	(0.003)	(0.004)
Days below Freezing	0.003	0.002**	0.001	0.004***	0.001*	0.001**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Estimator	OLS	OLS	IV	OLS	OLS	IV
R^2	0.902	0.973	0.970	0.825	0.955	0.949
No. of Observations	407	407	407	1,189	1,189	1,189

Notes: Old systems are bus-only systems that were in operation before 1960 or bus-and-rail systems where rail was in operation before 1960. All specifications include year fixed effects. The specification in columns (3) and (6) use log population as an instrument for log passenger-miles. Robust standard errors clustered at the transit-system level are shown in parentheses. Statistical significance is indicated at the ***1%, **5%, and *10% levels.

**Chapter 4. Measuring the Benefits of Ridesharing Service to Urban Travelers:
The Case of the San Francisco Bay Area**

Hyeonjun Hwang, Clifford Winston, and Jia Yan

1. Introduction

Ridesharing consists of drivers who provide private trips with their own car without intervention from a regulatory authority; passengers who use an app on their smart phone to request transportation from the origin indicated on their phone to various destinations; and transportation network companies (TNCs), such as Uber Technologies, Inc. (hereafter Uber) or Lyft, which have developed a smart phone application that matches passengers' demand for trips with drivers' supply of automobile transportation.

Ridesharing services have grown rapidly since 2010 when Uber launched its first on-demand car service, UberCab, in San Francisco. In 2019, Uber completed some 14 million trips per day, serving more than 600 cities in 65 countries around the world, and its main competitor, Lyft, launched in 2012, completed more than one million trips per day, serving 300 US and two Canadian cities. Since 2017, the share of trips by ridesharing has exceeded taxi's share of trips. By January 2019, daily travel in New York City, for example, consisted of 462,113 trips by Uber, 149,142 by Lyft, and 271,135 by taxi.[45] The growth of platforms such as DiDi in China and Ola Cabs in India indicates that ridesharing services also are widely used in other countries.

Drivers have been attracted to ridesharing because they do not have to purchase medallion permits and meet occupational licensing requirements that apply to taxi drivers

---

[45] Taxi & Limousine Commission, New York City, NY, https://www1.nyc.gov

(Cramer and Krueger (2016), Angrist, Caldwell, and Hall (2021)). In addition, ridesharing drivers have flexible hours, which enable them to smooth income fluctuations (Hall and Krueger (2018)). Finally, capacity utilization, measured by the fraction of business hours or miles that a fare-paying passenger occupies a shared vehicle, is much greater for UberX drivers than for taxi drivers in, for example, New York City and Boston (Cramer and Krueger (2016)).[46]

The growth of ridesharing services has not been welcomed by all segments of society. Ridesharing has significantly threatened the financial viability of the taxi industry by causing significant declines in its revenue and in the value of operating medallions; for example, the recent value of a medallion in New York City has dropped 80% from its all-time high of $1.3 million. A backlash from taxi drivers has led policymakers in various countries to limit Uber operations.[47] In the United States, actions include limiting New York City (commercial) drivers' access to the Uber app and forcing Uber and Lyft to treat California drivers as employees, who are eligible for various benefits, instead of as independent contractors. Critics of ridesharing also claim that by creating more automobile travel, it is increasing congestion, emissions, and traffic accidents, while reducing the use of public transit.

Despite the growing popularity of ridesharing, there has been little empirical assessment of its benefits to travelers. Conceptually, the benefits reflect the value that travelers place on the ridesharing alternative compared with the value that they place on their next best transportation alternative, such as driving their personal vehicle or taking public

---

[46] UberX, the most frequently used Uber service, provides comfortable sedans for up to four people. See https://www.uber.com

[47] Al Goodman, Elwyn Lopez, and Laura Smith-Spark, "Spanish Judge Imposes Temporary Ban on Uber Taxi Service," *CNN*, December 9, 2014, report on efforts to limit Uber operations in Spain, and Hope King, "Uber Suspends UberX Service in South Korea," *CNN Business*, March 6, 2015, reports on efforts to limit Uber operations in South Korea.

transit. Cohen et al. (2016) estimate that the benefits generated by UberX in San Francisco, Los Angeles, New York City, and Chicago amounted to nearly $3 billion in 2015; however, their estimated gains do not account for the potential benefits to travelers provided by alternative modes, which causes the estimated benefits from UberX to be biased upward because Uber is not the only mode that provides benefits to travelers. At the same time, the authors do not account for certain non-price benefits provided by UberX, which causes a downward bias in their estimates.

In this chapter, we estimate a mixed-logit model of mode choice to measure the benefits of ridesharing services provided by Uber to travelers in the San Francisco Bay Area. Ordinarily, the benefits of a mode are obtained from a discrete choice model by deleting the modal alternative from the choice set, including its alternative specific constant and attributes, and calculating the loss in consumer surplus. However, we cannot use that procedure here because our data set from the 2017 National Highway Travel Survey (NHTS) includes a car-hire alternative that combines Uber and Lyft with taxi.[48] Thus, deleting the Uber alternative from the choice set would yield a biased estimate of travelers' value of Uber because taxi's alternative specific constant would be inflated. Accordingly, we calibrate the alternative specific constant for car-hire such that it equals the value of the alternative specific constant for taxi based on the 2009 NHTS when Uber was not available. We then compare travelers' total welfare when Uber service was available with travelers' total welfare when Uber service

---

[48] As discussed below, we were able to obtain data from Uber to conduct our analysis, but we could not obtain separate data on Lyft's operations. However, Uber and Lyft offer very similar services and fares and, according to several marketing consulting firms, Uber accounts for nearly three-quarters of the ride-sharing market in the San Francisco Bay Area. We therefore use data based on Uber's operations and fares to construct attributes of ride-sharing services in Bay Area markets.

was not available to obtain a welfare difference that is not biased downward by an inflated alternative specific constant for taxi.

Based on this procedure, we find that Bay Area travelers who have access to Uber gain $0.815 per day in non-fare benefits from higher service quality and personalized pricing, which traditional taxi services have not provided, but those travelers lose $0.005 per day in slightly higher fares. The $0.81 in daily benefits amount to roughly $1 billion annually; extrapolating this estimate to all US cities with ridesharing services indicates that travelers' annual benefits are at least several billion dollars. The available evidence indicates that ridesharing does not create social costs that offset those benefits and that the social costs of any form of automobile travel could be reduced substantially with efficient policies. We conclude that regulations that limit entry and operations by TNCs reduce travelers' welfare.

2. Transportation Markets and Modes

The geographic transportation market in this study consists of nine counties in the San Francisco Bay Area: Alameda, Contra Costa, Marin, Napa, San Mateo, Santa Clara, Solano, Sonoma, and San Francisco, which are depicted in light-gray shapes along with latitudes and longitudes on the axes in figure 1. We define transportation markets as (exclusive) origin-destination pairs of grids of approximately one mile by one mile within the nine counties, which resulted in a total of 6,530 grids.[49]

Compared with traditional disaggregate analyses of urban mode choice, our analysis is complicated by the availability of disaggregated data on mode choices that includes TNCs

---

[49] We drew a rectangle covering the nine counties and then formed the grids, but we could not include the northwest part of Sonoma County. Data obtained from Uber for this study did not include information about that area, so it is omitted from the analysis.

that do not distinguish between travelers' choice of Uber or Lyft from their choice of taxi service. We therefore employ the following procedure to determine the benefits of Uber.[50]

We use the 2017 National Household Travel Survey (NHTS), a national household survey conducted by the Federal Highway Administration, to estimate San Francisco travelers' mode choice in the 2016-17 environments where Uber is an available option. The NHTS contains information on daily trips made by household members, including time of day, the location of the origin and destination based on actual coordinates, trip purpose, mode of transportation, and vehicle occupancy. The five alternative modes that we include are walking, bicycling, driving, public transit (bus and rail), and car hire services (Uber, Lyft, and taxi). The NHTS does not contain data on the prices that individual respondents pay for the modes that they use and the service times of their trips. Thus, we constructed average values of those variables using data from other sources.

The trip data are linked with socio-demographic characteristics of both the household and individual household members who make the trips. Travelers can face different choice sets if they cannot access a mode (for example, respondents that do not use smartphones cannot use Uber or Lyft, so the car hire services for those respondents include only taxis), and if a mode does not serve an origin-destination pair. We form an origin-destination pair for each trip based on its latitude and longitude provided in the NHTS and we match the trip with the two most appropriate grids that we constructed for the nine Bay Area counties. San Francisco County contains a large share of trips, both work and leisure, in our sample. We obtain additional data to construct attributes for all the modes that provide trips over the origin-destination pairs in our sample.

---

[50] As noted, we obtained data for Uber's attributes but not for Lyft's attributes.

Because we cannot distinguish travelers' choices of car hire services between Uber and taxi, we use the 2009 NHTS to calibrate the estimated choice model to replicate the mode choice outcomes that occurred in the 2008-09 environment before the existence of Uber services. This is achieved by adjusting the alternative specific dummy variables or intercepts for the choice alternatives that we estimate using the 2017 NHTS. We expect the alternative specific constant for car hire services, which no longer includes Uber, to decrease. We then conduct counterfactual simulations that estimate travelers' benefits without Uber service and compare those benefits with the estimated benefits when Uber service was available to determine the benefits generated by Uber.[51]

The 2017 NHTS, conducted between April 2016 and April 2017, contains 3,498 households and 24,585 trips within the geographical market area of analysis, while the 2009 NHTS, conducted between March 2008 and March 2009, contains 3,460 households and 28,916 trips within the geographical market area of analysis. Table 1 presents the five modes in our analysis and their total trips and mode shares for both samples and previews some possible effects of the introduction of TNCs; namely, the share of driving decreased from 81% to 73%, while the share of car hire services increased more than ten-fold from 0.07% to 0.9%.

Although it is not possible to determine directly how many of the 221 trips are by a TNC, we use a travel survey, which we discuss later, conducted by the San Francisco Municipal Transportation Agency in the Bay Area to infer that taxi has not increased its number of trips, 20, during the period, thereby indicating that approximately 200 trips were taken on a TNC. Mode shares in the San Francisco Municipal Transportation Agency study

---

[51] This procedure assumes that the decline in car-service's alternative specific dummy is attributable to the absence of Uber. We report later that the growth of ridesharing accounts for the entire increase in car hire trips from the 2009 to the 2017 NHTS.

are like the ones in the 2017 NHTS, suggesting that the NHTS does not underrepresent ridesharing.

The modes, including Uber and taxi service, differ in their spatial coverage. Figure 2 uses the 2017 NHTS data to show, as expected, that travelers drive their own vehicle throughout the nine Bay Area counties. In contrast, travelers use public transit and car hire services primarily in San Francisco County. We obtained data on Uber trips in August 2016 from Uber engineers, and for comparative purposes we show in figure 3 that Uber's service area is more extensive than the service area indicated by car hire services.[52]

3. Modal Attributes

Travelers obtain utility from the existing modes' price and non-price attributes and from the price and non-price attributes of a new mode, such as ridesharing. The data sources we use to measure the attributes of for-hire services (Uber and taxi), public transit, private car, bicycling, and walking are as follows. As is standard in urban mode choice studies, unmeasured modal attributes are captured by the alternative specific dummy variables.

<u>Uber</u>

Uber engineers provided us with data for two Uber services: UberX, a private car, and UberPool, a carpool service where riders share the ride with strangers and split the cost.[53] Data were collected for the morning and evening of each day of the week for the average trip duration (excluding waiting time), distance, and average fare of trips taken during August 2016 in the San Francisco Bay Area. The average fares did not include tips, but Chandar et

---

[52] Uber's data indicate that Uber serves about 15% more markets than are served by car hire services in the 2019 NHTS data.

[53] We are grateful to Jonathan Hall, Andrew Salzberg, and Santosh Rao Danda for providing the data and preparing it for our purposes.

al. (2019) found that in 2017 only 1% of travelers always tip, 60% of travelers never tip, and the average tip came out to 50 cents per ride.[54] Because tips represent a voluntary transfer from travelers to drivers, accounting for them would change only the distribution of benefits from Uber but not the total benefits.

The origins and destinations of the trips conformed to the grids that we constructed for our analysis. Note that we were unable to measure the attributes for an Uber trip in an origin–destination market if an Uber trip was never made in that market. Thus, our analysis did not enable Uber to contribute any utility to a traveler's choice set in such a market. As shown in table 2, UberX operates in more markets than UberPool does. In addition, trips on UberX tend to cover a longer distance, cost more, and take less time than trips on UberPool.

<div align="center">Modal Attributes from Google Maps API</div>

The 2017 NHTS provides information on the distance of each trip, which we use to calculate the cost of certain modes, and the duration of the trip for the chosen mode, but it does not provide information on the duration of the trips for the non-chosen modes and on transit fares for both the chosen and non-chosen modes. Given the departure times of the trips in the NHTS and the centroids of the latitudes and longitudes of their origins and destinations, we fill in those gaps by using Google Maps API to identify the available transportation alternatives and their attributes.

Data were collected for 2017 for the distance and travel duration of trips by private auto, bicycling, and walking. We collected the same variables and fares for the public transit

---

[54] The authors also found that the default tipping suggestions that Uber build into its platform led to only a 2.5% increase in tips.

options, bus, subway (BART), and the combination of the two.[55] We assume that people who walk or ride a bicycle to their destinations do not incur out-of-pocket costs.

Travel Decision Survey

The 2017 Travel Decision Survey (TDS) that is conducted by the San Francisco Municipal Transportation Agency summarizes travel behavior in the Bay Area and provides data that we use to help distinguish taxi from TNC operations. According to the TDS, taxis account for roughly 9 percent of all trips using car-hire services based on a modest sample of trips between ZIP codes and based on a larger sample of trips in the Bay Area.[56] That share is consistent with the increase in the car hire alternative from the 2009 to the 2017 NHTS that came solely from the growth of ridesharing. The data's spatial patterns also are consistent with patterns that we discussed previously, which indicate that TNCs provide more extensive transportation coverage than taxis do.

Constructing Monetary Costs for Private and For-Hire Cars

The monetary cost per mile of driving a private car in 2017 is from the American Automobile Association.[57] We assume vehicles accumulate 15,000 miles per year and we match the vehicles that a traveler uses to the operating and ownership costs for the vehicle's classification. We do not include parking costs, which fluctuate greatly in the Bay Area depending on the time of day, area, and duration.

Within for-hire services, we distinguish between the monetary cost of TNCs and taxi. Because we have two different Uber services, we use the weighted average fare of UberX and

---

[55] The Google Maps API transit fare is the aggregated fare of using bus and/or subway. Our measure of the duration of transit trips does not include the transfer time between buses and subways because it is generally short. Similarly, wait time at the origin tends to be short during peak periods, although it can be long during off-peak periods, such as late in the evening. However, only a small fraction of the trips in our sample occurred during that time.

[56] See https://www.sfcta.org/projects/tncs-today.

[57] American Automobile Association, *Your Driving Costs: How Much Are You Paying To Drive?*, 2017 edition.

UberPOOL, where the weights are 80 percent and 20 percent, respectively.[58] Note that the average fares are actually paid by Uber passengers, so they include any increase caused by surge-pricing adopted by Uber. The taxi fare is calculated based on the common pricing policy adopted by the major taxicab firms in San Francisco as reported by the San Francisco Municipal Transportation Agency.[59] Regulated taxi fares are, in general, based on distance and location, and regulatory authorities have not adjusted them to respond to competition from TNCs. We use the weighted fare of Uber service and the fare of taxi service as the monetary cost of car hire services for markets that are served by both taxi and Uber; otherwise, we use the taxi or Uber fare, as appropriate, when only one of those transportation options serves a market.

<u>Summary of Modal Attributes</u>

We summarize the mean and standard deviation of the fare and travel time duration of each mode in table 3. Taxicab and Uber trips are assumed to have the same duration on a given origin–destination pair, which is plausible, although they do have different fares. We also show the average distance of trips that travelers take on each mode. As expected, driving is far less costly on a per-mile basis than are the car hire services and faster on a per-mile basis than all the other modes, which greatly contributes to its dominant share of urban travel.

---

[58] The percentage of Uber drivers working for UberX was nearly 75% in 2018; no official statistic for UberPOOL was reported. See https://www.ridester.com/2018-survey.

[59] See https://www.sfmta.com/getting-around/taxi/taxi-fares. The fares for a trip of a given distance are based on charging travelers $3.50 for the first 1/5 mile and $0.55 for each additional 1/5th mile. We also add $0.55 for each minute of wait time and traffic delay. Out-of-town trips from San Francisco that exceed 15 miles are charged a fare that is 150% of the metered rate. The same fare calculation is applied for trips originating outside of San Francisco because taxis operating in the San Francisco Bay Area generally have the same pricing policy.

4. Empirical Methodology

We quantify the welfare gain to travelers from the introduction of Uber by first using the 2017 NHTS data and the data on modal attributes to estimate travelers' mode choices when ridesharing services were available in San Francisco Bay Area markets. This is the base case scenario subject to the car hire option including Uber and taxi. We then construct a counterfactual scenario where ridesharing services do not exist by using the 2009 NHTS data to calibrate the estimated choice model in the base case. We then calculate consumer benefits in the base and counterfactual scenarios and compare the difference to obtain the welfare gain to travelers from Uber services.[60]

<div align="center">Panel Data Mixed Logit Mode Choice Model</div>

Assume households indexed by $h = 1, 2, \ldots, H$ are composed of individuals $i = 1, 2, \ldots, I_h$. The transportation mode for a trip is denoted by $j \in \Omega$, where Ω is the choice set in a market:

$$\Omega = \left\{ \underbrace{\{\text{walking, bicycling}\}}_{\text{non-vehicle nest}}, \text{public transportation, driving, car hire services} \right\}$$

Define $I(A)$ as an indicator function that equals 1 if A holds and 0 otherwise. We specify the indirect utility function of individual i choosing mode j for trip t as:

$$
\begin{aligned}
u_{ijt} &= v_{ijt} + \varepsilon_{ijt} \\
&= \alpha_i p_{jt} + \beta_i d_{jt} + \eta_i I(j \in \text{non-vehicle nest}) \\
&\quad + \sum_k \psi_{ik} I(k \in \Omega \text{ and } k \neq \text{driving}) + \varepsilon_{ijt} \\
\psi_{ik} &= \mathbf{Z}_i \boldsymbol{\lambda}_k \\
\alpha_i &\sim N(\mathbf{Z}_i \boldsymbol{\theta}, \sigma_\alpha^2) \\
\beta_i &\sim N(\mathbf{Z}_i \boldsymbol{\delta}, \sigma_\beta^2) \\
\eta_i &\sim N(0, \sigma_\eta^2)
\end{aligned}
\tag{1}
$$

---

[60] It would be more difficult to execute our analysis if we used the 2009 NHTS to construct the base case because we would have to obtain 2009 data for all the modal attributes, including Uber fares.

where p_{jt} is the fare for alternative j for trip t, d_{jt} is the travel duration, and $\mathbf{Z}_i$ is a vector of ones and sociodemographic variables of the individual and household, such as age, education, household income, and the like. Preference heterogeneity for the fare and travel duration is captured by a vector of random parameters, (α_i, β_i); $(\boldsymbol{\theta}, \boldsymbol{\delta})$ and $(\sigma_\alpha^2, \sigma_\beta^2)$ are the associated vectors of means and standard deviations, respectively, to be estimated; and the random coefficient η_i captures the correlation between the two non-vehicle alternatives.

This mixed-logit specification mimics a nested logit specification where a traveler first chooses between non-vehicle and vehicle travel and then an alternative from the chosen non-vehicle or vehicle nest. We choose driving as the base alternative and ψ_{ik} represents the alternative-specific effects that vary across individuals with different socio-demographic characteristics. Finally, ε_{ijt} is assumed to be distributed *iid* extreme value for trips, individuals, and alternatives.

Let $y_{ijt} \in \Omega$ denote individual i's mode choice for trip t and let T_i denote the number of trips made by individual i; thus, the joint mode choice probability conditional on the random parameters, $\Gamma_i \equiv (\alpha_i, \beta_i, \eta_i)$, is:

$$\Pr(y_{ij1}, \dots, y_{ijT_i} \mid \Gamma_i) = \prod_{t=1}^{T_i} \frac{\exp(v_{ijt})}{\sum_{j' \in \Omega} \exp(v_{ij't})} \tag{2}$$

The unconditional joint choice probability is:

$$L_i(\Lambda, \Theta) \equiv \Pr(y_{ij1}, \dots, y_{ijT_i}) = \int_{\Gamma_i} \Pr(y_{ij1}, \dots, y_{ijT_i} \mid \Gamma_i) \rho(\Gamma_i; \Theta) \, d\Gamma_i \tag{3}$$

where $\Lambda = \{\lambda_k\}_k$ denotes a vector of non-random parameters; $\Theta = (\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\eta^2)$ denotes the parameters of the random components; and $\rho(\cdot)$ indicates the distribution of the coefficients in the population.

Estimates of the parameters (Λ, Θ) for the mixed logit probabilities of the panel are obtained by maximizing the simulated log likelihood (Train (2009)). Note that the random parameters account for the error correlation that arises for travelers who take multiple trips.

Calibrated Model

As noted, the 2017 NHTS combined taxi and Uber in the car hire alternative; thus, we cannot conduct a counterfactual that simply eliminates Uber from the travelers' choice set because the alternative specific constant for the taxi alternative that remains is inflated given that its value was determined when Uber services were available. We use the 2009 NHTS data to calibrate a model for the counterfactual scenario in which Uber services do not exist. To do so, we first replace Uber fares, which are the fares of car hire services, with taxi fares in the 2017 NHTS data. Given those fares, we calibrate the intercepts in $\psi_{ik} = \mathbf{Z}_i \boldsymbol{\lambda}_k$, which are specified in equation (1), to replicate the mode shares in the 2009 NHTS data.[61]

Travelers' Welfare

Travelers' welfare is calculated by the log-sum rule in Choi and Moon (1997):

$$\text{CS} = \sum_{i=1}^{I} \int_{\Gamma_i} \frac{1}{\tau_i} \sum_{t=1}^{T_i} \ln \sum_{j \in \Omega} \exp\left(v_{ijt}(\Gamma_i) \right) \rho(\Gamma_i; \Theta) \, d\Gamma_i \tag{4}$$

where τ_i is the individual's marginal utility of income derived from Roy's identity.[62] We decompose travelers' benefits from Uber services into non-fare and fare benefits. Non-fare benefits come from sources such as higher service quality, more transparent fares, personalized pricing and services, competitive responses to Uber such as expanded taxi

---

[61] The share of taxicab (or car hire services) was only 0.09 in the 2009 NHTS data.
[62] Because we have several fare variables, the marginal utility of income is computed using the coefficients of those variables, not just the coefficient for fare when it enters by itself.

service into new markets, and the like.[63] We quantify those benefits by using the estimated and calibrated model to calculate travelers' welfare using equation (4) on the adjusted data, in which fares of car hire services in all markets are assumed to be cab fares. We measure non-fare benefits as the difference in welfare from the two models on the adjusted data, where the change in non-fare benefits is captured by the change in the alternative specific dummy variable for car hire services. We quantify the fare benefits by redoing the calculations on the original 2017 NHTS data in which the fares of car hire services in markets served by Uber are based only on Uber fares.

5. Empirical Findings

Travelers' utility from their choice of mode for a given trip is influenced by the modal attributes, fares, and trip duration. Those attributes were specified alone, interacted with the purpose of the trip, work, or shopping, and interacted with socioeconomic variables, household size, income, professional, managerial, or technical job occupation, young child in the household, and frequent smartphone user.

We include random parameters, assumed to be log-normally distributed, for fares and trip duration and for the non-vehicle dummy variable. Bicycling is the base mode in the non-vehicle nest and driving is the base alternative of the mode choice model. Finally, we control for the possible endogeneity of fares by including alternative specific dummy variables, which capture omitted modal attributes, including attributes that may be correlated with fares.[64]

---

[63] The benefit of expanded taxi service is reflected in the alternative specific dummy for car hire services, which includes both Uber and taxi.

[64] Using alternative specific dummy variables for this purpose has been common practice in disaggregate choice modeling since the 1970s.

We present the mixed logit parameter estimates in table 4. The coefficients of the mean random parameters have the expected negative sign and along with the standard deviations of the random parameters are statistically significant. All else constant, walking is preferred to bicycling, and non-vehicle modes are less preferred to driving. Transit and car hire services also are less preferred to driving, although the alternative-specific dummy for transit is not statistically significant.

The average value of time, VoT, based on the fare and duration mean coefficients is $42.42 per hour, which is broadly consistent with other high VoT estimates for travelers in urban areas with high household incomes (Small, Winston, and Yan (2005)). Note that our sample also has a large share of people who walk to work. As we report below, people with short trips have a very high VoT.

The estimates of the interaction terms indicate that the disutility of a higher fare is increased for shopping trips because it reduces the availability of money that could be spent on goods and services, and it also is increased for larger households because such households have a tighter budget constraint per household member. Similarly, the disutility from a trip with a longer duration that increases time costs is greater for shopping trips and larger households.

<div align="center">Sensitivity by Trip Distance</div>

Travelers' heterogeneity is reflected in sorting across different residential locations. All else constant, travelers with the highest value of travel time live closest to their workplaces, while travelers with the lowest value of travel time live farthest from their workplaces (Calfee and Winston (1998), Winston and Shirley (1998)). The non-motor-vehicle modes, bicycling and walking, are unlikely to be chosen by travelers for long-distance work

or shopping trips. Descriptive information in the NHTS indicates that the shares of bicycling and walking declined dramatically for trips greater than or equal to two miles, so we re-estimated the mode choice model with a sample that included only trips with a distance that was less than two miles.

The estimation results shown in table 5 indicate that travelers' relative valuation of fares is notably different for shorter trips compared with their valuation of those attributes for the full sample of trips. The effect of fare is much smaller, while the effect of duration does not change much. The estimated value of time, $151 per hour, is much higher than the value of time for the entire sample, which likely reflects the fact that the most affluent people tend to live closest to their workplaces because they value travel time so highly.[65] In addition, public transit is preferred to driving for short trips, all else constant, which may reflect the unavailability of free parking.

<center>Estimated Benefits</center>

We measure the benefits that Uber provides to consumers by first removing Uber as an alternative in 2017 Bay Area markets. To do so, we replace Uber fares, which are included in the fares of car hire services, with taxi fares, and we calibrate the constant for car hire services so that its market share drops from 0.9% to 0.07%.[66] We then calculate the total benefits that travelers obtain from the transportation choice set in the counterfactual scenario where Uber is not an available option and we compare it with their total benefits from the transportation choice set in the base case scenario where Uber is available.

---

[65] The high hourly value of time is plausible for people who live in the Bay Area close to their workplace and make say $300,000 a year and value travel time greater than the conventional value of 50% of the hourly wage.
[66] As expected, the constant for car-hire services, which no longer includes Uber, decreased from –9.08, as reported in table 4, to –23.

We find that Bay Area travelers in the NHTS sample, not just Uber passengers, gain $4,446 per day, which, given 5,471 travelers, amounts to an average gain per traveler of $0.81. Making the conservative assumption that only half of the Bay Area population has access to Uber,[67] we estimate that the daily benefits to the Bay Area are roughly $2.8 million ($0.81 × 3,484,823[68]) for an annual gain of $1.02 billion. Extrapolating the gain to all US cities suggests that an estimate of the annual benefits from Uber services would amount to several billions of dollars.

We offer four useful checks on the plausibility of the magnitude of the estimated benefits. First, Cohen et al.'s (2016) estimates of the average daily gain per traveler generated by Uber of $1.60 is roughly twice our estimate; however, as noted, their estimate is inflated because they do not account for the benefits provided by alternative transportation modes. Second, Gorback's (2021) estimates of the welfare gains from Uber are somewhat higher than our estimates, but they also account for the benefits from improving residents' accessibility and Core Based Statistical Areas' (CBSAs') amenities. Third, our estimate of the average daily gain per traveler is a modest share, roughly 5%, of Bay Area travelers' average daily expenditures on transportation. Given that our estimate is conservative, as noted previously and discussed further below, and given the high cost of driving in the Bay Area, this relationship is plausible.

Finally, the estimated benefits can be decomposed into non-fare benefits of $0.815 and very slight fare losses of -$0.005. Travelers incur a loss from fares because although Uber fares are roughly $2 less than taxi fares for the car hire trips that were actually taken, Uber fares are higher than taxi fares for the large remainder of trips that were taken on other modes.

---

[67] The assumption is conservative because Uber tends to serve the most densely populated areas in an urban metropolis.
[68] The figure for the Bay Area population is from the State of California, Department of Finance.

Because there is a positive probability, even if small, that Uber could have been chosen for those trips, its relatively higher fares reduce travelers' welfare.

The non-fare benefits reflect Uber's departure times and differentiated services, which, as captured by the alternative specific constant, are more closely aligned with travelers' preferences than are taxi's offerings. Uber also benefits travelers by causing taxi to increase its geographic coverage to compete more effectively with Uber.[69]

Because our analysis holds the size of the travel market constant, we underestimate the gains provided by Uber given that Uber's entry has expanded the San Francisco Bay Area travel market by attracting additional travelers who benefit from the service. In addition, we do not account for trips by travelers outside of the Bay Area, which include new trips generated by Uber to, for example, San Francisco and possibly San Jose airports.[70] Finally, we do not account for any changes in spatial economic activity attributable to Uber. As noted, Gorback (2021) finds that UberX's entry has caused house prices for the 34 largest CBSA's in the United States to increase 4% by improving residents' accessibility and the area's amenities. Similarly, if Uber has improved travelers' accessibility to Bay Area amenities, then its entry has provided additional social benefits.

Lui, Brynjolfsson, and Dowlatabadi (2021) point out that we may be understating the benefits from Uber because taxis have been able to overcharge passengers by taking a longer route to their destinations when passengers were unfamiliar with the area or were disinterested in the cost because their employer was paying for the trip. The authors find that taxi drivers

---

[69] The geographic coverage of car hire service in the 2017 NHTS data is greater than it is in the 2009 NHTS data, suggesting the competition from TNCs has caused car hire services to expand their coverage.

[70] We possibly overstate the gains because people must choose their next best alternative mode to Uber when they may get higher utility by choosing not to travel if Uber is not a viable option. But this upward bias is undoubtedly much smaller than the downward bias from holding the size of the travel market constant.

take longer routes that are more time consuming than Uber drivers take for matched airport routes by an average of 8%, with nonlocal passengers experiencing even longer routes and travel times. Thus, we may be understating the fare benefits from Uber by setting taxi fares based on standard taxi fare policies, as summarized previously, instead of using the taxi charges that travelers actually pay. Uber and other ride-sharing companies have addressed this information problem with a transparent pricing system, which requires a receipt for the passenger with the name of the driver and a map of the actual journey.

Other Effects of Uber

The available evidence suggests that a full social welfare analysis of Uber, which would go beyond its effect on travelers and account for its effect on automobile-related externalities, including safety, congestion, and emissions, and on other modes and the automobile industry, would not overturn our finding that its effect on society is positive.

Uber is likely to improve safety by providing trips during the evening to travelers who have been drinking or who have been working late and are too tired to drive. Empirical research of the effect of ridesharing on traffic fatalities has yielded inconsistent conclusions. Recently, Anderson and Lucas (2021) use proprietary data from Uber to measure rideshare activity more accurately than previous work and find that ridesharing has a robust negative impact on traffic fatalities that concentrate during nights and weekends and has decreased US alcohol-related traffic fatalities by 6.1% and reduced total US traffic fatalities by 4.0%.[71]

Uber's effect on congestion is controversial because although it reduces congestion caused by taxis that cruise for passengers in dense urban areas, it also could cause congestion if it generated more trips during peak travel periods. Li et al. (2022) find that the entry of

---

[71] Based on conventional estimates of the value of statistical life, the annual life-saving benefits range from $2.3 billion to $5.4 billion.

Uber has mixed effects on traffic congestion, decreasing it in sprawling urban areas and increasing it in compact urban areas.

In any case, the efficient policy is not to limit Uber operations, but for policymakers to set an efficient congestion toll for all motor vehicles that could simultaneously address multiple automobile-related externalities (Langer, Maheshri, and Winston (2017), Winston and Yan (2021)). Leard and Xing (2020) conclude that the availability of ridesharing has led to modest increases in total vehicle miles traveled and greenhouse gas emissions; but efficiency calls for setting emission charges on all motor vehicles to reduce the social costs of pollution.[72]

Ridesharing also has been criticized for plunging the taxi industry into a financial crisis and for displacing a significant portion of public transit trips in some large cities (Leard and Xing (2020)).[73] Taxi's financial problems reflect outdated regulations that limit its operations and its inability to cater effectively to travelers' preferences. Indeed, taxis have yet to develop an effective platform using a smartphone application in most major US cities. Starting in spring 2022, riders will be able to open the Uber app and choose a taxi. Uber will then refer the request to the two taxi technology companies, Curb and CMT, which will notify drivers to pick up the passengers. The fare will be based on Uber's pricing and policies, including surge pricing, which can significantly increase the cost at peak times.

As we discussed in the previous chapter, public bus and rail transit are highly inefficient and requires large taxpayer funded subsidies that are likely to exceed their benefits

---

[72] Ward, Michalek, and Samaras (2021) simulate replacing private vehicle travel with TNCs in six US cities and find mixed effects of TNCs on emissions.

[73] Leard and Xing (2020) show in their figure 2 the percentage of each mode, including transit and taxis, that travelers would use if ridesharing were not available. Meredith-Karam et al. (2021) find that transit and ridesharing barely complement each other, while Hall, Palsson, and Price (2018) find, on average, that Uber complements transit.

to users (Winston (2013)). Thus, ridesharing is increasing not decreasing social welfare when it captures some of transit's mode share because transit's ridership consists of highly subsidized passengers.

Uber is providing a benefit to transit by licensing its ride-hailing software to public transit agencies. For example, Denver's Regional Transportation District is using Uber's management software to manage its fleet of wheelchair-accessible vehicles. The Transportation Authority of Marin County is using Uber's software to facilitate requesting, matching, and tracking its vehicle fleet.

The introduction of Uber and ridesharing in general has provided benefits to the automobile industry and dealers because it has induced people to purchase cars so they can provide ridesharing services. Ward et al. (2020) estimate that ridesharing has resulted in at least 100,000 additional vehicles on the road, although this figure may be partially offset by people who decide to eschew car ownership or to own fewer cars because of the availability of ridesharing.

Ridesharing's financial condition also is a matter of controversy and uncertainty. Uber has yet to be consistently profitable, but its market valuation of $50 billion before COVID-19 suggests that investors believe that Uber will be highly profitable in the long run. Accordingly, ridesharing's producer surplus could offset its effect on the financial condition of the taxi industry and public transit.

Finally, ridesharing companies' relationship with labor has created uncertainty about their long-run profitability. The companies have resisted classifying drivers as employees, who can legally obtain certain benefits and protections. In 2018, New York City became the first US city to require Uber and Lyft to pay their drivers a minimum wage. Koustas, Parrott,

and Reich (2020) found that the policy led to a modest increase in drivers' earnings and passengers' fares and to a modest decline in Uber's and Lyft's commission rates. We await other actions by policymakers and additional research before reaching a conclusion about the long-run distributional effects of labor market policies on ridesharing companies' profits, labor's earnings, and travelers' fares.[74]

6. Conclusion

Utility-maximizing travelers have improved their welfare by shifting to a new mode, ridesharing services, which has grown rapidly by competing effectively against taxi and transit alternatives. We quantified the benefits from the leading rideshare company, Uber, in San Francisco Bay Area markets, accounting for travelers' complete set of transportation options. We find that travelers have gained roughly $1 billion annually from Uber's service, and we argue that a full welfare analysis is more likely to strengthen than to weaken the conclusion that Uber has provided positive social benefits.

Nonetheless, ridesharing faces opposition from special interests that is aided and abetted by some policymakers in the United States and in some other countries, which have introduced regulatory policies to limit ridesharing operations. Our findings suggest that such actions harm the public interest and that they should be resisted by the public because they

---

[74]Angrist, Caldwell, and Hall (2021) find that, compared with taxi, rideshare drivers' earnings improve considerably because they do not have to make a fixed payment for their weekly or daily medallion lease. Taxi medallions limit the supply of transportation service because drivers must own or lease a medallion granting them the right to drive. Cook et al. (2021) find that female Uber drivers earn 7% less than male drivers, but the difference is not due to discrimination. Instead, the difference can be explained by men driving faster, choosing to drive for more costly trips, and being more experienced, such as, knowing when to strategically cancel and accept trips. Uber's pay gap has no implication for "social justice" because drivers are treated equally.

unnecessarily interfere with their informed, self-interested transportation choices in a highly competitive environment.

After decades of inefficiency and technological stagnation in urban transportation, ridesharing is a welcome private sector innovation that many travelers prefer to government regulated taxis and public transit. In the coming decades, ridesharing's competitive inroads are likely to expand. As we discuss in chapter 6, ridesharing is likely to operate autonomously in the long run and further benefit travelers by replacing a large share of privately owned automobiles. Innovative efforts by entrepreneurs should not be discouraged by policymakers because they may produce substantial social benefits by displacing less efficient modes of transportation.

Figure 1. San Francisco Bay Area Counties (Light gray)
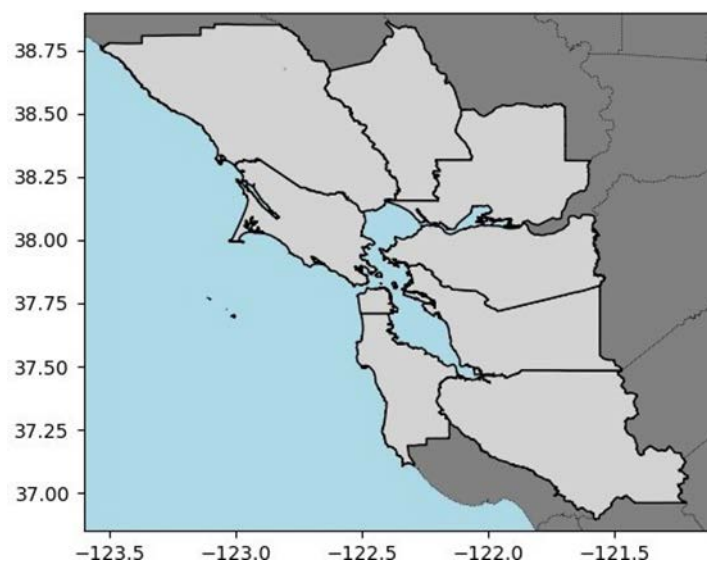
Figure 2. Spatial Distribution of Trips by Mode in the 2017 NHTS

(a) Car Hire Origin (b) Car Hire Destination

(c) Driving Origin (d) Driving Destination

(e) Public Transit Origin (f) Public Transit Destination
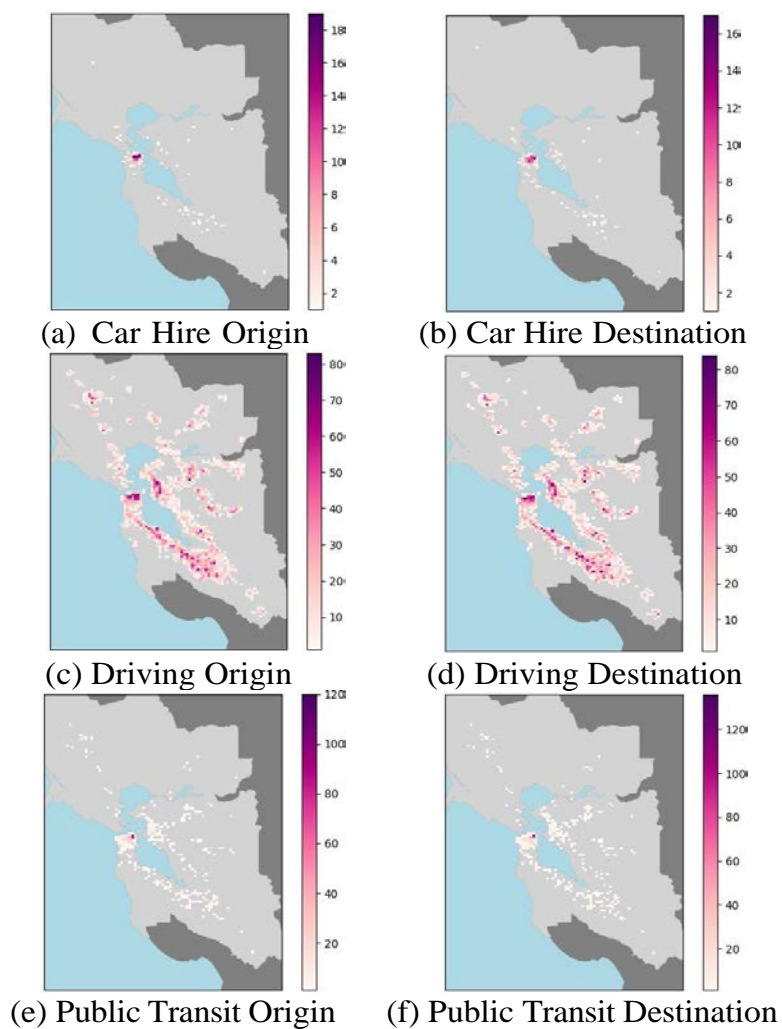
Figure 3. Spatial Distribution of Uber Trips

(a) Origin Grids (b) Destination Grids
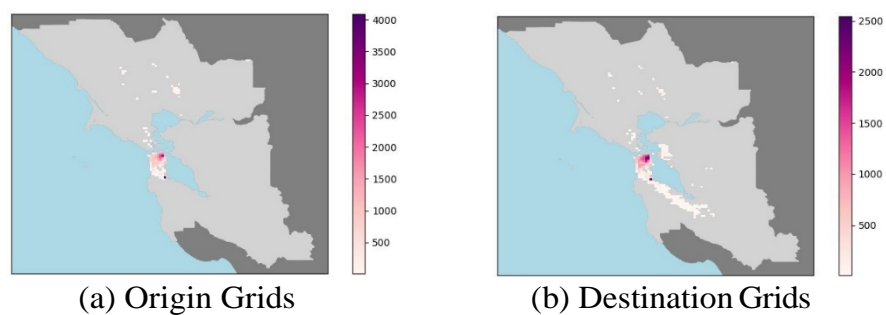
Table 1. Mode Shares in the 2009 and 2017 NHTS

Mode	Number of trips in 2009	Share (%) in 2009	Number of trips in 2017	Share (%) in 2017
Driving	23,477	81.19	18,055	73.43
Car hire services	20	0.07	221	0.9
Public transport	591	2.04	1,071	4.35
Bicycling	431	1.49	480	1.95
Walking	3,932	13.6	4,448	18.09
Other	465	1.6	310	1.26
Total	28,916	100	24,585	100

Notes: Uber accounts for the vast majority of TNC trips in the Bay Area so we use its attributes for TNC service and Fares.

Table 2. Summary Statistics of Uber by Type of Service

Mode	Attribute	Obs.	Mean	S.D.	Min	Max
UberX	Fare ($)	14,342	16.27	10.64	4.75	110.82
	Duration (minutes)	14,342	17.35	10.08	1.67	92.33
	Distance (miles)	14,342	5.33	5.19	0.03	46.81
UberPOOL	Fare ($)	12,868	10.89	5.74	4.27	68.20
	Duration (minutes)	12,868	20.38	10.22	2.00	75.50
	Distance (miles)	12,868	4.99	3.64	0.00	38.18

Table 3. Means (Standard Deviations) of Modal Attributes

Attribute	Car hire services		Driving	Public transport	Bicycling	Walking
	Uber	Taxi				
Fare ($)	24.24	25.32	4.12	4.50	0	0
	(27.14)	(27.15)	(5.98)	(3.02)		
Duration (minutes)	16.31		13.24	50.04	14.08	8.86
	(10.17)		(11.68)	(33.96)	(13.26)	(29.96)
Distance (miles)	6.18		7.18	12.88	2.35	0.43
	(7.96)		(10.35)	(13.46)	(2.35)	(1.49)

Table 4. Mode Choice Parameter Estimates

Variable	Coefficient estimate (standard error)
Fare	–0.526 (0.071)
Fare × work trips	–0.005 (0.013)
Fare × shopping trips	–0.109 (0.024)
Fare × household size	–0.055 (0.021)
Fare × high income	–0.104 (0.072)
Fare × household size × high income	0.034 (0.022)
Fare × frequent smartphone	–0.031 (0.052)
S.D. of fare coefficient	0.370 (0.017)
Duration	–0.372 (0.014)
Duration × professional	0.042 (0.006)
Duration × work trip	0.004 (0.004)
Duration × shopping trip	–0.070 (0.007)
Duration × household size	–0.015 (0.002)
Duration × young child	–0.011 (0.012)
S.D. of duration coefficient	0.178 (0.007)
Non-vehicle	–3.449 (0.897)
S.D. of non-vehicle coefficient	2.635 (0.115)
Walking	7.227 (0.742)
Public transport	–0.317 (0.723)
Car hire services	–9.081 (2.430)
Interactions between constants and sociodemographic variables	Yes
Number of observations	87,386

Table 5. Mode Choice Parameter Estimates for Distances Less Than Two Miles

Variable	Coefficient estimate (standard error)
Fare	–0.143 (0.089)
Fare × work trips	–0.015 (0.015)
Fare × shopping trips	–0.083 (0.029)
Fare × household size	–0.078 (0.024)
Fare × high income	–0.121 (0.076)
Fare × household size × high income	–0.195 (0.082)
Fare × frequent smartphone	0.078 (0.027)
S.D. of fare coefficient	0.224 (0.019)
Duration	–0.360 (0.030)
Duration × professional	0.055 (0.013)
Duration × work trip	0.011 (0.006)
Duration × shopping trip	–0.058 (0.011)
Duration × household size	–0.003 (0.004)
Duration × young child	0.007 (0.022)
S.D. of duration coefficient	0.163 (0.013)
Non-vehicle	–11.362 (7.250)
S.D. of non-vehicle coefficient	21.491 (3.363)
Walking	14.404 (6.813)
Public transport	5.120 (1.454)
Car hire services	–22.800 (7.383)
Interactions between constants and sociodemographic variables	Yes
Number of observations	39,647

Chapter 5. Consumer Adoption of Electric Vehicles:
The Appropriate Role of Government and the Private Sector

Clifford Winston and Jia Yan

1. Introduction

The combustion of petroleum-based products, like gasoline, in internal combustion engines (ICE) is a major source of air pollution, which damages lung tissue and can aggravate and lead to repository diseases, such as asthma. In addition, ICEs account for some 10% of the world's greenhouse gas emissions, which trap heat, thereby causing climate change, and account for nearly 25% of greenhouse gas emissions generated by the United States.[75]

The private sector's innovation of electric vehicles (EVs) powered by an electric engine instead of an ICE can improve human health and the environment by significantly reducing the gasoline emissions' externality. At long last, the innovation is gaining measurable public adoption, as indicated by EVs' share of new car sales in the United States increasing from roughly 2% in 2020, 4% in 2021, and 5% in 2022. In addition, EVs have become a priority for automakers, as indicated by the $0.5 trillion the industry is on track to invest during the next five years to accelerate the transition to EVs (Ewing and Boudette (2022)).

Because there are strong elements in society that are rightfully concerned about the damage that cars inflict on the environment and because those elements generally do not trust the private sector to take steps on their own to improve the environment, federal policymakers have intervened to spur the public's adoption of EVs without drawing much criticism. It is reasonable for the government to take actions, such as implementing efficient emissions taxes

---

[75] According to the Environmental Protection Agency, transportation is the largest single source of greenhouse gases generated by the United States, representing 29% of the nation's total emissions. Within transportation, light-duty vehicles account for 58% of the greenhouse gas emissions and medium- and heavy-duty trucks account for 24% of the greenhouse gas emissions.

or command and control policies, to curb a significant externality like vehicle emissions. However, federal government policies toward EVs have eschewed those policies and included subsidizing the price for EVs paid by consumers and allocating funds to construct EV charging stations. The California Air Resources Board has gone a step further and banned the sales of new ICE vehicles by 2035, with intermediate requirements that 35% of new passenger vehicles sold by 2026 and 68% by 2030 produce no carbon emissions.

In this chapter, we review the history of EVs' development and the sector's competitive environment, and we discuss the factors that influence whether EVs will be adopted efficiently by households.[76] We then consider the efficient mix of market forces and government policy that determines the path that households and automakers should take to make the transition from gas-powered to electric vehicles, which will enable society to benefit from an innovation that curbs harmful vehicle emissions.

We conclude that it would be more efficient if government allowed competitive automakers to take the lead in resolving EVs' remaining technological problems, investing in fast EV chargers, and determining the pace that consumers throughout the nation are encouraged to adopt EVs. For its part, government should set efficient highway user taxes for EVs, enable lithium to be mined more quickly in the United States to facilitate critical lithium-ion battery projects, and ensure that the electric power grid is sufficiently reliable to power the growing fleet of EVs. Subsidizing EV consumers and investing in EV charging stations is likely to be counterproductive.

---

[76] We point out in chapter 12 that the trucking industry will eventually replace diesel fuel to power trucks with either electric batteries or hydrogen fuel cells.

2. A Brief History of Electric Vehicles and the Current Competitive Environment

Like many innovations, a series of breakthroughs—in this case, in batteries and electric motors—led to the development of the electric vehicle in the 1800s.[77] Practical versions first appeared in France and England, while the first successful electric car, with a top speed of 14 miles per hour, made its debut in the United States around 1890. By 1900, electric cars accounted for roughly one-third of all vehicles on the road, including steam and gasoline powered vehicles, and they continued to show strong sales during the next 10 years. Urban residents found electric cars suitable for short trips around the city; poor road conditions outside cities meant few cars of any type could venture farther.

Henry Ford's mass-produced Model T emerged as a formidable competitor to the electric car. Introduced in 1908, the Model T made gasoline-powered cars widely available and affordable. By 1912, the gasoline car cost only $650, while an electric roadster sold for $1,750. That same year, Charles Kettering introduced the electric starter, eliminating the need for the hand crank and giving rise to more gasoline-powered vehicle sales.

By the 1920s, the US had a more expansive system of roads connecting cities, and Americans wanted to explore the country. With the discovery of Texas crude oil, gas became cheap and readily available for rural Americans, and gas stations began popping up across the country. In comparison, very few Americans outside of cities had electricity at that time. By 1935, electric vehicles all but disappeared from passenger transportation.

During the early 1970s, soaring oil prices and gasoline shortages—peaking with the 1973 Arab oil embargo—revived interest in electric cars to reduce US dependence on foreign oil. Congress took note and passed the Electric and Hybrid Vehicle Research, Development,

---

[77] The historical summary that follows draws heavily on Matulka (2014).

and Demonstration Act of 1976, authorizing the Energy Department to support research and development in electric and hybrid vehicles.

However, it is not clear that a market failure existed that justified congressional involvement to spur the adoption of electric vehicles. For example, during the 1970s, the Japanese automakers were able to obtain a strong toehold in the US market by offering reliable, fuel-efficient cars that gained considerable popularity among consumers. By the time US automakers responded, the Japanese automakers had developed brand loyalty among purchasers of smaller vehicles (Mannering and Winston (1991)), and they would build on that loyalty to attract consumers of all vehicle sizes. Today, Japanese automakers' share of new car sales in the United States exceeds the US automakers' share of new car sales.

Congress should have allowed the market to determine whether automakers were able to offer an electric vehicle that would be sufficiently attractive to consumers to enable automakers to make money on the vehicle in the long run. If Congress wanted to incentivize consumer demand for electric vehicles efficiently, they should have implemented efficient taxes on vehicle miles, which included a charge for vehicle emissions that would have raised the relative operating costs of gasoline-powered vehicles (Langer, Mahesrhi, and Winston (2017)).

As it turned out, automakers began to explore options for developing alternative fuel vehicles during the 1970s, including electric cars. For example, General Motors developed a prototype for an urban electric car that it displayed at the Environmental Protection Agency's First Symposium on Low Pollution Power Systems Development in 1973, and the American Motor Company produced electric delivery Jeeps that the United States Postal Service used in a 1975 test program. However, those and other electric vehicles still did not compare favorably with gasoline-powered cars in terms of performance, their top speeds usually did not exceed 45

miles per hour, and range, their typical range was limited to 40 miles before needing to be recharged. Given their low scale of production, electric vehicles also suffered from a cost disadvantage, which was masked by subsidies provided in the 1976 electric vehicle act. Subsidies to consumers and automakers have continued to be a mainstay of electric vehicles.[78]

Government involvement in EVs and EV market forces have continued to evolve, but it is far from clear that government policy has played a constructive role in promoting consumer adoption of EVs. Toyota and Honda introduced hybrid gas-and-electric vehicles during the 1990s, which increased the credibility of electric vehicles becoming a viable part of the US vehicle fleet. A major change in the EV market occurred when Tesla announced in 2006 that it would start producing a luxury electric sports car. In 2010, Tesla received at $465 million loan from the Department of Energy's Loan Programs Office—a loan that Tesla repaid a full nine years early—to establish a manufacturing facility in California. It is likely that private markets would have loaned Tesla the money if it approached potential lenders with a business proposal that could be credibly interpreted as showing promise of its eventual market performance.

In any event, Tesla's subsequent success spurred many large automakers to accelerate work on their own electric vehicles. In late 2010, the Chevy Volt and the Nissan Leaf were released in the US market. The first commercially available plug-in hybrid, the Volt has a gasoline engine that supplements its electric drive once the battery is depleted, allowing consumers to drive on electric for most trips and gasoline to extend the vehicle's range. In comparison, the Leaf is an all-electric vehicle. Other automakers began rolling out electric

---

[78] Government support of innovations that lead to significant social benefits is a justification for subsidies. However, Winston (2021b) points out that many government efforts to spur innovations amounted to funding that would have been undertaken by the private sector if it believed that it would be profitable to do so.

vehicles in the US to the point where dozens of plug-in electric and hybrid models were available from several different automakers in a variety of sizes.

3. The Current Competitive Environment and EV Performance

Although electric vehicles are slowly gaining acceptance in the United States, their market share of roughly 3% of vehicles sold in the United States ranked tenth among the ten countries with the highest penetration rates.[79] Telsa dominates the US market with roughly 75% of sales, while GM, for example, has only a 6% share. From a global perspective, in the first half of 2022, BYD, based in Shenzen, China, sold more EVs and plug-in hybrid vehicles than Tesla sold.

In the coming years, Tesla will be challenged in the US by BYD, as it plans to sell its e6 in Southern California in 2022 and to expand sales in other US markets in 2023, and by established automakers and startup companies as "market conditions allow." For example, GM is planning to go all-electric by 2035 and is investing nearly $7 billion in Michigan to build a new battery plant and to overhaul an existing factory to make electric trucks. GM also has announced a partnership with Honda to co-develop electric vehicles to reduce production costs and introduce significant price competition to the EV market with new EVs priced below $30,000. The partnership centers around using the General Motors' Ultium EV platform and both automakers' robust manufacturing capabilities. The vehicles are expected to be available in North America in 2027.

---

[79] Norway leads all countries with electric vehicles accounting for roughly 75% of all new car sales. Norway promoted EVs by exempting purchasers from the country's vehicle purchase tax and its 25% Value Added Tax. Extensive charging infrastructure began with government funds, followed by private sector investment.

Ford has already attracted considerable attention by announcing the production of the all-electric F-150 Lightening. The gas-powered Ford F-150 is the best-selling vehicle in the US, so it is not surprising that more than 200,000 people signed up for the waitlist for the electric F-150 shortly after Ford announced plans for its production. Finally, many startups have entered the EV market, including Rivian, Lucid, Fisker, and several others. It is likely that at least a few will survive the shakeout and provide notable competition.

Electric vehicles still face the fundamental challenge of being cost-competitive with gasoline-powered cars, including the sticker-price and the cost of insurance, maintenance, deprecation, battery replacement, and purchasing electricity. EVs' have advantages in lower fuel and maintenance costs, but to promote efficient adoption their full costs must come down to match the full cost of gasoline cars without consumers receiving subsidies.

It appears, however, that EVs will continue to be subsidized by the federal and state governments. Although the $7,500 federal tax credit for the purchase of Tesla and General Motors electric vehicles was eliminated as of April 2020 and federal subsidies for other EV models depend on their battery capacity and are limited by their manufacturers' cumulative EV sales, the Inflation Reduction Act of 2022 provides new opportunities for savings on buying an EV. The bill extends the current $7,500 tax credit for a new vehicle and provides a $4,000 tax credit for a used EV. It also eliminates the current cap that cuts automakers off tax credits after they have sold 200,000 EVs, and it is written so buyers can get an immediate discount at the dealership, instead of waiting weeks or months for their tax credit to come through. An important caveat is that automakers and consumers are unlikely to be able to take advantage of this tax credit over the next few years because of the time it will take automakers to transform their operations to satisfy the tax credit's requirement that EVs are built in North America, and

EV batteries do not come from countries like China.[80]   Finally, states also provide varying

subsidies; for example, California offers rebates of up to $7,000 for plug-in hybrid and fully

electric vehicles.

Batteries are another important component of the EV and gasoline car competitive

interface. Most electric vehicle batteries do not provide enough range for trips beyond urban

travel, and they take far longer to charge than refilling a conventional car's gas tank.[81]   Range

anxiety exists because there is sparse provision of chargers in lightly populated areas. For

example, the 550 miles between Reno and Salt Lake City has only a few chargers. The private

sector is investing large sums in new kinds of vehicle batteries that will reduce the cost and

increase the range of an EV. However, no one can predict how soon and how much better the

next generation of EV batteries will be and the stability of raw materials prices.[82]

Technology that enables EVs to be charged while they are stationary or moving on a

road would calm EV range anxiety, but such dynamic charging requires a costly investment to

put electric coils underneath the pavement and to connect the road system to the electric grid to

become economically viable.[83]   Alternatively, the public and private sector intend to put more

EV charging stations on roads and to increase their speed, which again raises the question of

who should bear responsibility for this investment.

---

[80] The tax credit also is limited to trucks, vans, and SUVs under $80,000, and other vehicles under $55,000, and has an income threshold for consumers.

[81] EVs can be charged at home or at a charging station. Charging may take 30 minutes or several hours depending on the size of the battery and the speed of the charging point.

[82] For example, the Mercedes Vision EQXX may get as much as 600 miles on a single charge. However, it is not clear when it will be available with that range and its cost. China's Guangzhou Automobile Group has the Aion LX Plus, which it claims can drive more than 600 miles on a single charge. And Apple is targeting 2024 to produce a passenger vehicle that could include its own breakthrough battery technology, which could significantly reduce the cost of batteries and increase an EV's range.

[83] Michigan has awarded a contract to ElectReon to build the first electric vehicle charging road in the United States by 2023. The roadway would be in the Michigan Central district and could charge electric vehicles while they are in motion or stationery.

4. An Appraisal of Government Interventions in Electric Vehicles

Government efforts to advance automobile performance have incurred considerable waste when the market could have achieved the desired result more efficiently without costing taxpayers. A textbook demonstration of the point occurred in 1993 when the Clinton administration initiated the "Partnership for a New Generation of Vehicles" with domestic automakers in hopes of producing a high-gas-mileage car using a hybrid propulsion system by the decade's end. At a cost to taxpayers of $1.5 billion, the goal was never accomplished (Winston (2006)). However, the two automakers that began offering hybrid vehicles to American consumers in the 1990s, Honda and Toyota, received none of the subsidy.

Electric Vehicles' Market Share

A similar inefficient outcome has developed in electric vehicles. Despite decades of US taxpayer–funded subsidies to jump start motorists' adoption of electric vehicles, the EV market share is tiny. President Biden intends to spur a dramatic change in EVs' market share, as indicated by setting a national goal that 50% of all new cars in the United States will be electric by 2030 and the subsidies contained in the Inflation Reduction of 2022 will help to some extent to achieve that goal.

However, President Biden has failed to consider the role of market forces, even though in recent years, the incentives for automakers to develop electric vehicles that can displace gasoline-powered vehicles without federal subsidies have grown sharply, as indicated by GM's commitment to be fully electric by 2035 and Volkswagen's plans to be fully electric by 2040. Most other automakers are committing to make a share of their vehicles electric by a certain date. For example, Ford announced that 40% of its vehicles will be electrified by 2030. Even Tesla has been operating without federal subsidies.

Accordingly, government should stop trying to rush the transition from gasoline-powered vehicles to EVs by setting arbitrary targets, which erroneously serve to justify continued subsidization of automakers' production and sale of electric vehicles. Lieberman (2021) points out that government actions have been regressive because nearly 80% of federal tax credits for EVs have gone to households with gross income of $100,000 or more after adjusting for certain taxes and deductions.

It also is not yet clear that EVs improve the environment on net, because they may generate more emissions than similar fossil fuel–powered vehicles in certain areas, largely due to the production of batteries and the electricity they require. Rapson and Muehlegger (2021) and Holland et al. (2016) find there is considerable heterogeneity by location and by motorists' vehicle miles traveled when attempting to determine whether the environmental benefits of an electric vehicle compared with a gasoline-powered vehicle are positive or negative. This ambiguity is likely to change in favor of the superiority of electric vehicles as their batteries continue to improve and as electricity generation produces less pollution (Holland et al. (2020)).[84]

EV Charging Stations

To help achieve his target, President Biden signed into law a $1 trillion infrastructure package that he initiated, which includes $7.5 billion in subsidies to fund a nationwide network of some 500,000 EV charging stations by 2030, up from the fewer than 50,000 such stations across the entire country today. The stations will be built in Alternative Fuel Corridor designations in all 50 states, as well as Washington, DC, and Puerto Rico. Although well-

---

[84] Nunes, Woodley, and Rossetti (2022) show that household purchase behavior and use of EVs also contribute to their effect of emissions.

intentioned, such subsidies may be wasteful if they add considerable charging capacity in many areas long before the demand for it exists. In addition, car and truck drivers taking long-distance trips will be greatly inconvenienced if they do not have access to fast charging stations, which will require more acreage. Generally, urban and suburban motorists have been able to charge their EVs overnight at home or, in some cases, at workplaces.

Policymakers and some economists have argued that government subsidies are justified because a so-called "chicken and egg" problem (Li et al. (2017)) exists between electric vehicles and electric vehicle charging stations. The implication is that a suboptimal level of electric vehicle adoption could result because consumers will not buy EVs until charging stations are available enroute as needed and companies will not put in charging stations until there is sufficient demand for EVs.

To transportation economists, the problem is like the fundamental challenge facing transportation firms of aligning capacity with demand when capacity is committed in advance of demand and when capacity and demand influence each other. For example, airlines must commit to a level of capacity in advance of passenger demand by purchasing or leasing planes and choosing which routes to serve and how often to serve them. Those choices affect schedule delay (the difference between a traveler's preferred departure time and actual departure time), which is an important service quality variable that affects passenger demand for air transportation. At the same time, passenger demand affects airlines' capacity investment decisions.

Airlines have addressed the problem by forecasting macroeconomic performance and passenger demand and expanding (or contracting) their fleet, flight frequency, and cities served to align their capacity with expected demand. To be sure, airlines make forecasting errors, and

they adjust fares and capacity to reduce the financial cost of those errors (Morrison and Winston (1995)). However, to the best of our knowledge, no one has justified continual government intervention in the airline industry because carriers may make forecast errors. Airline fares and entry and exit were regulated because the industry was alleged to have characteristics of a natural monopoly that would lead to destructive competition. The airline industry became more competitive after it was deregulated and the competition spurred the industry to greatly improve its efficiency and to become more profitable (Morrison and Winston (1999)).[85]

The deregulation experience revealed that one must use caution when claiming market failures may prevent competition from working efficiently. The alleged market failures in transportation, such as economies of density and scale, have reflected a mix of carriers' technology and regulatory restrictions that prevented inefficient operations. When transportation firms were deregulated, they had the ability and incentive to reduce excess capacity and operate more efficiently thereby reducing costs and prices and increasing social welfare.

While the chicken and egg problem as applied to electric vehicles is a plausible theory and network externalities in the provision of electric vehicle charging stations may exist, we argue that those factors are not sufficient to justify government subsidies for electric vehicle charging stations for several reasons, including private sector incentives to correct the potential market failure; private sector provision in practice of electric vehicle charging stations; the

---

[85] Morrison and Winston (1986) assessed whether airline deregulation led to the social optimum, accounting for fares and service frequency, and found that it came close but that travelers originating at non hubs and small hubs could benefit from additional service. We are not aware of more recent assessments, which could show that airlines have provided more service to non and small hubs over time. Chapter 7 reports that low-cost and ultra-low-cost carriers currently serve routes that account for a large fraction of US passengers. Morrison and Winston (1986) also found that government's effort to increase service to small communities by providing subsidies through the Essential Air Service program resulted in less not more airline service to small communities because the program masked profit opportunities for carriers and discouraged additional entry to airports serving those communities.

significant potential for government failure; and the existence of an efficient approach, VMT taxes, to spur electric vehicle adoption.[86]

Private Sector Incentives. The electric vehicle industry has an incentive to spur EV adoption by increasing the availability of fast charging stations to help incentivize consumers to prefer EVs over a gasoline-powered car when they conduct a full cost comparison of those vehicles. Accounting for the cost of the gas and electricity and the value of time spent waiting to "fill up" a vehicle, it could cost drivers more to find an electric charger and charge an electric vehicle compared with filling up a gasoline-powered car unless they have access to a charger at home, which would allow them to charge the EV overnight.

 Fast charging stations that significantly reduce wait times when charging an EV could change the outcome of this comparison and possibly the full cost comparison.[87] Dorsey, Langer, and McRae (2022) find that motorists place a high value of time on having to refuel their vehicles and that motorists who adopt electric vehicles and rely on home chargers would receive significant benefits by eliminating travel and waiting time costs to refuel their vehicles. Importantly, they find that increasing the charging speed of the existing EV charging network yields nearly five times greater time savings than a proportional increase in the number of charging stations.

Private Sector Behavior. Historically, the major oil companies created and expanded the US network of gasoline stations *without government subsidies*. During the 1930s, the oil companies began to lease independently operated stations to local dealers. The oil giants are

---

[86] Li et al. (2017) find that the same amount of subsidy spent on building charging stations could result in twice as many EV sales as that spent on subsidizing demand. However, it is difficult to separate causality between charging stations and EVs, and the policy issue at hand is whether spending public money to build electric charging stations is justified because it would significantly improve social welfare.

[87] Fast charging may decrease battery capacity over time depending on the EV model and the climate where it is operated.

cognizant of the challenge to their business today as households transition from gasoline powered to electric vehicles and they oppose government efforts to subsidize electric vehicles and charging facilities.

At the same time, oil companies are taking constructive steps to play a role in an electric vehicle future. For example, BP claims the use of its pulse electric vehicle chargers is "on the cusp" of being more profitable for the company than filling up an internal combustion–powered car with gas, and they plan to grow their fast-charging business (Borras (2022)). Other oil companies are likely to do the same. Royal Dutch Shell Petroleum Company entered the EV charging industry by purchasing Greenlots, an American company that operates fast EV chargers, and plans to rebrand it as Shell Recharge Solutions and to operate some 500,000 charge points by 2025.[88]

Currently, most of the fast chargers in the US have been deployed by Tesla via its proprietary Supercharger network. Tesla plans to open its Supercharger network to non-Tesla EVs in the US in late 2022. In addition, General Motors is working to launch a program with its dealers to add more than 40,000 Level 2 chargers in local US and Canadian communities.[89] Other major providers include EVgo, which is partnering with GM and plans to deploy more than 3,250 Charge Point Stations across the country by 2025; Pilot, a travel operator, which is partnering with GM and plans to develop a national network of 2,000 electric vehicle charging stalls at travel centers to make it easier to recharge near highways; and Electrify America, the company investing $2 billion in US EV charging as part of Volkswagen's Dieselgate settlement

---

[88] https://insideevs.com/news/545338/greenlots-renamed-shell-recharge-solutions/
[89] Level 2 chargers double the charge speed of Level 1 units, and in most cases add around twenty or so miles of range to car batteries of a car for every hour that it is plugged in. Effectively that means the car must be plugged in overnight to travel a significant distance the next day.

agreement. Mobility startups, including, for example, Gravity Mobility, Revel, and Beam Global, are trying to make EV charging faster and easier in urban environments, such as New York City. Finally, TerraWatt Infrastructure Inc. has raised more than $1 billion to build electric vehicle charging stations for commercial fleets of cars, delivery vans, and trucks.

Commercial firms also see an opportunity to provide EV charging facilities to promote their product. For example, Starbucks is partnering with Volvo and ChargePoint, which operates a network of independently owned EV charging stations in 14 countries, to install EV chargers in its parking lots along a 1,350-mile route from Denver to Seattle, with stops available every 100 miles. By the end of 2022, there will be a total of 60 DC fast chargers installed at 15 Starbucks locations along the route. Other retailers are expected to follow the Starbucks experiment closely, and some are likely to replicate it.

One potential venue for EV charging stations that cannot be used is on Interstate highway rest areas.[90] EV charging stations also are not listed in the federal network of alternative fueling and charging infrastructure unless they are located less than five miles from an off-ramp. Federal law prohibits commercialization of rest areas to protect off-Interstate truck stop owners, which is prompting some states to close rest areas. Repealing the law could allow states to partner with private concessionaires to build modern travel plazas (as seen on toll roads that were grandfathered into the Interstate system) along with EV charging stations.

In any case, the major oil and automobile companies, commercial firms, and startups are helping to expand the nation's network of fast charging stations, although it is not clear that all those enterprises will continue to do so as the EV industry and charging network matures. As expected, EV chargers are currently concentrated in areas with high rates of EV adoption, which

---

[90] https://reason.org/wp-content/uploads/rethinking-interstate-rest-areas.pdf

typically means wealthier neighborhoods. However, like gas stations, the disparity in charging access should decline as companies find it profitable to put in new stations to cater to a broader spectrum of drivers, including drivers who work for rideshare companies like Lyft and Uber that have promised to make the shift to all-electric fleets over the coming decades.

Finally, automobile companies, such as the Chevrolet division of General Motors, are offering to obtain the required permits and to install chargers for its Bolt EVs at no cost in an owner's home or other location.[91] In addition, automobile companies are providing apps that indicate the location of the closest charger to a motorist's location.

Government Behavior. The claim that the US government can finance the creation and expansion of the electric vehicle charging station network efficiently assumes that policymakers have all the information that is needed to solve this optimal dynamic investment problem and the ability to implement the investment decision efficiently. The assumption is highly questionable because government has no established track record in making similar dynamic public investment decisions efficiently.

At the same time, government does have an established track record of misallocating large public expenditures to improve the highway network (Winston and Langer (2006)) and air traffic control (Morrison and Winston (2008)) by spending funds in areas of the country that yield low benefit-cost ratios. Accordingly, it is likely that government would waste public funds by building charging stations in places where travelers use them infrequently and by failing to build charging stations where travelers' demand is strong. In addition, government spending on electric vehicle chargers could generate social costs by providing disincentives for the private sector to invest in the charging network.

---

[91] https://www.chevrolet.com/electric/living-electric/home-charging

Responses to government actions by certain states and the private sector identify potential problems with those actions. For example, the Biden administration's plan to make fast EV charging stations available every 50 miles along major highways, including those in the rural western and midwestern states, is an example of building charging stations where travelers will use them infrequently. Some states that are receiving federal funds to as part of the 2021 infrastructure bill plan to request waivers of the requirement to build chargers every 50 miles. Private investors may reduce their investment in charging stations if states enable regulated utilities to own charging stations, which have competitive advantages because they are subsidized by ratepayer funds.

Optimal Policy. The issue of whether the government should spend public funds on electric chargers to spur EV adoption should be considered in the context of the welfare maximizing policy to reduce vehicle emissions externalities. Langer, Maheshri, and Winston (2017) provide evidence that a VMT tax would generate large social welfare gains including the reduction of vehicle pollution costs. Yet, proponents of government expenditures on charging stations claim that a VMT tax would produce a suboptimal level of welfare improvement because it would not address the chicken and the egg problem. However, in the absence of compelling evidence to the contrary, a VMT tax in combination with investments by the private sector to develop and expand electric vehicle charging facilities nationwide suggests that government spending would, at best, produce a small welfare gain and would more likely produce a welfare loss because of inefficiencies associated with its dynamic investment decisions and implementation of them.

Finally, as discussed in the next chapter, the chicken and egg problem will be eliminated when autonomous electric vehicles are fully integrated into a US transportation system, where

private companies provide mobility as a service (MaaS) and travelers share but do not own autonomous electric vehicles. In such an environment, private MaaS firms will ensure that the fast-charging network caters effectively to their operations, or they will invest in charging stations themselves.

5. The Appropriate Role for the Private Sector and Government in the EV Market

We have argued that consumers can adopt EVs efficiently without government intervention by responding to the evolving offerings of private auto companies. Evidence exists that some consumers are willing to pay as much as $7,000 extra for an EV than for a gasoline-powered vehicle because it contributes to reducing emissions (Gore (2021)). We also have argued that the private sector can supply EV fast chargers to meet and encourage consumers' demand for EVs and that government expenditures are not justified by the alleged chicken and egg problem. As EV technology improves, the growing share of EVs in the US vehicle fleet will enhance social welfare by reducing the costs of automobile emissions.

As noted, government has had the opportunity to and could still spur EV adoption by setting an efficient vehicle emissions charge as part of a VMT tax, which would increase the per-mile cost of gasoline-powered vehicles relative to the per-mile cost of electric vehicles. In response to concerns that such charges would enable EV users to avoid taxation, an efficient emissions charge also should include EVs' environmental costs. However, the draconian step, as some have proposed, of a government ban on gasoline-powered vehicles after a certain date would be highly inefficient (Holland, Mansur, and Yates (2021)) and would disrupt firms' production of and consumers' adoption to EVs and firms' provision of charging infrastructure.

Resources for Electric Vehicle Production and Transportation

Electric vehicles require significant resources to produce their batteries and considerable electricity to operate on US roads. One of the key minerals for EV battery production is lithium, the chief component of lithium-ion batteries that power electric vehicles. Currently, most of the world's raw lithium is mined outside of the United States, with China controlling roughly three-fourths of the lithium-ion battery megafactories in the world.[92] The United States is among the top five countries in the world in terms of its lithium reserves, but only one lithium mine, Albemarle's Silver Peak in Nevada, is operating in the country.

Recently, the richest known hard rock lithium deposit in the world was discovered just north of Plumbago Mountain in Newry, Maine (Cline (2022)). Other large potential domestic sources of lithium include extracting it from groundwater in the Salton Sea, California and processing it from spodumene ore from the Piedmont region of North Carolina. However, because of the lengthy permitting process to open a lithium mine and to extract lithium from its deposits, the domestic supply chain for EV battery production is slowed by several years. For example, permitting delays have prevented Lithium Piedmont from breaking ground for its lithium mine in North Carolina. Moreover, even after a permit had been granted many years ago to open the lithium mine in Nevada, environmental groups brought a lawsuit in 2021 to prevent it from operating.

President Biden invoked the Defense Production Act in 2022 to help accelerate the creation of a domestic supply chain for electric vehicle manufacturing. However, Biden's action will have little effect on getting critical lithium-ion battery projects running unless the permitting process for mining/raw material operations is significantly streamlined. Similarly,

---

[92] https://www.msn.com/en-us/money/markets/how-the-u-s-fell-behind-in-lithium-the-white-gold-of-electric-vehicles/ar-AASOtPx

Biden's Bipartisan Infrastructure Law signed in November 2021 includes more than $7 billion to help grow the battery supply chain, but it is not clear if those funds will be used constructively if regulatory constraints continue to impede lithium-ion battery projects.

Government also has a critical role in ensuring that the US has enough electricity generating capacity to supply EV transportation. For example, if 60% of all US cars were electric vehicles by 2050, the nation's electricity capacity would need to double (Groom and Bellon (2021)), which would require either a massive government investment in the electricity grid or a mechanism whereby private quasi-utilities could undertake those investments. At the same time, the government must respond effectively to various environmental groups that are likely to align with local NIMBY groups to block new transmission lines and other technology.[93]

<div align="center">Autonomous Electric Vehicles</div>

Automakers and technology firms will eventually combine electric vehicles with autonomous vehicles and offer those to the public. We have argued here that the government has inappropriately intervened in the adoption of electric vehicles and in the availability of charging stations. We argue in the next chapter that government must take certain actions that will greatly benefit society by expediting the adoption of autonomous vehicles, which in turn will expedite the adoption of autonomous electric vehicles.

---

[93] Governments also must make a careful decision about whether to spend public funds to construct electric vehicle charging roads. As noted, such roads could address the problem of range anxiety and eliminate the time costs of charging an EV because the EV could be charged while it uses the road. On the other hand, the cost of constructing EV charging roads may be considerable and the technology could be obsolete if technological improvements in batteries increase the range of an EV and if fast EV chargers are widely available for public use.

Chapter 6. Consumer Adoption of Autonomous Vehicles and the Role of Government

Clifford Winston and Jia Yan

1. Introduction

Autonomous vehicles (AVs) represent a significant innovation in transportation, which could eventually solve transportation problems that the public sector has failed to solve for many years, including the automobile externalities of nerve-wracking congestion, millions of fatal and nonfatal accidents, and violent police confrontations with drivers. In addition, in a world where COVID-19 may be a persistent concern, AVs can facilitate greater economic activity without increasing the spread of the virus. Yet, despite those and other potentially significant benefits from AVs and the fact that automakers and technology companies are planning to eventually combine electricity-powered engines and autonomy to produce autonomous electric vehicles, the government has shown little interest in expediting the adoption of autonomous vehicles.

In this chapter, we provide an overview of the history of autonomous vehicles and its technology, the competitive environment of the emerging AV industry, and AVs' potential benefits and costs. We argue that government has a critical role to play in enabling the public to realize the full benefits from autonomous vehicles by establishing testing and public adoption protocols and by upgrading the infrastructure to facilitate AV operations.[94] Government's failure to embrace this role will delay the adoption of autonomous vehicles in the United States and put America behind other countries that are prioritizing autonomous vehicles.

---

[94] Parts of this chapter draw heavily on Winston (2021c) and Winston and Karpilow (2020).

2. An Overview of Autonomous Vehicles

The Pentagon has long been interested in developing autonomous tanks for the battlefield and autonomous trucks to deliver food, fuel, and other logistics to soldiers engaged in combat. In 2004, through the Defense Advanced Research Projects Agency, it sponsored its first autonomous vehicle competition. However, there was little interest among car companies in making a large investment to build and sell autonomous vehicles. Google, and subsequently Waymo, its self-driving car project, took the lead in introducing autonomous vehicles to the world in the 2010s. A few years later, US and foreign automakers, non-auto technology firms, and various start-up ventures were in hot pursuit, either by themselves or in a partnership. Today, the global autonomous vehicle industry is engaged in advancing the technology and hundreds of new companies are developing various components of the technology.

An autonomous vehicle drives itself by using a combination of onboard technologies to measure the distance of the car from various objects, including pedestrians, bicyclists, and other cars. GPS, supplemented with highly detailed up-to-date digital maps, locates vehicles, and indicates which has the right of way. Communications among AVs, as well as between AVs and roadway infrastructure, determine the location and intention of other vehicles, the condition of the roadway, and the status of traffic signals. Cooperative automated driving, in which multiple (or even all) vehicles are effectively plugged into a local network, has the potential to smooth traffic flow, improve travel time and its reliability, and virtually eliminate accidents along with the loss of life and damage to property.

The technologies are advancing at such a rapid clip that important breakthroughs in different parts of the world occur regularly. It has become very difficult to keep up with the latest

technological advances, new players, new partnerships, and new services in the United States and abroad. Some of the most important recent developments include:

— MonoCon's enhancement of AVs' visual capabilities by extracting three-dimensional data from two-dimensional camera views to cut an AV's reliance on LIDAR (Light Detection and Ranging) while maintaining safety and slashing operating costs.

— X-ray style vision technology that enables autonomous vehicles to see pedestrians and cyclists through obstacles, and ground positioning radar that allows AVs to image a road's subsurface for precise positioning regardless of poor weather or visibility, or faint road markings. There is growing recognition that AVs will rely on a combination of sensing modalities, including visible light, near-infrared, thermal imaging, radar, and lidar.

— Techniques that enable AVs to navigate complex traffic situations, such as merging into heavy traffic when lanes disappear on a highway (Tajalli, Niroumand, and Hajbabaie (2022)).

— Autonomous vehicle industry participants, including but not limited to Google's Waymo, GM's Cruise, Amazon's Zoox, and Aurora, are committed to offering full Level 4 or 5 autonomy.[95] GM's Cruise has petitioned the National Highway Traffic Safety Administration (NHTSA) for permission to put the driverless Level 5 Cruise Origin into commercial service, with delivery of the vehicle in 2023.

---

[95] The Society of Automotive Engineers has created a widely accepted scale of automation systems, which ranges from level 0 (no autonomy) to level 5. Level 4 indicates systems that are self-driving but operate only under well-specified conditions, such as certain road types or geographic areas. The Operational Design Domain, spanning all scenarios that an AV may encounter, determines the difference between Level 4 versus Level 5 systems. Motorists have already grown accustomed to some level of independence from their cars thanks to advanced driver-assistance systems. Many vehicles, for instance, have collision-avoidance systems or parking assistance features that place them at a level 1 or 2 on the scale. At levels 3 to 5, the amount of control the system has increases.

— In 2022, Mercedes-Benz is expected to launch the first mass-production Level 3 car using its Drive Pilot technology and BMW is widely expected to roll out Level 3 technology in the new 7 Series. At Level 3, the driver can allow the car to operate under automation, but the system may prompt the driver to take over driving if it detects conditions outside of its automation capabilities. Mercedes is pushing for its Level 3 technology to be certified in United States, while BMW is not likely to try to certify a Level 3 car in the US until uncertainties about liability are resolved.

— Sony is entering the global industry with an electric vehicle with autonomous features.[96]

— Apple is accelerating work on an electric car project, aiming for a fully autonomous vehicle by 2025.[97]

— Mobileye, an Israeli subsidiary of Intel, and Geely, China's largest automaker, are teaming up to build a self-driving car that will supposedly go on sale in China in 2024. Waymo is planning to work with Geely on a fleet of electric robotaxis.

— Autonomous delivery services are being provided by Nuro to deliver Domino's Pizza in Houston and elsewhere and Waymo has launched its self-driving grocery delivery service in California.

— Fully autonomous passenger services also are being offered. In the United States, GM's Cruise received a permit to charge for fully driverless rides at nighttime in San Francisco, while Waymo also is operating its cars without any human control in San Francisco and offering free rides to its staff. In China, Baidu is operating ride hailing

---

[96] https://www.sony.com/en/SonyInfo/vision-s/safety.html
[97] The Apple car chip is the most advanced component that Apple has developed internally and is made up primarily of neural processors that can handle the artificial intelligence needed for autonomous driving.

services without a driver or a person overseeing safety in the cities of Wuhan and Chongqing, AutoX, an Alibaba startup, is operating autonomous robotaxis in Shenzhen, and Baidu and Pony.ai, which is backed by Toyota, are starting robotaxi service in Beijing. The Chinese robotaxi companies also are testing their vehicles on California roads to draw the attention of global investors.

— The United Kingdom is working through legal issues to be resolved in advance of the first iterations of self-driving features on its driverless cars that are expected to be on its roads by 2025; a uniform law in Germany is helping German companies test AVs throughout the country; China is adopting 5G technology—car to car communication—to help AV adoption; and Toyota's Woven City at the foot of Mount Fuji in Japan will have only autonomous cars, where the city's hand-picked residents will use them in a living laboratory (Davis (2021)).

— Because there may be instances when humans must take over and drive an autonomous vehicle, for example, when the AV is operating in an environment that it cannot navigate, General Motors has submitted a design for an autonomous vehicle system designed to measure and train new drivers. It is an open question whether such training will exist after AVs are fully adopted and whether it will be required for drivers who use AVs in certain environments.

— Finally, AV companies are thinking about passengers' experience and exploring ways to prevent motion sickness that some people may experience when they focus on something within the cabin of the car, such as a book, and the senses in their brain

cannot tell them if they moving or stationary. GMs' Cruise has patented a system that can sense a passenger's motion sickness and prevent it from occurring.[98]

Notwithstanding the abundant recent activity and much more to come, the skeptics are still skeptical. C.Winston and J.Winston (2022) report that AVs are described as "one of the most hyped technology experiments of this century," which are incapable of reversing the growing death toll on American roads for many years to come, if ever. The skeptics ignore that aviation in the early 1900s experienced many setbacks enroute to its development as a successful commercial industry, with each setback identifying another problem to be fixed and that eventually was fixed. In any case, the continuing evolution of autonomous vehicles calls for increasing optimism that global competition and cooperation among auto companies and technology firms will succeed in developing a commercial AV industry that will overcome any serious technological challenges. We return to the skeptics later.

3. Benefits and Costs of Autonomous Vehicles

Autonomous vehicles' most important economic impact is likely to derive from their greater travel speed and reliability, compared with non-autonomous vehicles, which results from a smoother traffic flow as vehicles travel faster and closer together without causing accidents that can tie-up traffic for hours, especially during rush hours. Note, too, that unlike drivers of non-autonomous vehicles, the software managing AVs will not be tempted to compound traffic slowdowns by rubbernecking past wreckage.

An improvement in average travel time and a reduction in the variation in travel time are likely to generate efficiency gains beyond the transportation sector by benefiting urban areas that

---

[98] https://www.luxury-minds.com/gm-is-developing-anti-motion-sickness-tech-for-autonomous-vehicles#

are now facing limits linked to traffic and parking woes. Shorter and more reliable commute times would expand individuals' choices of employers and employers' choices of workers, adding a welcome increase in potential competition in local labor markets. As discussed in chapter 12, the cost of moving goods by truck would fall, raising productivity, increasing the flexibility of businesses to buy from the best or cheapest suppliers, and lowering logistics costs because firms would be able to reduce their inventories. By reducing firms' costs, autonomous trucks would reduce consumer prices.

Winston and Karpilow (2020) caution that no one knows with great precision how much the widespread adoption of AVs could raise the nation's annual economic growth rate. However, using conservative assumptions, they estimate that AVs could raise the nation's annual growth rate by at least one percentage point. Given that the construction of the highway network greatly increased national welfare (Allen and Arkolokis (2014)), it is intuitively plausible that AVs could generate enormous benefits by significantly reducing travel costs, which would greatly increase the value of the highway network by exponentially increasing access to the nation's human and physical resources, and enabling more people to work, shop, trade, and produce goods.

Importantly, by improving accessibility for disadvantaged travelers, the benefits from autonomous vehicles also would have positive distributional features. The improvements in accessibility would reduce the persistent differences in travel times between Black and White commuters that exist in large, segregated, congested, and—especially—expensive cities (Bunten et. al (2022)) and would reduce the long commutes endured by carless households (Winston (2013)). Shorter and more reliable commutes would reduce differences in racial employment disparities, increasing single mothers' employment and number of hours of work per week, and reducing poverty and the use of welfare (Scribner (2021a)).

<center>Additional Benefits</center>

Autonomous vehicles also could produce a host of other benefits, including improvements in public health, better access to leisure activities, more efficient land use, and even assistance in addressing two leading social problems at this moment: the COVID-19 pandemic and violent police confrontations.

Public Health. AVs have the potential to produce improvements in public health, which are valued in the trillions of dollars, by making highway travel far safer and by relieving drivers from the stress of driving in congested conditions. According to NHTSA, the social costs from motor vehicle crashes approach \$1 trillion annually, accounting for the loss of life of some 40,000 people, injuries to roughly 4 million people, and economic losses from damage to 25 million vehicles.[99] Surveys indicate that commuting is the leading daily activity for which individuals' dominant attitude is negative (Kahneman and Krueger (2006)). Indeed, automobile commuting in congested conditions can damage emotional health by causing stress that leads to road rage and household violence. AVs also could generate substantial health and productivity gains by converting stressed-out commuters to more relaxed passengers.

Greater Access to Leisure in and out of the Vehicles. Manufacturers will have incentives to change the interiors of driverless vehicles to enable consumers to enjoy a variety of activities while traveling—among them, dining, watching movies, and sleeping during long intercity trips. Perrine, Kockelman, and Huang (2020) point out that autonomous vehicles could increase competition in intercity passenger transportation by capturing market share from aviation for trips that take an hour or two by air. We return to this issue in chapter 10. Finally, autonomous vehicles

---

[99] https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813298

will increase mobility for people who cannot drive because of age, infirmity, or lack of a driver's license without threatening public safety.

Land Use. Currently, the United States has more parking spaces than cars. Because it is widely expected by participants in the autonomous vehicle industry that people would generally share, not own, autonomous vehicles, valuable land that was once used for parking cars in central business districts and for home garages would be freed up to, say, build new housing or expand existing homes. The reduced parking footprint for suburban malls also could make land available for housing or other beneficial uses because AVs would leave passengers at designated drop-off zones and would then be directed to parking facilities where they could be stacked vertically and easily retrieved when summoned. Importantly, AVs would weaken the justification for restrictive land-use policies such as parking requirements that aim to reduce congestion but which, in practice, have little effect on congestion and raise housing prices by reducing the quantity of available housing.

Fewer Disamenities from Highways. Brinkman and Lin (2022) discuss the disamenities associated with downtown freeways, which are reflected in population declines in census tracts nearest to freeways. Autonomous vehicles could potentially reduce the disamenities associated with living close to a freeway by reducing the presence of and noise associated with congestion; decreasing emissions as cars move more smoothly instead of idling (autonomous electric vehicles would produce an even greater decrease in emissions); and virtually eliminating accidents, which would further reduce congestion and the noise created by ambulance and police car sirens.

Pandemics. The world is increasingly accepting the likelihood that COVID-19 will be with us for the foreseeable future. AVs will prove invaluable by maintaining economic activity while facilitating social distancing for commuters and delivery services. Lower-paid essential service

workers would especially gain by having access to sanitized vehicles for their sole use at manageable cost, and as discussed in chapter 2, by traveling in much safer highway conditions.

Safe and Efficient Policing. A large fraction of police work consists of enforcing traffic laws, which may result in violent encounters with drivers. On average, more than one officer per week is killed in a highway accident, accounting for nearly one-third of all police officer deaths in the line of duty. There also have been high-visibility motorist deaths at the hands of the police that resulted from an alleged traffic violation. AVs could virtually eliminate the need to use the police to enforce traffic laws, eliminating violent encounters with vehicle occupants, and either sparing taxpayers the expense or allowing for police redeployment to ameliorate public safety concerns that are not being addressed.

<div align="center">Costs</div>

The potential costs of autonomous vehicles include congestion and sprawl, job losses, rising highway and transit deficits, safety and liability concerns, and technology failure. Efficient government policy can play an important role by mitigating some of those costs.

Congestion and Sprawl. Critics claim that AVs are likely to induce more automobile travel and congestion, and to create sprawl by making longer commutes more practical so people can live in less expensive homes on larger lots farther from their workplaces. What are the counterarguments? First, by eliminating vehicle accidents, AVs will eliminate incident delays, which account for one-third of all delays. Second, AVs will enable highways to accommodate a greater number of vehicles because AVs can safely travel closer together at high speeds; highways can be reconfigured with more, but narrower, lanes; and highways can eliminate the breakdown lane and use it for vehicle travel because AVs will undergo a "safety check" before they can enter a highway and travel en masse with other vehicles.

Finally, policymakers can help reduce congestion and sprawl by implementing congestion pricing, which would affect households' decisions of when to travel and where to live. Efficient road pricing would be more politically palatable in the new autonomous driving environment because, as noted, people would reduce the cost of automobile travel by not owning cars and instead would hire an autonomous vehicle when they need transportation. Hence, travelers would perceive congestion tolls as like a toll that they pay when they use a taxi or surge charges that they pay when they use shared transportation, such as Uber or Lyft, and would therefore be likely to accept those tolls as a valid part of the lower full cost of autonomous vehicle travel, including travel time costs.

Jobs. There is no doubt that the widespread adoption of AVs would force hundreds of thousands of people who drive for a living to look for new work. But it is likely that those job losses would be more than offset by additional jobs created by the demand for workers in everything from maintaining AVs to updating the very accurate maps needed for them to operate safely and, more importantly, by creating new job opportunities throughout a more efficient economy.

Highway and Transit Fiscal Deficits. Autonomous electric vehicles (AEVs) will virtually eliminate gasoline tax revenues that are the backbone of highway finance. Efficient per-mile road pricing that charges AEVs for emissions created in the production of electricity and for congestion that AEVs create on the road would simultaneously raise highway revenue while reducing congestion and emissions.[100]

By the same token, AVs will probably increase deficits in urban mass transit by contributing to its long-run decline in ridership. However, the efficient response is to automate

---

[100] Adler, Peer, and Sinozic (2019) study the effects of AEVs on public finance.

transit and see if it can compete effectively for passengers in the autonomous urban transportation environment. If not, transit's operations should be curtailed because it would effectively be replaced by a personalized transit system provided by autonomous vehicles.

Safety Concerns. Instead of focusing on autonomous vehicles' potential safety benefits, some commentators have raised concerns that AVs will reduce safety as exemplified by incidents involving Tesla vehicles and its Autopilot feature where a driver hit another car and killed two people, and where another driver was killed while playing a video game. Both accidents indicate the importance of the federal government setting national testing and safety standards, which, as we discuss below, it has thus far failed to do.

Liability Concerns. The legal community has just begun to confront the issue of who pays when AVs have accidents. Liability has yet to be firmly established, and, as with other new products and technologies, the law is likely to evolve as AVs are adopted and the various stakeholders—insurance companies, the tort bar, and, of course, the public—gain greater experience with them.

It will be important to clarify liability, regardless of how it is assigned, because the AV industry will need certainty on this matter to thrive. At the same time, liability should fade markedly as an issue of contention because AVs will be so much safer than conventional vehicles and because, like airlines, manufacturers will continue to solve their residual safety issues. Note, too, that the data automatically collected by AVs will be very useful in determining the cause of an accident and for providing guidance on preventing future accidents attributable to that or related causes.

Technology Failure. Critics of AVs assert they will never achieve Level 5 or even lower automated levels of operation. Economists have stressed the difficulty of explaining technological

change and innovation, with Solow (1957) claiming that it is exogenous and Romer (1990) attempting to identify endogenous influences including but not limited to investments in human capital and research and development. In the specific case of AVs, Bishop (2022) cautions that it is premature to make any strong predictions about what will happen when autonomous vehicles are deployed. The critics have not offered any dispositive arguments for why automakers and technology firms cannot learn from their mistakes and why AV technologies cannot improve over time. In the concluding chapter, we discuss concerns about the likely success of autonomous transportation technologies in more detail.

4. Government Actions to Expedite the Adoption of Autonomous Vehicles

AVs will operate on a publicly owned and managed road system. Hence, policymakers and regulators have a critical (and unavoidable) role to play to ensure that society will realize the potential benefits from AVs. Specifically, policymakers must establish a framework for vehicle testing and adoption, invest and modernize the highway infrastructure to facilitate safe vehicle operations, and reform highway policies to encourage efficient operations.   However, policymakers have made little progress in fulfilling those responsibilities, which will slow the adoption of autonomous vehicles.

Testing Framework

In 2018, Congress drafted, but failed to pass, important AV legislation that would have clarified and expedited national vehicle testing and adoption. Four years later, Congress has still not passed a version of this legislation, and a recent hearing showed no movement, as members raised concerns about protection from job losses and differentiating between advanced driver assistance systems and AVs. Republicans assert that AV legislation will advance in 2023 if the

Republicans take over the majority in the US House of Representatives following the midterm elections in 2022.

In the absence of new laws tailored to autonomous vehicles, the National Highway Traffic Safety Administration (NHTSA) has put forward voluntary guidelines and in 2021 required companies to report accidents involving automated driving systems (Shepardson, Jin, and White (2022)). But the agency has not issued comprehensive standards for robot-driven cars or trucks.

Motor vehicle manufacturers are free to certify for themselves that a vehicle feature is safe; NHTSA steps in if new features turn out to be a safety hazard. For example, NHTSA has pressed Tesla to improve its vehicles, so they do not keep moving through stop signs rather than coming to a complete halt. Recently, it opened an investigation into Tesla's Autopilot. Generally, however, NHTSA approaches Elon Musk, Tesla's CEO, using a combination of pressure, flattery, and threats to persuade him to comply with federal safety measures instead of instituting and strongly enforcing unambiguous guidelines for the autonomous vehicle industry to test vehicles safely (Siddiqui (2022)).

NHTSA also has amended the occupant protection Federal Motor Vehicle Safety Standards (FMVSS) to account for autonomous vehicles that do not have the traditional manual controls associated with a human driver because they are equipped with Automated Driving Systems (ADS). The rule makes clear that, despite their innovative designs, vehicles with ADS technology must continue to provide the same high levels of occupant protection that current (non-autonomous) passenger vehicles provide. For example, the rule works through the kind of airbags and seatbelts autonomous vehicles should install.[101]

---

[101] Automobile and technology companies can petition NHTSA to waive requirements for manual controls on their autonomous vehicles. GM's 2021 Cruise has applied for one, and autonomous delivery vehicle developer Nuro received a waver in 2020.

States and localities have adopted regulations allowing testing within their borders, sometimes with an eye toward attracting new high-tech businesses. However, federal legislation is necessary to jump-start the formal adoption process and to ensure that vehicles do not have to shift from autonomous to manual operations because they cross a state line.

<div align="center">Government Investments</div>

In a fully optimized system, AVs would be connected to other vehicles and their surroundings, including pedestrians, infrastructure, and the road network. Some states are making investments to upgrade their infrastructure to align it in advance with AV technology. For example, Panasonic and Utah's Department of Transportation are jointly developing a system that enables vehicles, roads, and traffic signals to continuously communicate about conditions, location of obstacles, signal timing and the speed, direction, and position of all cars in the network.

Michigan, for its part, is working with a Google-funded startup to transform a stretch of road between Detroit and Ann Arbor into a connected highway for testing purposes. Recently, Governor Gretchen Whitmer signed the first legislation adopted in the United States specific to the deployment and operation of technology-enabled roadways for connected autonomous vehicles. The legislation authorizes the Michigan Department of Transportation to designate automated vehicle roadways throughout the state, enter into agreements with technology partners to operate them, and charge users of Level 3 autonomous vehicles for using the roadways.

Several states are installing fiberoptic lines in roads that can send electronic warnings to AVs about hazards ahead and other information to keep them aware of their surroundings. By contrast, other states seem content to wait for the federal government to provide guidance before moving forward. Regardless of who is at fault, states that delay upgrades to their infrastructure could delay the use of autonomous vehicles on their roads.

The adoption of 5G wireless technology is critical to launching the era of autonomous vehicle connections, with ultrafast speeds, greater capacity, and ultra-low latency. Rollouts have begun, but delays have been caused due to battles between Congress, the Federal Communications Commission, and the Federal Aviation Administration. Government's lack of regulatory agency coordination has become another potential source of delay for the nation's adoption of autonomous vehicles. Autonomous vehicle companies also are interested in satellites to update information on road conditions and to have more accurate information about vehicle locations. Government regulators will have to balance the AV industry's desire to use satellites with other interests that wish to use them.

Finally, there are over 30 million lampposts in the United States, which could be used to mount an electronic communications router or a 5G transmitter, to assist autonomous vehicles' communications with the infrastructure and, for example, to warn of impending delays and indicate alternative routes.

<u>Reforming Highway Policies</u>

AVs are not immune to highway inefficiencies. As noted, their operations could be disrupted, for example, by congestion and damaged pavement that requires vehicles to slow down or change lanes. Thus, it is important for highway policies to complement their operations by improving traffic flows.

As summarized in Winston (2021a) and Winston and Mannering (2014), highway performance could be significantly improved by adopting efficient policies, including:

— Congestion pricing for cars and trucks and, as discussed in chapter 12, pavement and
    bridge wear pricing for trucks.

— Optimal investment in highway durability and in pavement design to increase pavement durability.

— Variable speed-limits and real-time traffic signaling to reduce delay and improve traffic safety.[102]

— Adopting the latest technologies to improve highway design and reduce maintenance costs.

— Reducing project costs by expediting the procurement process, reforming environmental and other regulations to reduce the extent that they delay projects, and selecting socially beneficial projects based on cost-benefit analysis.

Unfortunately, the federal government has shown little interest in reforming highway policy along the preceding lines to improve the system's performance. President Biden's Investment Act increases highway spending more than $100 billion above baseline levels, but no acknowledgment is made of the various ways that accumulated policy inefficiencies substantially compromise highway travel and reduce the returns from highway expenditures. Increasing spending does not constructively address those problems and the Act does not contain any other useful actions.

To make matters worse, the US Department of Labor is proposing to further increase the cost of federal and federally assisted highway construction projects covered by the Davis-Bacon Act, which mandates the "prevailing wage" be paid on applicable projects. Research indicates that the Act has significantly inflated the labor costs of highway projects (Winston (2013)). The Department's proposed change will make it easier for a union to have its higher collective

---

[102] The artificially intelligent Surtrac system (Scalable Urban Traffic Control) is one of the first to gather information on vehicular flow and use it to adjust traffic lights in real time. Such a system could be used for the entire road system.

bargaining agreement rate set as the prevailing wage in the locality. In addition, President Biden has signed executive orders that increase the stringency of National Environmental Policy Act requirements, which will increase the time and cost of project delays to satisfy environmental mandates.

The states are making modest progress in exploring efficient pricing policies as some 14 states have run piloted VMT-based taxation programs. Oregon, Utah, and Virginia have implemented VMT taxes, albeit voluntary. As in the case of testing, states currently cannot implement a nationwide program that would not be disrupted because a vehicle crosses a state line. However, the International Fuel Tax Association, authorized by the Intermodal Surface Transportation Equity Act of 1991, is examining a way that states potentially could implement a nationwide road usage charge program.

Even when policymakers are interested in instituting an efficient highway policy, they may encounter difficulties in implementing it promptly and efficiently. For example, in 2019, New York City's legislators approved a congestion pricing plan for motorists driving into Manhattan south of 60th Street. Because some of the roads were built with federal funds, city officials needed to receive federal approval, which was expected to be granted after officials responded to the Federal Highway Administration's technical questions about how the plan would be implemented.

Supporters of the plan have become frustrated because The Metropolitan Transportation Authority has delayed naming the six-person review panel to guide the state through the congestion pricing process and the federal government has dragged its feet in approving an NYC congestion pricing plan. The latest projections are that congestion pricing in NYC will be implemented sometime in 2024, or five years after the city's legislators approved the plan. However, as pointed

out by (Gelinas (2022a)), some New Yorkers may still be asking themselves in 2024: Whatever happened to congestion pricing?

The delays in implementing congestion pricing in NYC are costly because congestion pricing is likely to improve travel by discouraging motorists from driving on the most congested routes during the busiest times of day. At the same time, the revenues generated by congestion pricing are likely to be wasted because state legislators have mandated $1 billion of annual toll revenues to support $15 billion in loans for its highly inefficient public transit system instead of using those revenues to improve the road system.

5. Highway Privatization

If the lack of constructive reform in highway pricing and investment policies compromises the performance of autonomous vehicles, privatization of the road system could be considered as an alternative approach to improve the system's efficiency and technology. Although privatization is in keeping with the book's theme of competition and innovation to improve the transportation system, it would be a radical departure from government provision of highway services that would raise significant challenges that we cannot ignore.

Friedman and Boorstin (1996)[103] argue that it would be feasible for private enterprise to provide certain US highways and benefit users. Based on government provision, highways appear to be highly profitable, although the public is not fully benefiting from its expenditures on them. The public expresses its demand for highway services by paying for them in the form of license fees, gasoline taxes, and tolls on the toll roads that have been mushrooming in recent years.[104] But a portion of public expenditures is used to subsidize other services, such as transit, and, as noted,

---

[103] The draft was written in the early 1950s but was not published until 1996.
[104] Some tolls have disappeared when the road has been "paid off."

costs are significantly inflated, which reduces the amount of funds that are used to maintain and expand highways.

By comparing the size of public expenditures on highways with the larger size of other private transportation industries, the provision of highways is clearly not too expensive nor too large-scale an activity to be conducted by private enterprises. And thanks to technical advances in automated tolling and collection, it is possible to set efficient prices for cars and trucks for their use of the roads without disturbing the flow of traffic.

Highways do have elements of a natural monopoly, which pose a challenge for allowing them to be provided by private enterprise and not subjecting users to monopoly charges. For example, city streets and roads should continue to be provided by governmental authorities. However, the evolution of the US transportation system facilitates competition for long-distance highway travel provided by alternative modes, such as airlines and passenger rail in some corridors, and for the potential of competition created by alternative intercity routes.

Winston and Yan (2011) explore the possibility of private highway competition empirically by designing a competitive private highway market in California on State Route 91 consisting of two routes with equal lane capacities that would be operated by two different private highway companies. A third party would represent motorists, negotiating tolls and capacities with the private companies to obtain a contract equilibrium. We found that motorists could gain from highway privatization when the contract equilibrium consisted of differentiated prices and service—that is, when motorists were given a choice of paying a high toll to use lanes with little congestion or paying a low toll to use lanes that were highly congested.

In practice, the United States has had little experience with highway privatization. It is uncertain whether policymakers could effectively design a competitive private highway system

and whether sufficient managerial talent exists to operate several competing highway companies that would remain profitable and provide highway services capable of improving motorists' welfare. The public sector has allowed some privatization to occur to raise revenue to fund highway projects as part of long-term leases and to build new capacity. The most encouraging public–private partnerships have led to high-occupancy toll lanes, where the private sector has built, financed, and maintained express lanes that have benefitted travelers. However, highway public-private partnerships have a limited history in the United States as investments and efficiency gains have been small (Engel, Fischer, and Galetovic (2021)).

In theory, owners of private US highways would have the financial incentive to recruit workers from all over the world with the technological expertise to upgrade the infrastructure to facilitate autonomous vehicle operations and to improve the condition of the highways through efficient pricing. Foreign experience also may be encouraging. For example, many roads in Sweden and Finland are operated by the private sector and are maintained by local communities. The government works in conjunction with road owners to reduce the costs of repair and maintenance.

The privatization option should be explored with carefully designed experiments, which eventually could give policymakers valuable guidance on whether to privatize the entire or part of the US highway system and under what conditions. Although policymakers have not shown any interest in conducting highway privatization experiments with that goal in mind, the privatization option may gain attention when autonomous vehicles are adopted if public sector management and operation of the road system significantly limit AV operations.

6. Conclusions

The private sector and policymakers have separated the development of and policy toward autonomous and electric vehicles. Ultimately, autonomous and electric vehicle technologies will be combined to develop autonomous electric vehicles. AEVs will become a game changing innovation that is a product of global competition and cooperation and will enable urban transportation systems to significantly reduce the automobile externalities of congestion, pollution, and accidents, increase productivity, and benefit all socioeconomic groups, especially less affluent people who have limited mobility. However, the government must carry out its responsibilities if the United States is to realize those improvements; thus far, it has failed to show much interest in doing so.

Because AEVs are a global endeavor, policymakers will be incentivized to take constructive actions if other countries take the lead in adopting this technology. President Biden famously proclaimed that China was "eating our lunch" in electric vehicles as motivation for increasing funding for EV adoption and charging stations. Hopefully, Biden will have a similar reaction, and take appropriate actions, if it becomes clear that China is eating our lunch in autonomous vehicles.

Part I of this book has explained how private sector competition and innovation could evolve to transform the US urban transportation system from a highly inefficient network of highways, public transit, and taxis to a much more efficient system characterized by autonomous electric vehicles, whose operations are facilitated by a highway system with advanced communication capabilities. Government policymakers are in a critical position to determine how quickly this transformation can occur.

As we shall see in Parts II and III, private sector competition and innovation also have the potential to greatly benefit the public by transforming the intercity and freight transportation systems, where again autonomous vehicles can improve efficiency and government will play a critical role in determining the speed of the systems' transformation.

PART II

INTERCITY PASSENGER TRANSPORTATION

Deregulation of the airline industry in 1978 significantly improved the efficiency of the US intercity passenger transportation system, benefiting both airlines and travelers. Airline safety also continued to improve as no major airline has been involved in an accident in the United States since 2009. A recent summary of the post-deregulation intercity passenger transportation environment reported that in 2020[105]:

— The price to fly, including all the extra fees like baggage, was over 60% cheaper than in 1980.

— 87% of the US population said they had taken a commercial airline trip, the highest number ever recorded.

— Fewer people must take connections, as a record 90.1% of the top 2000 largest domestic markets had nonstop service, up from under 70% in 1990.

— Almost 90% of the US population has access to a low-cost carrier.

— Satisfaction with airline travel is at its highest point in the last 20 years, at over 75%, according to the American Consumer Satisfaction Travel report.[106]

— Deregulation of the intercity bus industry in 1982 benefited the industry and travelers as intercity bus experienced much greater passenger growth than Amtrak experienced.

In this part of the book, we discuss additional ways that competition and innovation could significantly improve the intercity passenger transportation system. We provide new evidence on the potential benefits to travelers from the US allowing foreign airlines to serve

---

[105] Specific figures for airlines are from *Facts from Airlines for America* unless indicated otherwise.
[106] https://www.theacsi.org/the-american-customer-satisfaction-index

domestic routes—that is, granting cabotage rights. For example, although British Airways can serve the international route comprised of New York Kennedy and London Heathrow airports, it—as well as any other foreign airline—is not permitted to serve the domestic route comprised of New York Kennedy (JFK) and Chicago O'Hare (ORD) airports. If British Airways and other foreign airlines were granted cabotage rights, they would be free to serve the JFK-ORD route and any other domestic routes that they deemed profitable and that had available gates and operating times.

Cabotage is an essential step toward global airline deregulation because it will allow carriers to provide "seamless" travel from many origins in one country to many destinations in another country, eliminating transfers to other airlines. Thus, global airline competition will spur additional competition in domestic markets as well as more competition in international markets.

We also analyze empirically public airport competition to motivate the potential benefits from privatizing US airports to spur greater airport competition, which could lead to greater airline competition. In addition, we show that private airports would have greater incentives than public airports to adopt socially desirable innovations, such as heated runways. Finally, we discuss how additional sources of competition and innovation could lead to new modes and services, such as air taxis, and enable the intercity transportation system to provide greater benefits to the public for many years to come.

Chapter 7. Would US Travelers Benefit from Entry by Foreign Airlines? Simulating the Effect of Cabotage Based on Low-Cost Carrier Competition in US and European Union Markets

Xinlong Tan, Clifford Winston, and Jia Yan

1. Introduction

Some forty years after Congress deregulated the US airline industry, controversy remains over whether the industry is sufficiently competitive and whether government should intervene to increase competition. The wave of airline mergers since 2005, which enabled the remaining large legacy carriers, American, United, and Delta, to solidify dominance of their hubs in large cities while they reduced service to smaller cities, has spurred recent concerns that fares will increase, and service quality will decrease.[107]

At the same time, low-cost carriers (LCCs), such as Southwest and JetBlue Airlines, and ultra-low-cost carriers (ULCCs), which include Allegiant, Frontier, and Spirit Airlines (Bachwich and Wittman (2017)), have gained a greater share of the US market and have offered new service in markets abandoned by legacy carriers.[108] For example, Delta made significant cuts in service at Cincinnati/Northern Kentucky International Airport, but Allegiant, Frontier, and Southwest have provided new service. Similar changes are occurring at other former hubs, including Pittsburgh, Cleveland, Memphis, and St. Louis.

Although some larger carriers have exited through consolidation and some smaller low-cost carriers have expanded their networks, no evidence suggests that those competitive dynamics

---

[107] Since 2005, the following mergers resulted in three major legacy carriers remaining in the industry: US Airways-America West (2005), Delta-Northwest (2008), United-Continental (2010), and American-US Airways (2014). In addition, Southwest acquired Air Tran in 2011 and Alaska acquired Virgin America in 2016. Luo (2014), Shen (2017), Carlton et al. (2019), and Zhang and Nozick (2018) have assessed the effects on fares of the mergers between the legacy carriers and have found that they vary significantly, depending on the carriers involved and the type of route.

[108] Three new ULCCs have emerged, including Breeze, Avelo, and Porter, and Spirit consummated its plans to merge with JetBlue.

are likely to lead to adverse long-run changes, on average, in fares and service quality. However, because many established airlines have failed and because it has been difficult for new entrants to succeed, policymakers are concerned about the decline in the number of large carriers at the national level and about the long-run financial strength of low-cost carriers.[109] Recent research has added to those concerns by characterizing strategic behavior by airlines that could increase fares (for example, Aryal, Cilberto, and Leyden (forthcoming), and Sweeting, Roberts, and Gedge (2020)), and by pointing out that the largest investors in US airlines own stock in multiple airlines and that common airline ownership raises fares (Azar, Schmalz, and Tecu (2018)).

Historically, the US transportation industry has shown that the most effective way for policymakers to protect and enhance consumer welfare is to stimulate competition by eliminating entry barriers; but history also has shown that policymakers have applied the lesson selectively (Winston (2013)). Thus, in the airline industry, policymakers benefited travelers within the United States by deregulating domestic markets (Morrison and Winston (1986, 1995)), and benefited international travelers by negotiating open skies agreements with some countries (Winston and Yan (2015)).[110] Nevertheless, policymakers have allowed public airports to establish exclusive-use gates and slot controls that restrict air carrier entry and raise fares (Morrison and Winston (2000)), and, importantly, they have prevented foreign airlines from supplying additional competition in domestic markets by not giving them cabotage rights, which is the right to operate within the domestic borders of another country.

Beginning with the growth of Southwest Airlines, competition provided by low-cost carriers (LCCs) has accounted for a large fraction of the fare savings from deregulating entry into

---

[109] JetBlue, which started in 2000, is the most recent large airline in the United States that has not disappeared through merger or bankruptcy.

[110] Open skies agreements are an international policy concept that calls for liberalizing rules and regulations of international aviation to promote competition.

U.S. markets (Morrison and Winston (2000)). Similarly, the growth of LCCs Ryanair and EasyJet contributed to the benefits of deregulating entry into European Union markets (Alderghi et al. (2012)), and the emergence of LCCs, such as Norwegian Air, are likely to increase the benefits of liberalizing entry into transatlantic markets.[111] Accordingly, allowing foreign LCCs into U.S. markets has the potential to increase travelers' welfare significantly.

Federal law has continued to justify restricting competition in the airline industry by prohibiting foreign airlines from serving US domestic travelers because the US military might need immediate access to all aircraft during times of war, and because airline labor would mount significant political opposition to any further threats to its earnings. However, given the nation's increasing reliance on unmanned military aircraft and the weakened political position of airline labor in the wake of deregulation, those arguments have lost much of their force. Of course, this is not to say that U.S. airline carriers, labor, and the US Department of Defense would not oppose a change to current restrictions regarding cabotage and it is not clear if any foreign carriers want to expand their networks soon to serve domestic US markets.

In any case, it would be useful to quantify the potential benefits from allowing cabotage rights because they could provide a new source of airline competition that policymakers have overlooked. We are not aware of previous research that has attempted to estimate the effects of cabotage probably because there are no distinct airline markets where policymakers allow and do not allow cabotage. Our approach here is to develop a specialize case to estimate the potential

---

[111] The potential threat to competition represented by Norwegian Air was implied by a 2016 House bill H.R. 5090 that would have prevented Norwegian Air International (NAI) from serving US markets because NAI is based in Ireland and not in the home country, Norway, of its majority owner or owners. Norwegian Air, like many carriers, is currently experiencing financial stress caused by the global pandemic and faces an uncertain future. Recently, it announced that it would end its service to the United States and concentrate on service to Nordic countries. However, Norse Atlantic Airways (a Norwegian startup), Aer Lingus (the Irish flag carrier), and La Compagnie (an all-business French carrier) are planning to offer transatlantic service. Finally, the Iceland-based newcomer Play is picking up where the defunct Iceland discount airline Wow left off, offering service to more than twenty European destinations with plans to operate from US cities.

benefits to travelers of allowing a foreign airline to enter US routes. Specifically, we consider the entry of an LCC that currently serves the European Union because the US and EU LCCs have adopted similar business models, shared similar cost characteristics, and been leading competitors in their respective markets despite the efforts of legacy carriers.

We first provide an overview of the operations and entry patterns of Southwest, the leading US LCC, and Ryanair and EasyJet, the leading EU LCCs. We then obtain evidence about the effects of their actual, potential (Morrison and Winston (1987)), and adjacent (Morrison and Winston (2000)) entry on fares in their respective U.S. and EU markets. Adjacent competition has gained attention because of the competitive power of Southwest Airlines, which was able to reduce other airlines' fares on, for example, the route comprised of Washington Dulles International Airport and San Francisco International Airport by entering and charging low fares on the route comprised of Baltimore-Washington International Airport and Oakland International Airport.

Finally, we extrapolate from those findings to simulate the effect that the entry of an EU LCC would have on fares and travelers' welfare in U.S. markets. Of course, it is premature to assert that Ryanair or EasyJet would actually adjust their operations and undertake the necessary investments to serve U.S. domestic markets if they were granted cabotage rights. However, it is likely that some foreign carriers would eventually serve those markets if US policymakers gave them the opportunity to do so; conducting the analysis with an EU LCC is a useful first step to illustrate the relevant empirical issues and potential welfare effects.[112]

---

[112] The global pandemic has dramatically affected the airline industry as carriers have had to adjust to large variations in demand and sought government financial assistance. It is not clear how the pandemic will affect the industry in the long run. However, during the pandemic, Southwest, Ryanair, and EasyJet were the only investment-grade-rated carriers left in the industry. In fact, Southwest has expanded into additional airports. Generally, industry observers believe that low-cost and ultra-low-cost carriers will recover from the shock caused by the pandemic more easily than legacy carriers will recover.

In the process, our analysis contributes to the literature on the effects of LCC competition on fares by developing a methodology, difference-in-differences (DID) matching with a regression adjustment, which addresses the endogeneity of an LCC's entry as either an actual, potential, or adjacent competitor.[113] The standard DID approach used in previous research does not address fully the potential bias from endogenous entry. Our approach also is suitable for investigating the effects of firm entry in other industries, such as banking, petroleum, and chain stores, where markets are geographically isolated and where firms' initial networks drive their expansion.

We find that Southwest Airlines reduces fares in U.S. markets by as much as 30% and that Ryanair and EasyJet reduce fares in EU markets by as much as 20%. We also find that the conventional DID approach tends to understate Southwest's effect on fares and to overstate the EU LCCs' effect on fares. We simulate a stylized effect of cabotage by assuming that an EU LCC enters only US markets that are not served by a US LCC or ULCC, and that its effect on fares is comparable to its effect on fares in EU markets, although not as large as Southwest's effect on fares in US markets.   Given those assumptions and our estimated models, we find that travelers would gain a modest $1.6 billion annually because some 80% of domestic passengers in the United States are already flying on routes that are served by at least one LCC or ULCC.

However, it is likely that those initial gains to US travelers would expand greatly under a more comprehensive cabotage policy that allows entry of all foreign carriers in all countries, not just the entry by an EU LCC in the United States. Competition in a fully deregulated and integrated global airline market would encourage both foreign and US carriers to restructure their networks,

---

[113] We are not aware of previous research that has used this methodology.

which would result in greater entry on international routes and on domestic routes that feed those routes, leading to lower fares and improved service.[114]

The case for allowing cabotage is strengthened significantly as the framework evolves from the static network we consider here to a network restructured by new foreign entrants and domestic carriers that facilitates and encourages seamless international travel. As discussed in the appendix to this chapter, seamless international travel would significantly reduce the large disutility associated with long transfer wait times in a foreign airport, reduce total travel time for a connecting flight abroad, and improve the reliability of travel time—service improvements that are highly valued by international airline travelers.

Importantly, competition in a global airline industry that grants cabotage rights would provide extra protection for travelers against excessive fares on US domestic routes because those fares would induce entry by foreign carriers. In addition, excessive domestic fares may reduce overall airline profits because they could discourage travel on US international routes.

From the carriers' perspective, an expanded global market could reduce the impact of the periodic unanticipated shocks to the airline industry, most recently the global COVID-19 pandemic, by enabling carriers to diversify and expand their "portfolio" of airline routes by increasing operations in some countries and regions and decreasing them in others.

2. An Overview of LCC Expansions and Entry Patterns in the US and EU

Before estimating the effects on fares of the LCCs of interest, Southwest in the United States and Ryanair and EasyJet in the European Union, we discuss how the LCCs developed their

---

[114] When airlines introduce service on new routes, they can generate other benefits to the public. For example, Wang, Zheng, and Dai (2021) find that the introduction of a new airline route increases the number of shared kidneys by 7.3%.

operations and networks to become major competitors in their respective markets. We then present descriptive evidence of patterns of LCC entry behavior, which guides our decomposition of their effects on fares as actual, potential, and adjacent entrants.

Southwest's Expansion in the United States

Southwest Airlines began as an intrastate carrier in Texas and became an interstate carrier following deregulation in 1978, which allowed it to serve cities in other states. Southwest developed its business model as a low-cost carrier during the 1980s and captured national and even international attention for its efficient operations that some other carriers tried to emulate. For decades, Southwest has flown one type of aircraft, the Boeing 737, which reduces its maintenance and pilot training costs. In addition, it has avoided congested airports, or airports with high passenger facility charges, in favor of secondary, less crowded airports that have lower passenger facility charges and that enable Southwest to provide quick turnaround service and increase aircraft utilization (Boguslaski, Ito and Lee (2004)).[115] Southwest's operating costs per available seat-mile have consistently been much lower than the other large US airlines' operating costs per available seat-mile, with the exception of JetBlue, a more recent low-cost entrant.[116] Southwest's lower costs have enabled it to charge lower fares and put tremendous competitive pressure on other carriers.

In addition to its low fares, Southwest has grown far beyond its Texas roots (and routes) to serve markets throughout the United States. As shown in figure 1, Southwest expanded its network

---

[115] In recent years, Southwest has expanded into markets comprised of a congested airport, such as Miami, Chicago O'Hare, Houston Intercontinental, and Newark.

[116] According to Southwest's 10-K annual reports filled with the Securities and Exchange Commission, operating costs per available seat-mile have fluctuated in real 2000 dollars around 8 cents from 1993 to 2011, while the operating costs per available seat-mile of the other large US carriers have fluctuated around 10 cents. In contrast with other air carriers, since the early 1990s, Southwest has reduced its operating costs by hedging the cost of jet fuel. Southwest aims to hedge at least 50% of its fuel every year. In some years Southwest has reported hedging losses, but overall Southwest's hedging program has saved the carrier billions of dollars.

from 148 non-stop markets (501 markets with connections) in 1994 to 490 non-stop markets (2,228 markets with connections) in 2011. Most of those markets tend to consist of short- and medium-haul routes. Although Southwest has entered new routes over time, it has generally not exited many routes.

Ryanair's and EasyJet's Expansion in Europe

Ryanair, an Irish airline established in 1984, followed Southwest's low-cost model and expanded its network when the European Union fully deregulated air transportation in 1997. Similarly, EasyJet, a British airline established in 1995, adopted a low-cost business model and expanded its operations following EU deregulation. As in the case of Southwest, Ryanair and EasyJet attempt to avoid congested airports by using Europe's (underused) secondary airports, which helps them to achieve fast turnaround times, reduce their operating costs, and compete more effectively against other European carriers (Barrett (2004)). The supply of slots at major European hub airports has been significantly restricted by grandfather rights given to other European carriers; however, like Southwest, EasyJet has expanded to serve major (congested) airports as slots become available. Finally, Ryanair and EasyJet serve most of their routes year-round as well as several seasonal routes, primarily to Croatia, Greece, and Turkey.[117]

Within the EU, airlines can serve other member countries from their home country and can operate within other member countries; that is, as part of a protracted process that began in 1990, public authorities fully enacted cabotage rights in 1997. Figure 2 shows the number of routes served by Ryanair and EasyJet in the EU from January 2005 to December 2013. Because the two LCCs have expanded rapidly during this period, they have increased competition on many routes.

---

[117] Ryanair and EasyJet offer charter service. However, we confine our analysis to scheduled commercial, not private, airline service where we can clearly identify entry and exit, and we can estimate the causal effect of a carrier's presence on fares.

At the same time, their networks do not overlap much because they tend to avoid competing with each other.[118]

<u>An Overview of LCCs' Entry Patterns</u>

We investigate the entry patterns of the US and EU LCCs to indicate their ability to affect fares through actual, potential, and adjacent entry. Our empirical overview uses the US Department of Transportation's Airline Origin and Destination Survey (DB1B) and T-100 Domestic Segment Data from 1993 to 2011, and the International Air Transportation Association's (IATA's) monthly data on airline operations and fares from 2005 to 2013.[119] The routes in our analysis are non-directional airport pairs.[120] In our sample, there are 13,590 such routes in the United States and 3,588 in the European Union. Appendix tables A1 and A2 present the means, standard deviations, minimum, and maximum values of the variables used in this and subsequent empirical analyses in the paper.

Actual Entry. Actual entry (exit) occurs when an LCC served (did not serve) a route in a month (or quarter) but did not serve (served) that route in the previous month (or quarter).[121] Based on our data and assumptions, the EU LCCs tend to enter and exit routes frequently, as Ryanair made 500 entries on 377 routes and 211 exits on 150 routes, and EasyJet made 438 entries on 323 routes and 231 exits on 134 routes.[122] Some of the entries and exits reflect seasonal operations,

---

[118] Wizz Air, a new ultra-low-cost carrier, is expanding rapidly in European markets and is increasingly competing with Ryanair on some routes.

[119] We are grateful to John Byerly and Douglas Lavin for helping us to obtain the IATA data. Because our sample ends in 2011, we do not confound our analysis with the structural effects of Southwest's merger with another LCC, AirTran Airways, in 2011. Our analysis of the European Union market includes the United Kingdom and other European countries, such as Norway and Switzerland, which are not members of the EU

[120] Morrison (2001) points out that using airport pairs enables one to account for competition from adjacent airports and adjacent airport pairs; for example, the airport pair Baltimore-Washington International Airport to Oakland International Airport provides adjacent competition for the airport pair Washington Dulles International Airport to San Francisco International Airport.

[121] More precisely, we assume that an airline entered a route if it provided non-stop or one-stop connecting flights for at least 6 quarters and offered at least one flight every two days; that is, 45 flights per quarter.

[122] Ryanair's and, to a lesser extent, EasyJet's high route turnover is partially a control on airport costs and airport charges.

but Ryanair and EasyJet do compete with other carriers on their seasonal routes. Figure 3 shows that both EU LCCs exhibit strong time patterns of entry and exit by adjusting their networks more in April and November than they do in other months of the year; about 60% of Ryanair's entries and exits and about 50% of EasyJet's entries and exits are made during those months.

In contrast, Southwest Airlines seldom exits an airport or a city-pair market. In our sample, Southwest ceased operations at only Detroit (1993), San Francisco (2001), and Houston Intercontinental (2005) airports. Because a metropolitan area may be served by multiple airports, an airline's exit from an airport does not necessarily imply that it no longer serves the area. For example, after Southwest exited San Francisco International airport, it continued to serve the San Francisco Bay Area by serving Oakland International airport.

Potential and Adjacent Entry. Potential entry occurs when an LCC serves one (Type 1 potential entry) or both (Type 2 potential entry) of the end-point airports but does not serve the airport pair. Adjacent entry occurs when an LCC enters an adjacent competitive route comprised of two adjacent airports. We define two airports as adjacent airports if the distance between them is less than 100 kilometers or 62.5 miles.[123] To avoid double counting, if an airline enters a given route (actual entry) and then enters an adjacent route, we measure only the effect of actual entry on fares.

<div align="center">LCCs' Patterns of Route Entry</div>

We estimate a probit model to characterize patterns that are associated with the US and EU LCCs' entry. We express the conditional probability of an LCC entering a route as $\Pr(d_{ijt} = 1 \mid X_{jt}, Z_{it}, Z_{i't})$, where d_{ijt} is a binary indicator equal to 1 if LCC i entered route j in

---

[123] Our assumption is midway between the 75-mile radius used by Morrison (2001) and the 50-mile radius used by Dresner, Lin, and Windle (1996) to construct adjacent competitive routes. When we measured an airport's catchment area based on travel time instead of distance, our primary results did not change.

time t for the first time and 0 otherwise ; X_{jt} is a vector of market characteristics such as distance and market size; Z_{it} is a vector of variables measuring the LCC's network; and $Z_{i't}$ is a vector of variables measuring the competitors' i' networks at the time of entry.[124]

The estimation results presented in table 1 indicate how US and EU LCCs' actual entry behavior is associated with their potential and adjacent entry. The LCCs are more likely to enter a route when they already serve both end-point airports and, in the case of Ryanair and EasyJet, the likelihood increases as the number of routes connected to the end-point airports increase.[125] Actual entry is also more likely to occur if the LCCs serve one of the end-point airports, but in the case of EasyJet that effect is statistically significant only as the number of routes connected to the airport increases. Actual entry by the EU LCCs is less likely if they already provide adjacent competition, which may reflect the difficulty of entering certain routes connected to major EU hub airports. In contrast, actual entry by Southwest is more likely if Southwest already provides adjacent competition. Finally, consistent with our summary of their operations, an EU LCC is less likely to enter a route that the other EU LCC has entered.

3. Difference-in-Differences Estimation of the Effect of LCC Entry on Fares

We begin our exploration of the effect of LCC entry on a route's average fare by using the difference-in-differences (DID) approach taken by many researchers and performing the estimations using the DOT DB1B and IATA data sets.[126] The fares in both data sets do not include

---

[124] In this exercise, we exclude re-entries of an LCC on a route (for example, on seasonal routes) because the re-entry decisions could be affected by an LCC's previous entries and exits.

[125] Our finding on Southwest's actual entry behavior is consistent with work by Zou and Yu (2020).

[126] We focus on average fares because our main interest is to obtain an estimate of the potential benefits to travelers on US domestic routes from the entry of an EU LCC. That estimate is obtained by using the estimates of the average effects of Southwest and the EU LCCs on fares on the routes they serve in the US and EU, respectively, to bound the effect of an EU LCC on average fares on US routes.

ancillary charges, which in recent years have become an important source of industry revenues because carriers have unbundled many services, such as checked luggage and ticket reservations by phone, which were included in the ticket price, and they have begun to charge additional fees for them. However, this change is unlikely to bias our analysis much. [127]

The DID approach is a useful starting point because both the U.S. and EU samples include many instances of LCC entry and many routes that an LCC has never entered. However, we later discuss the concerns with the standard DID approach and we develop an improved approach.[128]

We define a route's product as the combination of a carrier, itinerary, and ticket class (first, business, economy full, economy discount and others).[129] We implement DID estimation of the average fare on a route for the following specification:

$$\ln(y_{it}) = \alpha_1 \text{LCCroute}_{it} + \alpha_2 \text{LCCadjacent}_{it} + \alpha_3 \text{LCConeairport}_{it} \times \text{LCCconnectivity}_{it}$$

$$+ \alpha_4 \text{LCCtwoairport}_{it} \times \text{LCCconnectivity}_{it}$$

$$+ \alpha_5 \text{LCConeairport}_{it} \times \text{LCCconnectivity}_{it} \times \text{noLCC}_i$$

$$+ \alpha_6 \text{LCCtwoairport}_{it} \times \text{LCCconnectivity}_{it} \times \text{noLCC}_i + \alpha_7 \ln(\text{Pax}_{it})$$

$$+ \alpha_8 \ln(\text{Ncarriers}_{it}) + \mu_y + \mu_m + \mu_i + \varepsilon_{it} \tag{1}$$

---

[127] Although fares in the DB1B data do not include ancillary charges, Southwest does not charge travelers for services, such as checked luggage for the first and second bag, changing or cancelling reservations, and so on. In addition, Southwest's charges for oversized and overweight bags are generally lower than legacy carriers' charges. Thus, if anything, our analysis may understate Southwest's effect on the full costs to air travelers because it does not include its effect on the price of other optional services. Ryanair and EasyJet do unbundle services and charge for them. However, our IATA sample ended before the share of revenues that those carriers derived from ancillary services significantly increased. It is unlikely that we are understating the effect of the EU LCCs on the full costs to travelers because most full-service European carriers unbundled their fares and derived significant ancillary revenues after the end of our IATA sample.

[128] Previous studies of the effects of Southwest's entry on fares in US markets include Windle and Dresner (1995), Dresner, Lin, and Windle (1996), Morrison and Winston (2000), Morrison (2001), Goolsbee and Syverson (2008), and Brueckner, Lee, and Singer (2013).

[129] We use the fare code information in the IATA data to determine the fare classification for European carriers. However, we cannot use the fare code information in DB1B because each U.S. airline has its own definition of its fare codes. Thus, following common practice, we define the fare class based on the range of fares.

where,

y_{it}: Average fare on route i at time t;

$LCCroute_{it}$: A dummy indicating that route i is served by one of the LCCs at time t;

$LCCadjacent_{it}$: A dummy indicating that one of route i's adjacent routes but not route i is served by an LCC at time t;

$LCConeairport_{it}$: A dummy indicating that one of the endpoint airports of route i but not route i is served by an LCC at time t;

$LCCtwoairport_{it}$: A dummy indicating that both endpoint airports of route i but not route i are served by an LCC at time t;

$LCCconnectivity_{it}$: Number of routes that an LCC serves that are connected to the endpoint airports of route i during period t; in the EU case, if both LCCs operate at the endpoint airports, we use the larger number of routes that they serve;

$noLCC_i$: A dummy indicating that an LCC has never entered route i during the sampling period;

Pax_{it}: Number of passengers on route i at time t;

$Ncarriers_{it}$: Number of carriers serving route i at time t;

μ_y, μ_m, and μ_i: year, month/quarter, and route fixed effects;

ε_{it}: the econometric error term.

We present the DID parameter estimates of the effects of LCC entry on average fares in table 2. We use the geometric mean of the population of the end-point cities as an instrumental variable for the number of passengers, which is endogenous.[130] Note that population is correlated with passengers, but holding passengers constant, it is not correlated with fares. In addition,

---

[130] We address the endogeneity of airline presence on a route later.

population varies across routes and over time. The first two columns of the table present OLS and IV estimates for the EU LCCs and the third and fourth columns present OLS and IV estimates for Southwest.[131]

The IV parameter estimates indicate that the LCCs' actual entry has (approximately) reduced fares, on average, by 39% in the EU markets and by 26% in the U.S. markets. The estimates also indicate that the EU LCCs reduce fares when they are adjacent competitors, but that Southwest slightly increases fares when it is an adjacent competitor. Finally, the LCCs have similar effects on fares when they are Type 2 potential competitors that serve both airports but not the route, but Southwest has a larger effect on fares than Ryanair and EasyJet have when the LCCs are Type 1 potential competitors that serve one airport comprising the route. In sum, the LCCs have marked impacts on fares in their respective markets, with the EU carriers reducing fares by more than Southwest reduces fares when we account for actual, potential, and adjacent entry.

At first blush, the findings are plausible, and they are broadly consistent with previous estimates of Southwest's effects on fares. We are not aware of previous DID estimates of the effect of the EU LCCs on fares; thus, we cannot assess our comparative findings in the context of previous research. However, we have concerns about the conventional DID approach that raise doubts about the accuracy of our initial estimated effects of LCC entry.

Identification of the parameters relies on the assumption that the route and time dummy variables capture the unobserved route-specific factors and time effects that are correlated with the regressors. This assumption is questionable because both samples contain many routes that capture entry at different points in time over several years in different market environments. Unobserved factors affecting market outcomes are unlikely to be captured by the route fixed effects

---

[131] Although we present the average effects of Ryanair and EasyJet on fares in the table, their individual effects on fares were similar.

because they are unlikely to remain constant over time. In addition, the timing of an LCC's entry on different routes is affected mainly by the structure of its initial network; thus, we cannot simply compare the outcomes of markets entered by an LCC with the outcomes of markets not entered by an LCC across different years because many other factors also affect market outcomes. Finally, the DID approach does not separate the effects of actual entry, potential entry, and adjacent competition on fares because those actions may occur sequentially on a route. For example, an LCC's actual entry may occur only several months after its potential entry. If we have many routes where that occurs, identification of the coefficients for the actual and potential entry dummy variables would be contaminated.

4. An Improved Approach: DID Matching with a Regression Adjustment

We take advantage of our large panel data samples to design a new approach to address the identification problems with the standard DID approach. We use a DID matching approach to compare fares on treated routes (entered by an LCC) with fares on control routes (not entered by an LCC) in the same time window. The approach compares the difference in fares on routes entered by an LCC with the difference in fares on matched routes without LCC entry both before and after LCC entry on the treated routes. Importantly, we also separate the effects of LCC actual, potential, and adjacent entry on route fares by first quantifying the effect of actual entry conditional on potential entry and then quantifying the effect of potential entry, excluding the effect of an LCC's adjacent entry. Finally, we quantify the effect of adjacent entry, excluding the effects of an LCC's actual and potential entry.

We match a treated route to multiple control routes by satisfying a matching criterion that we explain in the next section. For a treated route i, y_i^{pre} and y_i^{post} denote the average fare before

and after the LCC's entry, respectively.[132] For a matched route i', $y_{i'}^{pre}$ and $y_{i'}^{post}$ denote the average fare before and after the LCC's entry on the treated route, respectively. The non-parametric DID comparison on a matched pair is given by $\tau_{ii'}$, which measures the net change rate of the route average fare caused by an LCC's entry:

$$\tau_{ii'} = \underbrace{\left(\frac{y_i^{post} - y_i^{pre}}{y_i^{pre}}\right)}_{\substack{\text{change rate of the} \\ \text{average fare on a} \\ \text{treated route}}} - \underbrace{\left(\frac{y_{i'}^{post} - y_{i'}^{pre}}{y_{i'}^{pre}}\right)}_{\substack{\text{change rate of the} \\ \text{average fare on a} \\ \text{matched route}}} \tag{2}$$

The first term on the right-hand side is the change rate of the average fare on a treated route, and the second term is the change rate of the average fare on a matched route, which captures the time trend of the fare change in the counterfactual scenario where LCC entry does not occur.

It is, of course, possible that a treated route and the matched control routes have different characteristics that are correlated with the average route fare. We therefore construct variables, such as the number of carriers, number of passengers, and population and income at the end-point airports, to control for the difference between the treated and control routes.[133] Given our computation of the net change rate of the route average fare caused by an LCC's entry in equation (2), we remove the effects of the control variables on the computed outcome.

Formally, we denote $\Delta X_{ii'} = \left(X_i^{post} - X_i^{pre}\right) - \left(X_{i'}^{post} - X_{i'}^{pre}\right)$ as the vector of the DID control variables on a treated route i and a control route i' and we run the following regression without a constant term:

$$\tau_{ii'} = \Delta X_{ii'} B + \varepsilon_{ii'} \tag{3}$$

---

[132] We average the route average fare over the quarters (or months) pre and post entry.
[133] Intercity rail competition is not an important issue in US markets. In European markets, our matching algorithm accounts for rail competition as long as the LCC entries followed similar decision rules over time. The Channel tunnel rail link connecting London and Paris has had an impact on air passenger volumes.

where B is a vector of parameters and $\varepsilon_{ii'}$ is the error term. If the DID of the route average fare is completely determined by the DID of the control variables, then $\varepsilon_{ii'} \equiv \tau_{ii'} - \hat{\tau}_{ii'}$—where $\hat{\tau}_{ii'}$ is the predicted value from the regression equation (3)—would be expected to be close to zero. We therefore interpret a non-zero $\hat{\varepsilon}_{ii'}$ as the net effect of the LCC's entry on the route average fare.

In sum, we estimate the average treatment effect of an LCC's entry on a route's average fare, excluding the effects of observed market characteristics, by calculating:

$$\delta = \frac{1}{N} \sum_{i \in \Psi} \frac{1}{M_i} \sum_{i' \in \Gamma_i} \hat{\varepsilon}_{ii'} \tag{4}$$

where Γ_i is the set of matched routes to the treated route, M_i is the number of routes in the set, Ψ is the set of treated routes, and N is the total number of matched pairs. We use the bootstrap technique to construct the confidence interval of the estimator in equation (4) by randomly sampling the matched pairs (i, i') with replacement to obtain a bootstrapped sample with the same size as the original sample of matched pairs. We then compute δ using the bootstrapped sample and we repeat the process 100 times to obtain the empirical distribution of δ over the bootstrapped samples from which we construct the confidence interval of the estimator.

A final concern with our methodology is that unobserved market characteristics that are correlated with route fares may influence an LCC to enter a route. We address this concern by restricting the set of matched control routes for a treated route to those that the LCC entered at least two years later.[134] We therefore assume that the timing of an LCC's entry on different routes is driven mainly by the structure of its initial network and not by route characteristics. In the case of Uber, this is analogous to assuming that its entry into new markets is negatively related to the

---

[134] Our basic findings were not affected when we conducted sensitivity analyses of the two-year threshold.

market's distance from Uber's headquarters in San Francisco and New York City. In our context, the assumption is plausible because the LCCs' networks have primarily evolved from their initial headquarters in Texas, England, and Ireland. In addition, as shown in table 1, an LCC tends to enter a route only when it already serves at least one of the end-point airports, and it is more likely to enter a route when it serves both endpoint airports.

5. Criteria for Matching Control Routes to Treated Routes

We structure our analysis to characterize the pre-entry and post-entry periods and to distinguish between the different types of LCC entry. We define the pre-entry period as 4 to 12 months before actual route entry and we define the post-entry period as 18 months after actual entry. Within the post-entry period, we define 0 to 6 months after entry as the short-run, 7 to 12 months as the medium-run, and 13 to 18 months as the long run. We summarize the timeline for a treated route in exhibit 1 below. Because potential entry occurs at least 18 months before actual entry, it is likely to be unrelated to the eventual market fare when actual entry occurs. We exclude the 3-month period before actual entry because a time lag may exist between an LCC's announcement that it plans to serve a route and when it does serve the route. [135]

Exhibit 1. Timeline (in Months) of an LCC's Actual Entry on a Treated Route

The LCC serves at least one of the end-point airports at least 18 months before actual entry and this status remains unchanged before entry.	Pre-entry period	Post entry short-run effect	Post entry medium-run effect	Post entry long-run effect

-18   -12   -3   0   6   12   18

---
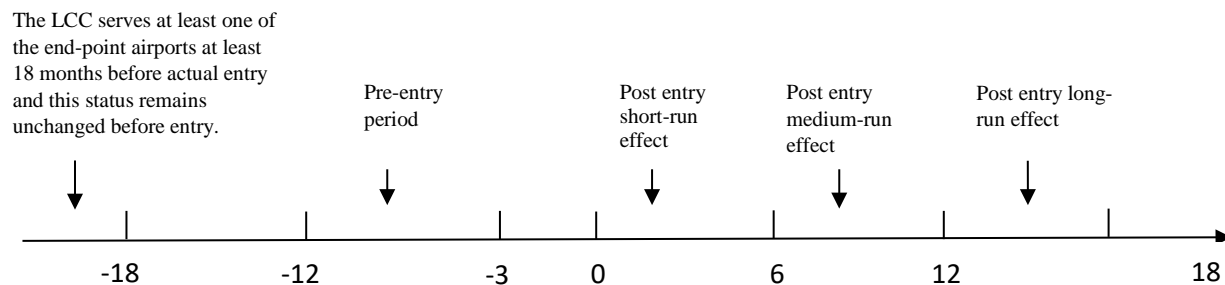
[135] We observed that the time gap between an entry announcement and actual entry was, on average, more than one quarter. Thus, in the US case, we excluded 2 quarters before Southwest's actual entry in a market.

Exhibit 2 shows the timeline for determining the set of control routes that match the treated route. The LCC serves at least one of the endpoint airports on a control route at least 18 months before actual entry on the treated route. However, its actual entry on a control route occurs at least 24 months after its actual entry on the treated route, thereby avoiding any effect its entry on a control route may have on its long-run effect on the treated route.

Exhibit 2. Timeline (in months) for Defining a Control Route and a Treated Route for Routes Entered by a Given LCC

| Potential entry of the LCC on the control route at least 18 months before its actual entry on the treated route | Pre-entry period in DID | Actual entry on the treated route | Short-run post-entry period in DID | Medium-run post-entry period in DID | Long-run post-entry period in DID | Actual entry on the control route is at least 24 months after actual entry on the treated route |

-18   -12   -3   0   6   12   18   24

Selection of the Sample of Routes

We now summarize our assumptions to select treated routes and control routes to estimate the distinct effects of the LCC's actual, potential, and adjacent entry on average fares. We construct four modules:

— The effect of Southwest's actual entry without potential entry, excluding the effect of its adjacent entry—this module is specific to Southwest Airlines because it arises in its US domestic route network;

— The effect of an LCC's actual entry conditional on both types of an LCC's potential entry, excluding the effect of an LCC's adjacent entry;

— The effect of both types of an LCC's potential entry, excluding the effects of its actual and adjacent entry; and

— The effect of an LCC's adjacent entry, excluding the effect of its actual and potential entry.

In table 3, we list the conditions for a given module that a route must satisfy to be a treated route and that it must satisfy to be a control route. The conditions are aligned with the timelines that we presented in exhibits 1 and 2. We also indicate certain robustness checks that we conducted.

<div align="center">Balancing Test</div>

The preceding conditions to define treated and control routes assume that the *timing* of an LCC's entry into different routes is exogenous. We test this assumption formally with a balance test proposed by Rosenbaum and Rubin (1985) by testing whether routes entered by an LCC in earlier years were like those entered by the LCC in later years.

The test statistic, $B(X)$, is based on the standardized difference:

$$B(X) = 100 \cdot \frac{\bar{X}_F - \bar{X}_M}{\sqrt{\dfrac{S_F^2(X) + S_M^2(X)}{2}}} \tag{5}$$

where for each covariate, $\bar{X}_F$ and $\bar{X}_M$ are the sample means for the full (F) affected routes (i.e., all the treated routes in our sample) and the matched (M) treated routes (i.e., the routes entered later by an LCC that are compared with each treated route); and $S_F^2(X)$ and $S_M^2(X)$ are the corresponding sample variances. Intuitively, $B(X)$ measures the difference in the means of a conditioning variable, scaled by the square root of the variances in the samples. Rosenbaum and Rubin (1985) define imbalance for a covariate as exceeding 20 in absolute value.

We found that the standardized differences in both the US and EU samples of the routes' average fares, distances, endpoint populations, and endpoint incomes per capita had absolute values less than 20, which indicates that those route characteristics pass the balancing test. Other

route characteristics, namely the number of carriers and Herfindahl-Hirschman Index (HHI) of the city-pair market, had standardized differences that exceeded 20 in some modules. However, we can use our regression adjustment to remove the effect of that imbalance on the results.

6. Estimation Results

Our approach identifies the distinct effects of LCCs' actual, potential, and adjacent entry on average fares in the short, medium, and long run; thus, we report the estimation results for each type of entry in separate tables. As noted, Southwest entered some routes without prior potential entry, but our data indicated that the EU LCCs entered routes only with prior potential entry. Southwest's entry without prior potential entry, shown in table 4, reduces fares, on average, roughly 10% in the short run and nearly 14% in the long run. The second column of the table shows that the regression adjustment is important because we would have significantly over-estimated the effect of Southwest's entry on fares without it. Hereafter, we report results only with the regression adjustment. The third and fourth columns of the table report findings for entry only on connecting routes and show that they are very similar to the findings for entry on all routes.

Both Southwest and the EU LCCs entered routes conditional on potential entry. As shown in table 5, their entry has different effects on fares, which suggests that the EU and US incumbent airlines respond differently to an LCC's actual entry. Entry by an EU LCC, conditional on its potential entry at one or both end-point airports, reduces fares roughly 15% in the short and medium run and 10% in the long run, possibly because an LCC cuts prices substantially in the short and medium run to capture market share, but gradually adjusts prices after the initial entry shock. In contrast, the last two columns of the table show that Southwest's entry, conditional on it being a potential competition at one endpoint airport, reduces fares by nearly 18% in the short

run and nearly 19% in the long run.[136] However, its effect on fares is much less, roughly 4%, when it is a potential competitor at both end-point airports and in all likelihood has a strong effect on fares before it actually enters a market.

As expected, we find in table 6 that when an LCC is a potential competitor, it reduces fares more if it serves both endpoint airports (Type 2) than if it serves one of the endpoint airports. Generally, Southwest has a much larger effect for this type of entry than the EU LCCs have. In fact, Southwest's effect as a Type 1 potential competitor (column 3) is somewhat greater than the EU LCC's effect as a Type 2 potential competitor.

We previously speculated that Southwest's effect as an actual competitor is less when it is a Type 2 potential competitor than when it is a Type 1 potential competitor because it already has a strong effect on fares when it provides Type 2 potential competition. Column 4 of table 6 is consistent with this speculation because we find that when Southwest is a Type 2 potential competitor, it reduces fares by roughly 8% to nearly 10% in the short and medium run and by more than 7% in the long run. Finally, we tested whether Southwest's effect as a potential competitor is consistent with its behavior, noted previously, of primarily entering short and medium haul markets. We found that Southwest reduced fares as a Type 1 potential competitor by 8% on routes less than 500 miles and that this effect declined with distance, becoming negligible on routes greater than 1500 miles because incumbent carriers were presumably less concerned that Southwest would eventually enter those routes.

Southwest also has a larger effect on fares as an adjacent competitor than the EU LCCs have, especially in the long run. We show in table 7 that adjacent competition provided by all the

---

[136] We found that the fare effect of Southwest's actual entry conditional on being a Type I potential competitor was somewhat greater when it entered routes that were more concentrated and when it entered hub routes.

LCCs reduces fares roughly 3% in the short and medium run; however, in the long run, Southwest reduces fares 5% and the EU LCCs reduce fares by slightly more than 1%.

In sum, we have developed a methodological approach, DID matching with a regression adjustment, which enables us to provide a more detailed and more accurate empirical assessment of the effect of LCC competition on fares than the conventional DID approach does. Importantly, we can identify distinct effects of actual, potential, and adjacent competition on fares and to show variations in those effects for Southwest and the EU LCCs. Our major findings include:

— Depending on the type of potential entry that an LCC provides, Southwest's actual entry in a market reduces fares more than when an EU LCC enters a market.

— When Southwest is a potential or adjacent competitor, it reduces fares notably more than when an EU LCC is a potential or adjacent competitor.

— Considering the effect of all types of entry, Southwest reduces fares, on average, by as much as 30%, with potential and adjacent entry accounting for a non-trivial amount of the decline in fares.

— Considering the effect of all types of entry, an EU LCC reduces fares, on average, by as much as 20%, with actual entry accounting for most of the decline in fares.

— Competition provided by an LCC appears to be more intense in U.S. markets than in EU markets, likely because entry into U.S. airports is less constrained by slot controls and by the limited availability of gates.[137]

---

[137] For example, in 2016, of the 180 slot-controlled airports in the world, 93 were in the EU and 2 in the U.S. See: http://www.europarl.europa.eu/RegData/etudes/IDAN/2016/585873/IPOL_IDA(2016)585873_EN.pdf
In the EU, slots are allocated based on the "grandfather rights" rule, which tends to benefit incumbent, mainly national, airlines by restricting the slots available to potential new entrants (Barrett (2004)).

— Compared with our main findings, the estimates that we obtained using the traditional DID approach overestimate the full effect of EU LCCs on fares and underestimate the full effect of Southwest on fares, which indicates that the methodological improvements provided by our approach are important for assessing the effect on fares of allowing cabotage and for other applications.

7. Additional Robustness Checks

In addition to Southwest, other LCCs and ULCCs serve U.S. markets, including Allegiant Air, Frontier Airlines, JetBlue, Spirit Airlines, Sun Country Airlines, and Virgin America (now part of Alaska Airlines), which could affect fares and potentially bias our estimates of Southwest's effect on fares if they provide service on the control routes. As a robustness check, we performed calculations where we considered only those routes where the preceding airlines did not serve the treated or the control routes for at least 8 quarters after Southwest actually entered the treated routes. We present the results in table 8.[138] As expected, we find that without the presence of other LCCs and ULCCs as actual competitors, the effect of Southwest's entry on average fares under alternative states of potential entry is greater, but the quantitative effect is not much greater than the effect in the base case, which can be interpreted a lower bound.

Ryanair and EasyJet account for most of the traffic carried by LCCs in the European Union during our sample. Given that Southwest's effect on fares was modestly affected when we controlled for other LCCs and ULCCs in the US, Ryanair's and EasyJet's effect on fares is likely to be even less affected if we controlled for other LCCs on EU routes.

---

[138] Because it is difficult to exclude all routes where other LCCs and ULCCs were potential competitors, we focused on the effect of Southwest's actual entry.

Because LCCs are growing, the number of routes that connect to their endpoint airports varies over time; greater connectivity increases the probability that an LCC will enter a route. In response, incumbent carriers may employ different pricing strategies as the probability of an LCC's entry changes with its connectivity. As a robustness check, we reevaluated the effect of Southwest's potential entry controlling for the number of routes that it served from its endpoint airports; however, we found that Southwest's effect on fares as a potential competitor increased only slightly.

8. Travelers' Potential Gains from Cabotage

We have found that Southwest's actual, potential, and adjacent entry has reduced fares, on average, 30% in U.S. markets. Although Southwest does not serve every U.S. route, especially long-haul routes with highly congested airports as endpoints, it does serve many short-haul and medium-haul routes. Without Southwest's entry, those routes are more likely than long-haul routes to have elevated fares because they are more likely to be underserved.

Southwest's entry has also generated significant gains in consumer surplus. To quantify those gains, we assume a simple constant elasticity demand function:

$$Q = aP^e \tag{6}$$

where Q is the demand for air travel, a is a positive constant, P is the average fare, and e is the constant elasticity of demand. We assume a route elasticity of demand of –1.4 based on Smyth and Pearce's (2008) analysis of US city-pair routes and we derive a as the mean of $Q/P^{\hat{e}}$ for each route.[139] We then calculate consumer surplus using the following expression:

---

[139] We do not distinguish between short-haul and long-haul markets here because Smyth and Pearce (2008) obtained very similar elasticity estimates for both types of markets.

$$\Delta CS = -\sum_t \Delta CS_t = -\sum_t \sum_r \int_{p_c}^{p_0} \hat{a}_r p_{rt}^{\hat{e}} \, dp \tag{7}$$

where p_c is the counterfactual average fare following Southwest's entry in category c (actual, potential, or adjacent), p_0 is the average fare before Southwest's entry, and r is the route affected by Southwest's entry.[140] We appropriately account for the sequential pattern of Southwest's entry as an actual, potential, and adjacent competitor, and we find that travelers gained roughly $5 billion from Southwest's entry in US markets during the last year of our sample, 2014, and gained $67.6 billion (2000 dollars) from its entry during the entire sample period, 1994–2014.

The recent motivation for a policy that grants cabotage rights to foreign airlines has arisen from concerns that the reduction in the number of airlines at the national level following several mergers has reduced competition. Given our findings about the large effects of Southwest, an LLC, on fares and travelers' welfare, the entry of another effective LCC might be desirable to address those concerns.

We assume that Ryanair or EasyJet would take advantage of the opportunity to serve U.S. domestic markets and we simulate the effect that such entry could have on fares. This simulation captures to a notable extent the effect that a Southwest-type foreign entrant would have on fares for three reasons. First, the EU LCCs' business plan was modeled after Southwest's business plan; thus, they primarily serve short-haul to medium-haul markets and tend to avoid highly congested airports. Second, like Southwest, the EU LCCs' costs per seat-mile are notably lower than their competitors' costs per seat-mile. Third, EU LCCs' entry has reduced fares significantly despite airport entry barriers that limit competition.

---

[140] Because the demand elasticity has a minus sign, we include a minus sign in the equation (7) so consumer surplus is positive if the average fare following Southwest's entry is lower than the average fare before Southwest's entry.

Because the EU LCCs, as well as any other foreign carrier, do not serve U.S. domestic routes, we make some plausible assumptions to execute the simulation. First, we assume that an EU LCC would not enter routes that Southwest or another US LCC or ULCC currently serve. This assumption is consistent with Ryanair and EasyJet's tendency to avoid competing on the same routes. Second, we assume that an EU LCC's entry on US routes would not reduce fares by as much as Southwest's entry has reduced fares, but we assume its entry would reduce fares more than its entry in EU markets has reduced fares because it would be less constrained by airport entry barriers. Thus, we assume that an EU LCC's entry would reduce fares 25%, on average, in U.S. markets.

Under those assumptions, we find that an EU LCC's entry in US markets would generate a modest, but non-trivial, $1.6 billion annual gain to travelers. Our qualitative finding would not change if we used the upper and lower bound assumptions of -2.5 and -0.7 for the elasticity of demand and if we assumed that an EU LCC's entry reduced fares by, say 5%, more or less than 25%.

The new entrant would address a potential gap in airline competition by entering city-pair markets that in our sample have, on average, 2.58 carriers, compared with the 3.94 carriers, on average, that serve the city-pair markets that LCC carriers currently serve. At the same time, the welfare gain is modest because competition on U.S. routes is generally very intense, as indicated by the fact that Southwest and the other LCCs and ULCCs serve routes that account for 80% of U.S. domestic passengers. Policymakers should be reassured that the state of airline competition is healthy, but they should not be diverted from the possibility that travelers could still benefit significantly if foreign airlines were granted cabotage rights.

Cabotage would usher in a new era of airline competition, which would affect domestic and international markets, because airlines would focus on providing seamless service in a global market. For example, companies such as Google and airlines such as Emirates, are envisioning seamless travel/lifestyle experiences, which rely on one airline to transport passengers from their domestic origins, possibly in low-density markets, to their foreign destinations, also possibly in low-density markets, with customized amenities.

Travelers' value of potential service improvements indicates that seamless travel could generate significant welfare gains to travelers. For example, Yan and Winston (2014) estimate that travelers' willingness to pay for non-stop flights to avoid connecting flights exceeds $200 per hour. The high valuation reflects passengers' disutility associated with transfer wait time in US airports. American travelers would likely have an even greater value of transfer wait time at foreign airports because they may fear that they could miss critical announcements about their connecting flight, and they may be less comfortable waiting for a flight in an unfamiliar environment. Total transfer wait time for international trips also is long, typically more than three hours. So, seamless travel that tightens connections and reduces the disutility associated with transfer wait time at a foreign airport could generate several billions of dollars in benefits to airline travelers. A rough calculation that provides supporting evidence for this claim is provided in the appendix to this chapter.

Generally, all carriers would have the incentive and ability to restructure their domestic and international networks to compete effectively in a deregulated global airline market. As in the case of domestic deregulation and open skies, travelers would then accrue gains from lower fares and improved service on international routes and on domestic routes of all hub classifications that could help feed international routes. Policymakers' concerns about competition also would be

addressed because the airlines that comprise the deregulated global air transportation industry would be unlikely to "leave any domestic route behind."[141]

9. Conclusions

Given the absence of foreign airlines that serve US domestic routes, we have argued that LCC competition in the US and EU is a useful starting point to explore the potential effects of cabotage because the LCCs have adopted a similar business model, shared similar cost characteristics, and been leading competitors in their respective markets. After estimating the fare effects of actual, potential, and adjacent entry by Southwest, the leading US LCC, and by Ryanair and EasyJet, the leading EU LCCs, we have extrapolated from those estimates to simulate the effect of entry by an EU LCC in U.S. markets.

We designed a new methodological approach, DID matching with a regression adjustment, to estimate the fare effects of LCC entry, which addresses the bias from endogenous entry that is not addressed by the conventional DID approach used in previous research to estimate the fare effects of Southwest's entry. The potential bias is important because we found that the conventional DID approach underestimates the fare effect of Southwest's entry and overestimates the fare effect of the EU LCCs' entry. In general, our methodology could overcome the bias from using the conventional DID approach to analyze the effects of entry in other industries, which are characterized by markets that are geographically isolated and firms with an initial network that drives their expansion.

---

[141] We noted that Azar, Schmalz, and Tecu (2018) have argued that the largest investors in US airlines own stock in multiple airlines and that common ownership raises fares. If so, entry by foreign airlines may also reduce fares by diluting the extent of common ownership of competing carriers. Currently, U.S. law prevents foreign companies and individuals from owning more than 25% of a US airline, but policymakers have not considered ownership restrictions on companies and individuals investing in a foreign airline operating in the United States. Thus, if cabotage rights were granted, it is not clear whether the largest investors in US airlines could or would also attempt to own foreign airlines.

We found that Southwest's entry reduces fares 30% in US markets, with potential and adjacent entry having important effects, and that the EU LCCs reduce fares 20% in EU markets, with actual entry accounting for most of the effect. To the best of our knowledge, the latter is the first estimate of the effect of the EU LCCs' entry on fares. We then found from our simulation that an EU LCC's entry in US markets would generate a modest $1.6 billion annual welfare gain to travelers.

We have acknowledged that we do not know whether an EU LCC or other foreign carriers would enter US markets if the US government granted cabotage rights to foreign airlines. However, we are confident that some U.S. airlines would enter European markets if European governments reciprocated and granted those airlines cabotage rights; thus, it is reasonable to believe that, in practice, cabotage would eventually lead to foreign entry in US markets that would benefit travelers.

We also have qualified our finding because it captures the effect of only one carrier's entry in US markets and does not account for the change in the global airline network, including competition from other carriers on domestic and international routes. We argued that changes in the global network would be the largest source of gains from cabotage because carriers would seek to provide seamless air travel throughout the world. We therefore conclude that the potential gains from allowing cabotage are likely to be substantial and that the influx of new entry should finally assuage policymakers' fears of insufficient airline competition in US markets. At the same time, the arguments against cabotage have greatly weakened.

Hopefully, this chapter is the beginning of a strong empirical case for cabotage, which policymakers throughout the world will consider and will eventually act on to benefit the world's airline travelers by deregulating the global airline industry.

Figure 1. Southwest's Expansion in the United States

Southwest Route Maps: 1993-2010
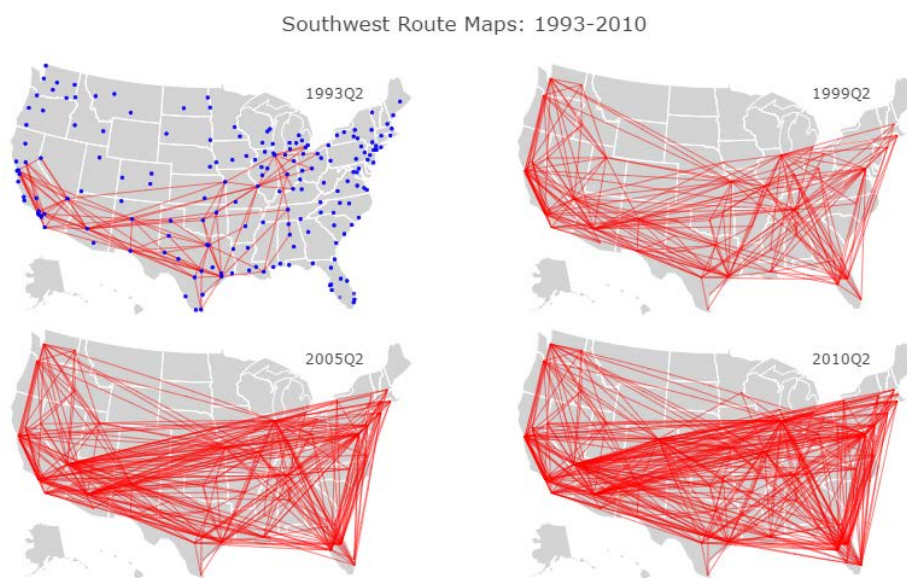
Notes: Blue dots and red lines denote airports and non-stop routes, respectively. Data Sources are the Origin and Destination Survey (DB1B) and the T-100 Domestic Segment Data (T100) from 1993 to 2011.

Figure 2. Expansion of Ryanair and EasyJet from January 2005 to December 2013
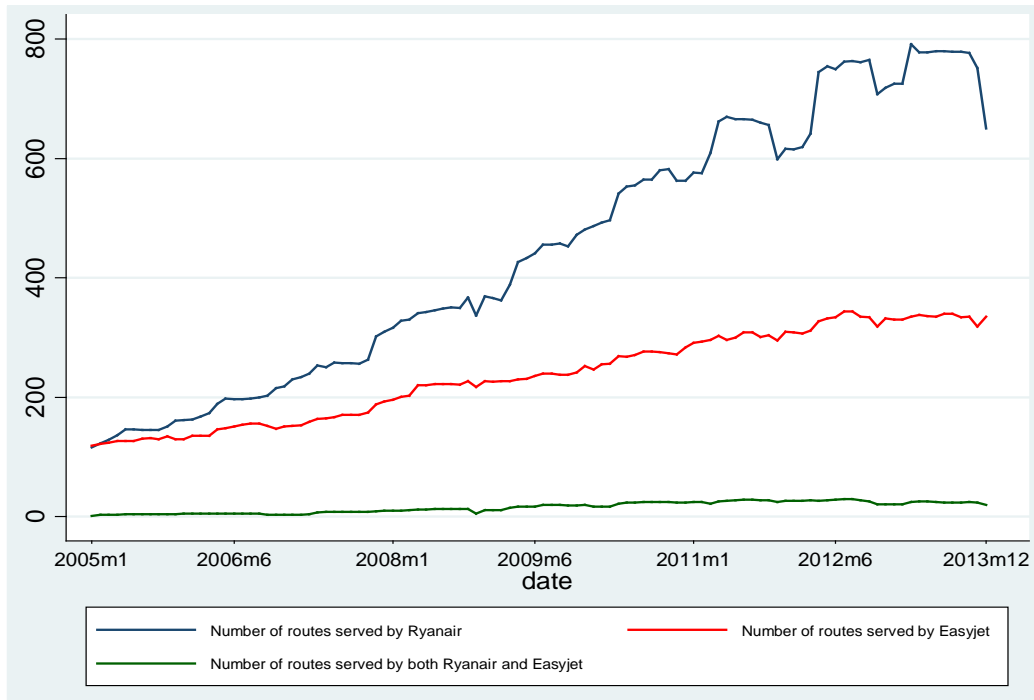
Source: International Air Transportation Association (IATA).

Figure 3. Monthly Entries and Exits by Ryanair and EasyJet, 20–2013

Table 1. Spatial Entry Patterns of LCCs from Probit Regressions (Dependent Variable: Dummy for First-Time Route Entry)

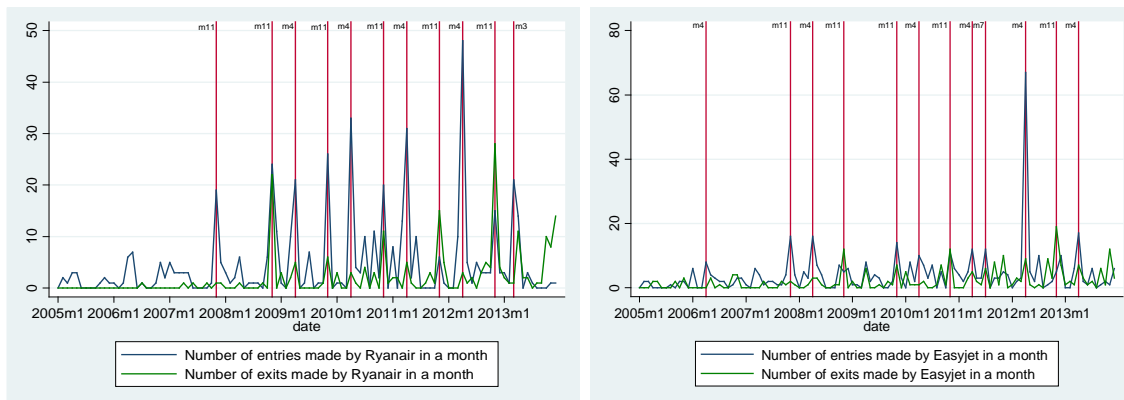| Variables	Note	Ryanair (1)	EasyJet (2)	Southwest (3)
Constant		–3.2791	–3.9192	–4.8442
		(0.2175)	(0.2147)	(0.1435)
Geometric mean of population of end-point catchment areas	The value of current period	–0.0228	0.0129	–0.0067
		(0.0150)	(0.0132)	(0.0088)
Geometric mean of income-per-capita of end-point catchment areas	The value of current period	–0.0817	0.0481	0.0095
		(0.0245)	(0.0248)	(0.0030)
Route distance		0.0910	0.0337	–0.0115
		(0.2886)	(0.0346)	(0.0163)
Percentage of first and business passengers in the regional market	The average value in previous year	0.0117	–0.0599	--
		(0.0928)	(0.0187)	
Dummy of airport presence at one of the end-point airports	The status of previous period	0.6136	0.1450	1.1388
		(0.1801)	(0.1625)	(0.1027)
Dummy of airport presence at one of the end-point airports × Number of served routes connecting the airport	The value of previous period	0.0198	0.0259	–0.0015
		(0.0015)	(0.0023)	(0.0009)
Dummy of airport presence at both end-point airports	The status of previous period	1.6303	0.7785	2.1022
		(0.1760)	(0.1558)	(0.1078)
Dummy of airport presence at both end-point airports × Number of served routes connecting the two airports	The value of previous period	0.0032	0.0149	–0.0136
		(0.0013)	(0.0019)	(0.0009)
Dummy of adjacent route presence	The status of previous period.	–0.7568	–0.8550	0.3298
		(0.0596)	(0.0879)	(0.0238)
Dummy of route presence of the other LCC	The status of previous period	–0.1129	–0.3585	--
		(0.1035)	(0.1738)	
Dummy of adjacent route presence of the other LCC	The status of previous period	–0.0113	–0.2159	--
		(0.0804)	(0.0848)	
Dummy of airport presence of the other LCC at one of the end-point airports	The status of previous period	0.2355	0.1774	--
		(0.0747)	(0.0631)	
Dummy of airport presence of the other LCC at one of the end-point airports × Number of served routes connecting the airport by the other LCC	The value of previous period	–0.0061	–0.0141	--
		(0.0041)	(0.0038)	
Dummy of airport presence of the other LCC at both end-point airports	The status of previous period	0.2619	0.1973	--
		(0.0873)	(0.1034)	
Dummy of airport presence of the other LCC at both end-point airports × Number of served routes connecting the two airports by the other LCC	The value of previous period	–0.0029	–0.0073	--
		(0.0030)	(0.0038)	
Average flights-to-runway ratio at end-point airports	The average value in previous year	–0.2421	0.0512	--
		(0.2543)	(0.2444)	
Maximal flights-to-runway ratio at end-point airports	The average value in previous year	–0.1233	–0.0680	--
		(0.1656)	(0.1682)	
Number of seats in regional market	The average value in previous year	0.0265	0.0112	--
		(0.0045)	(0.0042)	
Number of flights in the regional market	The average value in previous year	–0.0043	–0.0012	–0.0034
		(0.0007)	(0.0005)	(0.0022)
Number of carriers in the regional market	The average value in previous year	–0.0053	0.0054	0.0106
		(0.0062)	(0.0060)	(0.0078)
HHI (Herfindahl-Hirschman Index) of the regional market	The average value in previous year.	–0.5900	–0.2046	0.0332
		(0.0967)	(0.1047)	(0.0649)
Number of legacy carriers in the regional market	The average value in previous year	--	--	0.1246
				(0.0108)
Vacation dummy	1 if at least one of the two airports is located in Florida or Nevada	--	--	0.0585
				(0.0238)
Hub route dummy	1 if only one airport is the hub of some major full service airlines	--	--	–0.0300
				(0.0208)
Double hub route dummy	1 if both the two airports are hubs of some major full service airlines	--	--	–0.0431
				(0.0386)
Pseudo R^2		0.18	0.16	0.14
Number of routes		3,573	3,573	13,569
Number of Observations		258,322	258,322	666,371

Notes: period = month for Ryanair and EasyJet, and = quarter for Southwest respectively; Adjacent routes are the nearby parallel routes to the one under consideration. The end-point airports of an adjacent route are located within 100km of the end-point airports of the route under consideration; The HHI is calculated based on the seats (passengers in Southwest case) of carriers in a market.

Table 2. DID Parameter Estimates of the Effects of the LCCs on Average Fares.

Variables	Ryanair and EasyJet		Southwest	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
LCC route presence	–0.2976	–0.3895	–0.2080	–0.2607
	(0.0035)	(0.0066)	(0.0028)	(0.0033)
LCC adjacent presence	–0.0308	–0.0330	0.0183	0.0199
	(0.0030)	(0.0032)	(0.0035)	(0.0038)
LCC one-airport presence × LCC	0.0052	0.0089	0.0500	–0.0605
connectivity	(0.0126)	(0.0134)	(0.0081)	(0.0093)
LCC two-airport presence × LCC	–0.0435	–0.0491	0.0242	–0.0672
connectivity	(0.0109)	(0.0116)	(0.0047)	(0.0056)
LCC one-airport presence × LCC	0.0011	–0.0018	–0.2159	–0.1549
connectivity × dummy of no LCC entry	(0.0004)	(0.0004)	(0.0083)	(0.0092)
in the sample period				
LCC two-airport presence × LCC	–0.0001	–0.0062	–0.1344	–0.0520
connectivity × dummy of no LCC entry	(0.0003)	(0.0005)	(0.0101)	(0.0111)
in the sample period				
Log of number of passengers	–0.0539	0.1165	–0.0807	0.0641
	(0.0009)	(0.0101)	(0.0004)	(0.0041)
Log of number of carriers	0.0078	–0.0445	0.00004	–0.0999
	(0.0016)	(0.0035)	(0.0008)	(0.0029)
Year dummies included?	Yes	Yes	Yes	Yes
Month (quarter) dummies included?	Yes	Yes	Yes	Yes
Route dummies included?	Yes	Yes	Yes	Yes
Number of routes	3,573	3,573	13,590	13,590
Number of observations	289,546	289,546	762,534	762,534
R^2				
within	0.15	0.03	0.20	0.07
between	0.47	0.17	0.16	0.02
overall	0.35	0.13	0.19	0.01

Note: In the IV estimations, we use the geometric mean of the population of the endpoint cities as an instrument for log passengers.

Table 3. Conditions a Treated Route and a Control Route Must Satisfy

TREATED ROUTE	CONTROL ROUTES
Module 1 (on US markets only): The effect of Southwest's actual entry without potential entry, excluding the effect of its adjacent entry.	
1. Southwest entered a route after 1993 and served the route for at least 8 quarters. 2. Southwest was not a potential competitor on the route before its actual entry. 3. Southwest did not serve an adjacent route before its actual entry.	1. It is not adjacent to the treated route. 2. Southwest did not serve the route and the route's adjacent routes for at least 8 quarters after it actually entered the treated route. 3. Southwest was not a potential competitor on the route before and at least 8 quarters after it actually entered the treated route. 4. Southwest eventually entered the control route in later years without potential entry before the actual entry.
Module 2 (on both US and EU markets): The effect of an LCC's actual entry conditional on both types of an LCC's potential entry, excluding the effect of an LCC's adjacent entry.	
EU markets 1. The LCC entered the route in a month after July 2006 and continued to provide service. 2. The LCC was a potential competitor on the route for at least 18 months before actual entry. 3. The LCC did not serve an adjacent route before the actual entry. 4. After the LCC entered the route, the other LCC did not enter the route and an adjacent route for at least 24 months. 5. Before the LCC entered the route, the other LCC was not a potential competitor on the route.	EU markets 1. It is not adjacent to the treated route. 2. An LCC did not serve the route and the route's adjacent routes for at least 24 months after the actual entry on the treated route. 3. It was potentially entered by an LCC at least 18 months before the actual entry on the treated route. 4. The same LCC that entered the treated route entered the route in at least 24 months after the actual entry on the treated route.
US markets 1. Southwest entered the route after the second quarter of 1994 and continued to serve the route for at least 8 quarters.	US markets 1. It is not adjacent to the treated route. 2. Southwest did not serve the route and the route's adjacent routes for at least 8 quarters after the actual entry on the treated route.

2. Southwest was a potential competitor on the route for at least 6 quarters before the actual entry. 3. Southwest did not serve an adjacent route before actual entry.	3. It was potentially entered by Southwest at least 6 quarters before the actual entry on the treated route. 4. Southwest eventually entered the route at least 8 quarters after the actual entry on the treated route.

Module 3 (on both US and EU markets): The effect of an LCC's potential entry, excluding the effects of its actual and adjacent entry.

EU markets 1. An LCC started to operate continually at one or both endpoint airports, but not the route itself after July 2006. 2. Following the potential entry, an LCC did not actually enter the route for at least 24 months. 3. Before the potential entry, an LCC did not serve the endpoint airports of an adjacent route. US markets 1. Southwest became a potential competitor after the second quarter of 1994 and continued to serve the endpoint airport or airports. 2. If Southwest served both endpoint airports, the time between entering them sequentially was not less than 8 quarters. 3. After it became a potential competitor, Southwest did not enter the route for at least 8 quarters. 4. Before it became a potential competitor, Southwest did not enter an adjacent route. 5. After it became a potential competitor, Southwest did not serve an adjacent route.	EU markets 1. It is not adjacent to the treated route. 2. It was free of actual, potential and adjacent entry before the potential entry on the treated route. 3. The same LCC potentially entered the treated route at least 24 months after the potential entry on the treated route. US markets 1. It is not adjacent to the treated route. 2. After Southwest became a potential entrant on the treated route, it did not serve the control route and the adjacent routes for at least 8 quarters. 3. After Southwest became a potential entrant on the treated route, it did not become a potential entrant on the control route for at least 8 quarters. 4. The distance of the control route was similar to the distance of the treated route. Because there are so many routes for potential entry, we use distance as an additional criteria to control for the number of controlled routes and for heterogeneity.

Module 4 (on both US and EU markets): The effect of an LCC's adjacent entry, excluding the effect of its actual and potential entry.

EU markets
1. An LCC operated on one of the adjacent routes after July 2006 for at least 24 months.
2. After adjacent entry, the LCC was not a potential competitor for at least 24 months.
3. The other LCC was not an actual or adjacent competitor of the route.

US markets
1. Southwest served at least one of the adjacent routes after the second quarter of 1994 for at least 8 quarters
2. After Southwest's adjacent entry, it was not a potential competitor for at least 8 quarters.

EU markets
1. It is not adjacent to the treated route.
2. An LCC did not serve the route and its adjacent routes for at least 24 months after actual entry on the treated route.
3. An LCC was a potential competitor on the route at least 18 months before entry on the treated route.

US markets
1. It is not adjacent to the treated route.
2. After Southwest's adjacent entry on the treated route, it did not serve the control route and its adjacent routes for at least 8 quarters
3. After Southwest's adjacent entry on the treated route, it was not a potential competitor on the control route for at least 8 quarters
4. The distance of the control route was similar to the distance of the treated route.

Table 4. DID Matching Results on the Effect of Southwest's Actual Entry without Prior Potential Entry on Route Average Fare

	All routes entered		Connecting routes entered only	
	Excluding the effects of observed market characteristics (1)	Without excluding the effects of observed market characteristics (2)	Excluding the effects of observed market characteristics (3)	Without excluding the effects of observed market characteristics (4)
Short-run effect (1–2 quarters after entry)	–10.2% [–10.5%, –9.8%]	–23.8% [–24.2%, –23.4%]	–11.3% [–11.6%, –10.9%]	–20.4% [–20.8%, –20.0%]
Medium-run effect (3–4 quarters after entry)	–13.2% [–13.6%, –12.7%]	–24.3% [–24.8%, –23.8%]	–13.1% [–13.5%, –12.5%]	–21.0% [–21.6%, –20.5%]
Long-run effect (5–6 quarters after entry)	–13.7% [–14.1%, –13.2%]	–24.6% [–25.1%, –24.2%]	–13.4% [–14.0%, –13.0%]	–20.8% [–21.4%, –20.4%]
Number of treated routes	227	227	161	161
Number of observations	9,093	9,093	5,204	5,204

Notes: We report median along with [5%-ile, 95%-ile] for each of the effects. The confidence interval is calculated using the bootstrap technique.

Table 5. DID Matching Results on the Effect of a LCC's Actual Entry with Potential Entry on Route Average Fare

	EU Conditional on both types of potential entry (1)	US conditional on Type 1 potential entry (2)	US conditional on Type 2 potential entry (3)
Short–run effect (0–6 months after entry)	–14% [–16%, –12%]	–17.7% [–18.7%, –16.9%]	–4.0% [–4.6%, –3.4%]
Medium–run effect (6–12 months after entry)	–15% [–17%, –12%]	–17.5% [–18.2%, –16.8%]	–5.0% [–5.7%, –4.4%]
Long–run effect (12–18 months after entry)	–10% [–13%, –8%]	–18.8% [–19.9%, –17.8%]	–3.8% [–4.5%, –3.0%]
Number of treated routes	120	159	136
Number of observations	477	2,925	1,800

Notes: Southwest did not potentially enter the controlled route up to 8 quarters after the actual entry on the treated route. Type 1 potential entry = present at only one end-point airport; Type 2 potential entry = present at both end-point airports. The effects of observed market characteristics are excluded in all results. We report median along with [5%-ile, 95%-ile] for each of the effects. The confidence interval is calculated using the bootstrap technique.

Table 6. DID Matching Results on the Effect of a LCC's Potential Entry on Route Average Fare

| | EU | | US | |
| | Type 1 | Type 2 | Type 1 | Type 2 |
	(1)	(2)	(3)	(4)
Short-run effect (0–6 months after entry)	–0.1% [–0.2%, –0.016%]	–1.3% [–2.8%, –0.1%]	–2.3% [–2.9%, –1.9%]	–8.3% [–8.7%, –7.9%]
Medium-run effect (6–12 months after entry)	–0.3% [–0.08%, –0.44%]	–2.2% [–3.6%, –0.6%]	–3.3% [–3.9%, –2.9%]	–9.7% [–10.0%, –9.1%]
Long-run effect (12–18 months after entry)	0.6% [–0.1%, 1.1%]	–0.3% [–1.3%, 0.8%]	–3.2% [–3.8%, –2.7%]	–7.2% [–7.7%, –6.8%]
Number of treated routes	180	82	2,287	224
Number of observations	4,025	1,198	73,889	7,944

Notes: In column (4), the time interval between carriers entering the two airports is not less than 6 quarters. Type 1 potential entry = present at only one end-point airport. Type 2 potential entry = present at both endpoint airports. The effects of observed market characteristics are excluded in all results. We report median along with [5%-ile, 95%-ile] for each of the effects. The confidence interval is calculated using the bootstrap technique.

Table 7. DID Matching Results on the Effect of a LCC's Adjacent Entry on Route Average Fare

	EU (1)	US (2)
Short–run effect (0–6 months after entry)	–2.8% [–4.4%, –1.2%]	–3.0% [–3.4%, –2.6%]
Medium–run effect (6–12 months after entry)	–3.5% [–5.2%, –1.9%]	–3.9% [–4.3%, –3.5%]
Long–run effect (12–18 months after entry)	–1.3% [–2.7%, 0.01%]	–5.1% [–5.5%, –4.6%]
Number of treated routes	77	441
Number of observations	823	7,348

Notes: The effects of observed market characteristics are excluded in all results. We report median along with [5%-ile, 95%-ile] for each of the effects. The confidence interval is calculated using the bootstrap technique.

Table 8. DID Matching Results on the Route Average Fare Effect of Southwest's Actual Entry on Routes without Other LCCs' Entry

	Direct entry without potential entry (1)	Southwest operated at only one of the end-point airports before entry (2)	Southwest operated at both the end-point airports before entry (3)
Short–run effect (1–2 quarters after entry)	–10.8% [–11.4%, –10.3%]	–20.0% [–21.4%, –18.8%]	–4.9% [–5.7%, –3.9%]
Medium–run effect (3–4 quarters after entry)	–13.7% [–14.4%, –13.0%]	–20.7% [–22.5%, –19.1%]	–6.2% [–7.3%, –4.9%]
Long–run effect (5–6 quarters after entry)	–14.8% [–15.4%, –14.1%]	–23.0% [–24.6%, –21.0%]	–5.3% [–6.7%, –4.2%]
Number of treated routes	146	133	89
Number of observations	3,617	811	736

Notes: No other LCCs served either the treated or the controlled routes up to 8 quarters after the actual entry of Southwest on the treated route. The effects of observed market characteristics are excluded in all results. We report median along with [5%-ile, 95%-ile] for each of the effects. The confidence interval is calculated using the bootstrap technique.

Appendix

Table A1. Summary Statistics of US Data

Variable	Mean	S.D.	Min	Max
Dummy of route presence of Southwest	0.113	0.317	0	1
Dummy of presence at the origin airport of Southwest	0.352	0.478	0	1
Dummy of presence at the destination airport of Southwest	0.424	0.494	0	1
Number of Low-cost carriers on the route	0.172	0.464	0	5
Number of legacy carriers on the route	2.545	1.715	0	8
Number of carriers in the city market	3.941	2.917	1	18
Number of legacy carriers in the city market	3.002	1.923	0	8
Number of route passengers (1,000)	0.701	2.25	1	50.54
Route average fare ($)	215.1	76.89	16.55	1,501
Number of passengers in the city market (1,000)	2.6	10.8	1	243.4
HHI of the city market	0.610	0.279	0.111	1
Number of flights in the city market (1,000)	3.21	6.41	45	67.13
Number of routes served by Southwest connecting the origin airport	10.17	16.22	0	64
Number of routes served by Southwest connecting the destination airport	12.46	17.44	0	63
Origin hub dummy	0.211	0.408	0	1
Destination hub dummy	0.194	0.395	0	1
Geometric mean of population of end-point cities (Millions)	1.622	1.855	0.068	16
Geometric mean of income-per-capita of endpoint cities (1,000 $)	30.21	3.87	14.83	54.35
Route distance (1,000 miles)	1.045	0.648	0.025	4.004
Number of observations	762,534			

Notes: We do not include taxes in the fares because the tax does not vary across domestic markets.

Table A2. Summary Statistics of EU Data

Variable	Mean	S.D.	Min	Max
Route average fare (euros)	141.8	70.28	26	4,430
Route distance (1,000 km)	1.141	0.796	0.1	4.996
Number of carriers in the regional market	7.002	5.862	1	48
Dummy of the route presence of Ryanair	0.167	0.373	0	1
Dummy of the route presence of EasyJet	0.088	0.283	0	1
Number of passengers in the regional market (1,000)	11.48	20.45	1	227.9
Percentage of first-class passengers in the regional market	0.016	0.051	0	1
Percentage of business class passengers in the regional market	0.037	0.074	0	1
Number of served routes by Ryanair connecting the origin airport	7.026	13.790	0	99
Dummy of presence at the origin airport of Ryanair	0.406	0.491	0	1
Number of routes served by Ryanair connecting the destination airport	7.132	16.29	0	99
Dummy of presence at the destination airport of Ryanair	0.431	0.495	0	1
Number of routes served by EasyJet connecting the origin airport	5.084	8.574	0	69
Dummy of presence at the origin airport of EasyJet	0.563	0.496	0	1
Number of routes served by EasyJet connecting the destination airport	4.687	8.750	0	69
Dummy of presence at the destination airport of EasyJet	0.506	0.500	0	1
Number of flights in the regional market	146.9	236.4	1	2,832
Number of seats in the regional market (1,000)	19.28	33.33	19	388.3
HHI of the regional market	0.639	0.308	0.070	1
Geometric mean of population of end-point catchment areas (Millions)	2.581	1.926	0	13.95
Geometric mean of income-per-capita of end-point catchment areas (1,000 euros)	35.26	11.39	3.79	97.04
Number of observations	289,546			

Notes: We obtained tax information in the European data but to be consistent with our treatment of the US data we did not include taxes in the fares.

Chapter 7 Appendix. Potential Benefits of Seamless International Airline Travel

Austin J. Drukker and Clifford Winston

Airline passengers prefer direct flights over connecting flights, especially for international journeys where the flight times and transfer wait times for connecting flights are considerably longer than those for most domestic flights. We provide a rough estimate of the potential benefits from international seamless travel—that is, travel on the same airline for the entire international journey—that could result from globally deregulated and better integrated airline networks, which would reduce transfer wait times for connecting flights. We compare the difference between the total current disutility that US international travelers incur from transfer wait times for their connecting flights and the lower disutility that they would incur from seamless travel.

We perform the comparison based on a sample of US residents departing the US for an international flight, which we obtain from the Airlines Reporting Corporation (ARC). The ARC data are derived from online credit card transactions for flights booked through travel agencies, including all subsidiaries of Expedia Group and Booking Holdings. We use scheduling data from the Official Airline Guide (OAG) for all flights in 2018 coupled with routing choice data from the ARC.

Worldwide scheduling data from OAG were used to construct possible itineraries for the flight routes chosen by individuals in the ARC sample. We considered all individuals in the ARC data who departed the US on an international flight with at most one layover. Transfer wait times were constructed to be at least as long as a minimum feasible transfer wait time, which we assume to be 3 hours to account for the time it takes an international traveler to arrive at the gate for the connecting flight and the time the traveler must wait for the connecting flight to depart from the gate.

For example, consider a traveler flying from the United States to Pisa, Italy, and connecting at London Heathrow Airport (LHR). LHR advises travelers to allow 90 minutes to get to the gate for the connecting flight to Pisa. The traveler then must wait for the flight to Pisa to arrive at and depart from the gate. British Airways offers the only nonstop flight to Pisa from LHR, leaving once a day in the late afternoon. So, it is likely that, on average, a traveler must wait at least 90 minutes before the connecting flight to Pisa arrives at and departs from LHR.

Our sample contains 145 destinations and indicates that a significant share of travelers, nearly 40%, takes connecting flights. We assume travelers' hourly value of transfer wait time is $200, based on Yan and Winston's (2014) analysis of air travelers' carrier and routing choices for US domestic flights. This figure is a conservative estimate for our purposes because travelers are likely to place a higher value on transfer wait time in a foreign airport than in a US airport given that they may be concerned about missing important announcements about their flight or they may be less comfortable waiting in an unfamiliar environment. The total loss in utility from transfer wait time for passengers traveling on each connecting route is computed as the number of passengers on each route multiplied by the average transfer wait time (in hours) on each route multiplied by $200 per hour.[142]

Table A3 presents the economic loss to transferring passengers, total wait time, average wait time, and the total number of passengers for the top 20 airports in terms of economic losses

---

[142] The data are representative of the universe of US leisure and unmanaged business travel. To expand our sample to the population, we compared our direct and transfer passenger counts for some large airports with publicly available data reported on their websites. The numbers indicated that we have an approximately 3.5% sample; thus, we inflated our passengers counts assuming our sample accounts for 3.5% of US international travelers. Our assumption is consistent with internal documentation from ARC. Specifically, ARC reports that approximately 35% of domestic travel is booked through travel agencies, of which 20% is sent to a credit card processing company to extract details about the purchaser, which would imply an approximately 7% random sample for domestic travel. ARC also reports that a smaller share of international travel is booked through travel agencies compared with domestic travel. Our tabulations suggest about half as many travelers book through travel agencies for international travel compared with domestic travel.

incurred by connecting international travelers. Based on all the airports in our sample, the total annual economic loss incurred by travelers from transfer wait time amounts to $24.5 billion.

Table A4 shows the underlying source of the loss in terms of the share of travelers transferring at airports with various transfer wait times in 30-minute increments. About 60% of transferring passengers have wait times between 3 and 3.5 hours, 40% of transferring passengers have wait times exceeding 3.5 hours, 20% of transferring passengers have wait times exceeding 4.5 hours, and 7% of transferring passengers have wait times exceeding 8 hours. Thus, travelers' losses are attributable to all travelers experiencing transfer wait times of at least three hours and to many travelers experiencing wait times that are greater than three hours.

Global airline deregulation would give airlines the incentive and ability to provide seamless travel that significantly reduces transfer wait times at international airports because carriers could better align their connecting flights with their originating flights coming from the United States. Of course, no one knows how much airlines could reduce transfer wait time in the globally deregulated environment. A plausible assumption is that airlines could optimize their scheduling and operations such that passengers do not have to spend more than 3 hours to reach their gate and wait for the connecting flight to depart to its final destination. The assumed layover would not put passengers at much risk of missing their connecting flight because the incoming flight was late. Based on this assumption, we replace all the current average transfer wait times that exceed 3 hours in our sample with an assumed value of 3 hours, calculate the total economic loss, and subtract it from the current loss to obtain an annual gain in travelers' welfare of $6.9 billion.[143]

Table A5 shows the counterfactual economic gain from seamless travel, total extra wait time, average extra wait time, and total passengers for the top 20 airports, based on travelers'

---

[143] Even if we assume airlines could reduce all transfer wait times to be no more than 4 instead of 3 hours, the total gain in travelers' welfare would be $5.4 billion.

potential benefits. Several of those airports currently have average transfer wait times that exceed 4.5 hours, which should not be difficult for carriers to reduce.

The preceding calculation understates the benefits from seamless travel for three reasons. First, it does not include the potential benefits associated with reducing stochastic delay, which could arise if a US traveler misses their connecting flight in a foreign country because the originating domestic flight is delayed or the connecting flight is cancelled. Currently, in such cases, the traveler may have to wait for a connecting flight that departs the next day or possibly in the next few days. A carrier competing in a globally deregulated environment that provides all the flights on an itinerary would have the incentive and ability to adjust its flight schedules and operations to reduce the length and cost of stochastic delay.

Second, as noted, our assumed value of transfer wait time for international travel is conservative. In the case of stochastic delay, the disutility associated with the unanticipated additional transfer wait time that causes travelers to arrive at their destinations at least a day late could be much greater than the value of the disutility we have assumed for transfer wait time, which is based primarily on frequency delay, because travelers could miss the first day or few days of a long-awaited vacation or an important business meeting.

Finally, seamless travel is likely to reduce the incidence of lost luggage that occurs on connecting flights when luggage transported by the originating airline must be transferred to a different airline, which may depart from a different terminal. Lost luggage, and travelers' exasperation from trying to find it and get it delivered to their hotel or residence, is less likely to occur when the same carrier is transporting the luggage for the entire itinerary.

Table A3. Top 20 Airports in Terms of Annual Economic Losses from Transfer Time

Country	Airport	Annual economic loss ($ billions)	Annual passengers transferring (millions)	Annual wait time (millions of hours)	Average wait time (hours)
United Arab Emirates	Dubai International Airport (DXB)	1.96	2.09	9.81	4.7
South Korea	Incheon International Airport (ICN)	1.53	2.19	7.63	3.5
Turkey	Istanbul Airport (IST)	1.48	1.53	7.41	4.8
Great Britain	Heathrow Airport (LHR)	1.40	2.18	6.98	3.2
United Arab Emirates	Abu Dhabi International Airport (AUH)	1.23	0.85	6.15	7.2
Germany	Frankfurt Airport (FRA)	1.21	1.81	6.06	3.3
China	Hong Kong International Airport (HKG)	1.09	1.57	5.45	3.5
Japan	Narita International Airport (NRT)	0.91	1.47	4.56	3.1
China	Shanghai Pudong International Airport (PVG)	0.87	1.01	4.33	4.3
China	Beijing Capital International Airport (PEK)	0.83	1.03	4.16	4.0
Taiwan	Taiwan Taoyuan International Airport (TPE)	0.82	1.13	4.08	3.6
France	Paris Charles de Gaulle Airport (CDG)	0.81	1.19	4.07	3.4
Netherlands	Amsterdam Airport Schiphol (AMS)	0.79	1.20	3.93	3.3
Qatar	Hamad International Airport (DOH)	0.69	0.44	3.45	7.8
Canada	Pearson International Airport (YYZ)	0.67	0.95	3.33	3.5
Germany	Munich Airport (MUC)	0.65	0.90	3.26	3.6
Singapore	Singapore Changi Airport (SIN)	0.57	0.43	2.85	6.6
Iceland	Keflavík Airport (KEF)	0.51	0.35	2.52	7.2
China	Guangzhou Baiyun International Airport (CAN)	0.49	0.52	2.44	4.7
Switzerland	Zürich Airport (ZRH)	0.48	0.38	2.38	6.2

Table A4. Share of Passengers with Various Transfer Wait Times

Transfer Wait time	Share of passengers
3–3.5 hours	0.601
3.5–4 hours	0.132
4–4.5 hours	0.060
4.5–5 hours	0.042
5–5.5 hours	0.026
5.5–6 hours	0.014
6–6.5 hours	0.022
6.5–7 hours	0.016
7–7.5 hours	0.008
7.5–8 hours	0.007
8+ hours	0.072

Table A5. Top 20 Airports in Terms of Counterfactual Welfare Gain from Reducing Transfer Times to 3 Hours

Country	Airport	Annual economic gain from reducing wait time ($ billions)	Annual transfer passengers (millions)	Annual wait time above 3 hours (millions of hours)	Average wait time above 3 hours (hours)
United Arab Emirates	Abu Dhabi International Airport (AUH)	0.72	0.85	3.59	4.2
United Arab Emirates	Dubai International Airport (DXB)	0.71	2.09	3.54	1.7
Turkey	Istanbul Airport (IST)	0.56	1.53	2.81	1.8
Qatar	Hamad International Airport (DOH)	0.42	0.44	2.12	4.8
Singapore	Singapore Changi Airport (SIN)	0.31	0.43	1.56	3.6
Iceland	Keflavík Airport (KEF)	0.30	0.35	1.48	4.2
China	Shanghai Pudong International Airport (PVG)	0.26	1.01	1.30	1.3
Russia	Sheremetyevo International Airport (SVO)	0.26	0.33	1.27	3.9
Switzerland	Zürich Airport (ZRH)	0.25	0.38	1.23	3.2
Austria	Vienna International Airport (VIE)	0.25	0.27	1.23	4.6
China	Beijing Capital International Airport (PEK)	0.22	1.03	1.08	1.0
South Korea	Incheon International Airport (ICN)	0.21	2.19	1.04	0.5
Poland	Warsaw Chopin Airport (WAW)	0.19	0.38	0.97	2.6
China	Guangzhou Baiyun International Airport (CAN)	0.17	0.52	0.87	1.7
China	Hong Kong International Airport (HKG)	0.15	1.57	0.74	0.5
Taiwan	Taiwan Taoyuan International Airport (TPE)	0.13	1.13	0.67	0.6
Germany	Frankfurt Airport (FRA)	0.13	1.81	0.62	0.3
Belgium	Brussels Airport (BRU)	0.12	0.22	0.61	2.8
Germany	Munich Airport (MUC)	0.12	0.90	0.58	0.6
France	Paris Charles de Gaulle Airport (CDG)	0.10	1.19	0.51	0.4

**Chapter 8. Public Airport Competition and Some Potential Benefits of
Private Airport Competition**

Austin J. Drukker and Clifford Winston

1. Introduction

Commercial airports in the United States were private enterprises until the Great Depression, when they experienced serious financial problems because of plummeting passenger demand. Government could have given airports financial assistance so they could remain in the private sector. Instead, they were put under the control of state and local governments, which had the sole authority to issue bonds to pay for airport facilities and operations. Today, almost all commercial airports are owned and operated by public authorities.

When faced with competition, private enterprises are pushed to innovate, improve product quality, and enhance operational efficiency, or be driven out of the market by more efficient firms. But when profit is not a motive, as is the case for subsidized government run airports, do competitive forces still improve efficiency that could reduce airline fares?

Competition between multiple airports in a metropolitan area may be limited because new airports rarely enter a market to compete with existing airports and because airports have little financial incentive to attract additional airlines and travelers.[144] In addition to federal subsidies provided by the Airport Improvement Program, airports obtain funds in a monopoly environment from automobile parking fees, retail store rents, advertising display charges, rents from terminal facilities such as counters and gates, and weight-based landing fees promulgated by US Department of Transportation policy. Long-term contracts negotiated between airlines and airports before deregulation have deterred new airline entry by containing "majority in interest"

---

[144] Some airports have tried to encourage Southwest Airlines to serve them and there are examples of smaller airports trying to increase the number of airlines that provide service.

clauses, which give airlines that sign long-term lease agreements the right to approve spending on terminals and gates (Winston (2010)).[145]

In general, public airports do not fail financially and exit the industry. During major shocks to air travel, including the period after September 11, 2001, the Great Recession, and the COVID-19 pandemic, airports have been in much better financial shape than airlines, several of which have gone into bankruptcy, and a few have been liquidated. In contrast, the aviation trust fund can run and (currently is running) large budget deficits. Financial analysts point out that airports tend to have high bond ratings because they have multiple revenue streams that service their debt.

The possibility that public airports engage in competition to some extent is motivated by the hypothesis that air travelers originating from multi-airport metropolitan areas pay lower fares, all else constant, than air travelers originating from single-airport metropolitan areas. The reason is that more airlines serve multi-airport metropolitan areas than serve single-airport metropolitan areas—especially low-cost and ultra-low-cost carriers—that engage in direct competition and facilitate effective adjacent competition (Cho, Windle, and Hofer (2012)). Adjacent competition, which we found in the last chapter to have a measurable effect on airline fares, is possible only in multi-airport metropolitan areas.

Additionally, low-cost carriers are attracted to metropolitan areas with multiple airports because they can choose the airport that is best suited to their operations. For example, Southwest Airlines preferred for many years to use less-congested airports and would bypass the most congested airport in a metropolitan area. Cho, Windle, and Dresner (2015) argue that the mere

---

[145] As these agreements expire, airports are increasingly replacing them with agreements that lack the "majority in interest" clauses. Airport projects also can facilitate entry. For example, Denver International Airport is building many new gates over a protracted period that will be used by its two largest incumbent carriers, United and Southwest Airlines. New entrants may possibly use the older gates that United and Southwest abandon. In addition, some airlines reshape their networks by moving their operations to different airports. For example, JetBlue shifted its West Coast service from Long Beach Airport to Los Angeles International Airport.

presence of a low-cost carrier at an airport may attract passengers, even to competing carriers, as travelers expand their search activity.

In this chapter, we study the effects of airport competition on airline prices by exploiting exogenous variation in the number of airports serving different metropolitan origins that serve the same metropolitan destination. Airport presence is plausibly exogenous because almost all US airports currently in operation were built when airline fares were still regulated by the Civil Aeronautics Board. In fact, only one new major airport, Denver International Airport, has been built in the United States since 1973.[146]

To the best of our knowledge, ours is the first empirical analysis to examine whether travelers flying out of multi-airport metropolitan origins pay lower fares than travelers flying out of single-airport metropolitan origins. This is an important question because the existence of airport competition suggests that travelers could potentially be made better off if: (1) additional public airports are constructed; (2) public airports are privatized and compete more vigorously for travelers and airlines than public airports currently compete for them; and (3) new private airports enter metropolitan areas that can accommodate additional airports.

We find that travelers flying out of metropolitan origins with at least *three* airports pay lower fares, *ceteris paribus*, than travelers flying out of metropolitan origins with only one airport; but that fares from metropolitan origins with exactly *two* airports are higher compared with fares out of single-airport metropolitan origins. In other words, our results suggest that the presence of two airports is not sufficient to induce competitive pressures to reduce fares (and may actually be

---

[146] One small airport, Northwest Florida Beaches International Airport, was built in 2010 to replace Panama City–Bay County International Airport and San Juan International Airport has been operated and managed under a public–private partnership since 2013. In addition, some large general aviation airports, such as Orlando Sanford International Airport, have added passenger service. Finally, Paine Field, north of Seattle, resumed commercial service, currently anchored by Alaska Airlines, at a new privately operated terminal in 2019.

counterproductive), but that three or more airports generate sufficient competition among airlines to reduce fares.

This finding is somewhat surprising *a priori*, and we investigate possible explanations. Our preferred explanation is that three or more airports competing in a metropolitan origin facilitate more entry by low-cost and ultra-low-cost carriers and facilitate adjacent competition. In other words, the channel by which three or more airports reduces fares is by generating additional *airline* competition. We test for additional channels by which three or more airports could reduce fares, namely *airport* competition—which is affected by eliminating common ownership and by expanding airport capacity, gates, and runways that could reduce delays and airlines' operating costs—but we find no evidence that supports those channels.

In contrast, when two airports serve a metropolitan origin, they tend to differentiate their service by developing distinct business models consisting of network airlines primarily catering to international and domestic markets and point-to-point airlines primarily catering to domestic markets. The distinct business models do not reduce fares because they do not encourage additional airline competition in the same markets.

We estimate that if metropolitan areas that are currently served by two airports were served by three airports, where such an increase would be feasible based on the metropolitan areas in question having comparable populations to the metropolitan areas served by three airports, travelers would benefit by more than $4 billion in annual fare reductions. A cost–benefit analysis that accounts for the cost of constructing new airports suggests that several large metropolitan areas would benefit from building an additional airport. Because our analysis is based on public airports, we argue that new private airports, which would have greater incentives to compete

vigorously to attract and provide facilities for additional airlines, could generate even greater reductions in fares.

2. **Characterizing Airport Competition**

Airports serve the same originating market if they draw travelers from the same geographic region. Brueckner, Lee, and Singer (2014) attempt to define a multi-airport airline market by using regression analysis to correlate route-level fare changes at a metropolitan area's primary airport with the number of competing airlines at other airports in the area. But anecdotal evidence, as well as work by McWeeny (2019), suggests that Brueckner, Lee, and Singer's (2014) market definition might be too conservative because it neglects some airports in a broad metropolitan area that travelers use.[147]

Hence, we report results based on conservative and liberal market definitions. The conservative definition is aligned with Brueckner, Lee, and Singer's (2014) definition, which includes primary and core airports.[148] The liberal definition includes primary, core, and fringe airports, as defined by Brueckner, Lee, and Singer, as well as airports that are more than 75 miles from the city center, and small airports that primarily serve low- or ultra-low-cost carriers. We also constructed definitions within the preceding extremes (for example, by not including small airports); however, we report results based only on the most conservative and liberal definitions because the main findings were not affected by small changes to the list of airports that we included in the market definitions.

---

[147] We thank John Heimlich of Airlines for America, Kenneth Strickland of Raleigh–Durham International Airport, Devon Barnett of Tampa International Airport, and Christopher Birch of San Francisco International Airport for helpful conversations regarding airline market definitions from an industry perspective.

[148] We also include Richmond International Airport (RIC) and Newport News/Williamsburg Airport (PHF) in the conservative definition. Brueckner, Lee, and Singer (2014) do not include those airports, but our results do not change if we do not include them in the conservative definition.

As shown in the first panel of table 1, the airports that comprise multi-airport metropolitan areas in the conservative market definition are within an average radius of 17 miles from the centroid of the metropolitan area, with a maximum radius of 51 miles. Under the conservative market definition, the average farthest driving distance between any two airports is 46 miles, with a maximum distance of 105 miles (Manchester-Boston Regional Airport (MHT) to Providence Airport (PVD), about 1 hour and 45 minutes apart).

As shown in the second panel of table 1, airports included in the liberal market definition are within an average radius of 28 miles from the centroid of the metropolitan area, with a maximum radius of 68 miles. The average farthest driving distance between any two airports is 68 miles, with a maximum distance of 140 miles (MHT to Bradley International Airport (BDL), about 2 hours and 15 minutes apart).

Some descriptive evidence suggests that the liberal market definition is the more appropriate multi-airport airline market definition. For example, according to survey data collected by Ipsos on behalf of Airlines for America, nearly 40 percent of travelers originated their trip during 2018 from an airport that was not the closest airport to their home or office.[149] McWeeny (2019) takes a close look at the behavior of airline travelers in the San Francisco Bay Area and reports that the majority bypass airports that are closer to their homes to take advantage of lower fares out of San Francisco International Airport (SFO), and that, on average, leisure travelers are willing to drive up to 2 hours to save $100 on fares.[150] A significant share of travelers drives from as far as Sacramento, bypassing Sacramento International Airport (SMF), while Sonoma County Airport (STS) even provides a trip calculator that includes parking fees at and

---

[149] We thank John Heimlich for providing us access to these data.

[150] Generally, the tradeoff is affected by gasoline prices. Higher gasoline prices would reduce the amount of time leisure travelers would be willing to spend driving to save money on lower fares.

travel costs to SFO to discourage travelers from taking long drives that bypass the flights it offers.[151] Although public airports are not profit-maximizing enterprises, they do have an interest in attracting passengers to generate revenues from ancillary services and to justify government subsidies.

Figure 1 shows the multi-airport metropolitan areas in our sample. The light blue dots correspond to the conservative market definition, and the dark blue dots plus the light blue dots correspond to the liberal definition, which are encompassed by circles. The 15 largest single-airport metropolitan areas are represented by yellow dots.

Table 2 presents descriptive evidence of the differences between fares on flights from metropolitan areas served by a different number of airports. Specifically, the table reports average yields (passenger-revenue per passenger-mile) for markets served by one, two, and three or more airports and shows that they are lower for markets served by three or more airports compared with those served by fewer airports, a finding that is robust to the alternative market definitions. However, yields for markets served by two airports are slightly lower than yields for markets served by one airport under the conservative definition while they are notably higher under the liberal definition. Recall, that the liberal definition includes several more metropolitan areas served by two airports than are included under the conservative definition.

Table 3, which disaggregates yields by the number of airports in a metropolitan area and by the hub classification (or size) of the airports, shows that differences in yields are more pronounced.[152] Yields for metropolitan areas with three or more airports are much lower than

---

[151] According to data collected at Tampa International Airport (TPA), a nonnegligible number of international passengers flying out of TPA drive from Orlando, which is roughly 90 miles away. We thank Kenneth Strickland and Devon Barnett for sharing this insight with us.

[152] A large hub handles at least 1 percent of national enplanements, a medium hub from 0.25 to 1.0 percent of enplanements, a small hub from 0.05 to 0.25 percent of enplanements, and a nonhub has less than 0.05 percent of enplanements but more than 10,000 annual enplanements.

yields for metropolitan areas with a smaller number of airports when the metropolitan area in question contains a large hub airport. Yields for medium hubs show an interesting pattern where they are lowest for metropolitan areas with one airport and highest for metropolitan areas with two airports. Yields for small hubs and nonhubs are much higher for metropolitan areas with one airport than for metropolitan areas served by multiple airports.

In sum, the lower yields for metropolitan areas with three or more airports may be driven by their lower yields at large hubs that reflect intense competition or lower costs from economies of distance and aircraft size. Metropolitan areas with one airport have lower yields than metropolitan areas with two airports for large and medium hubs, but not for small hubs and nonhubs, in all likelihood because those markets may be quite isolated and able to charge high fares or less able to benefit from economies of distance and aircraft size. Because the descriptive comparisons do not hold other important influences on fares constant, we turn to a formal econometric analysis to estimate causal effects.

3. **Estimating the Effect of Airport Competition on Fares**

We estimate an airline fare regression to determine the causal effect of airport competition on fares. In its basic reduced form, airline fare regressions specify the (log) yield in a market as a function of airline competition, route characteristics such as distance to capture costs and intermodal competition, distance squared to capture economies of length of haul, and population and income to capture exogenous measures of demand (Morrison and Winston (1995)).[153] It also

---

[153] As an additional demand variable, a greater share of business travelers could lead to higher fares, but it would be an endogenous variable in a fare regression. Income at the origin is an exogenous variable that can capture travelers' willingness to pay. As pointed out below, we also restrict the sample to include only destinations that are served by both single- and multi-airport metropolitan origins to help control for different characteristics across routes. As an additional cost variable, airports that have higher costs per enplanement could lead to higher fares. However, the relevant issue is how airport costs affect airlines' costs, whose airport operations are subsidized because aircraft are charged only for their takeoff but not landing operations. A representative landing fee per passenger for typical

is common to include route (airport-pair or city-pair) fixed effects to control for any time-invariant route characteristics (Brueckner, Lee, and Singer (2014)).

As we discussed in our analysis in chapter 3 of the effect of bus and rail transit competition on bus costs, it is important to carefully consider which variables should and should not be held constant so that we are able to identify the causal effect of the number of airports on fares. We specify the number of airports serving a metropolitan area as dummy variables for two or three-or-more airports with one airport as the base. Given that this variable is time-invariant and specific to the origin market, we cannot include route fixed effects or origin fixed effects because the origin component of the route fixed effect will be perfectly collinear with the multi-airport dummy variables. Thus, instead of specifying route fixed effects, we include destination fixed effects to control for unmeasured time-invariant influences on fares at the destination.

Including multi-airport dummy variables in the specification to capture airport competition raises the question of whether *airline* competition should be held constant because it is one of the mechanisms by which *airport* competition—through its ability to attract low-cost competitors and to facilitate adjacent competition—could affect fares. Thus, *airport* competition may be prevented from affecting fares if *airline* competition is held constant. Additionally, airline presence is likely to be endogenous because carriers' entry will be influenced by revenue and average fares on a route, while airport presence is arguably exogenous because, as noted, most US airports were built during the period when fares were regulated.

We therefore specify and estimate fare regressions of the following form:

_____

passenger loads is less than three dollars (Morrison and Winston (2007)) and any variation would have only a small effect on fares. Importantly, aircraft also are not charged for their contribution to congestion and delays during takeoffs and landings.

$$\ln \text{yield}_{ijyq} = \alpha + \beta_1 \, \text{two}_i + \beta_2 \, \text{three}_i + \gamma_1 \ln \text{inc}_{iyq} + \gamma_2 \ln \text{pop}_{iyq} + \delta_1 \ln \text{dist}_{ij} +$$

$$\delta_2 \ln \text{dist}_{ij}^2 + \xi_j + \xi_y + \xi_q + \varepsilon_{ijyq},$$

where $\ln \text{yield}_{ijyq}$ is the log of the yield from origin airport i to destination airport j in year y and

quarter q. The key variables of interest are two_i and three_i, dummy variables that equal 1 if origin

airport i is part of a two- or three-or-more-airport metropolitan origin, respectively; 0 otherwise.

The base metropolitan origin is served by a single airport. The remaining control variables in the

specification include log of income at the origin, $\ln \text{inc}_{iyq}$; log of population at the

origin, $\ln \text{pop}_{iyq}$; log of distance and log of distance squared between the origin and destination,

dist_{ij} and dist_{ij}^2; destination fixed effects, ξ_j; and year and quarter fixed effects, ξ_y and ξ_q.[154] We

also specified slot-control dummy variables for JFK, LGA, and DCA airports, but they were

statistically insignificant. We discuss the possible effect of airport capacity on fares later.

We estimate the model using data on fares and flight distances from the US Department of

Transportation's Airline Origin and Destination Survey (DB1B), a 10% quarterly sample of

domestic itineraries. Data on population and income are from the US Census Bureau and Bureau

of Economic Analysis, and data for the number of airports in the metropolitan area are from the

Bureau of Transportation Statistics. Our sample covers the period from 1993 to 2019 and consists

of approximately 3,400 routes and 149,000 observations, depending on the market definition used.

The year fixed effects control for national-level structural changes in the airline industry that

occurred during our sample period, such as the September 11, 2001, terrorist attacks and the Great

---

[154] One could argue that population should be divided by the number of airports at the origin. However, population is a measure here of potential airport demand throughout the metropolitan area. We have made the point that it is reasonable to assume that travelers consider all airports in a metropolitan area because they frequently use airports that are not closest to their residence or place of business. In addition, dividing population by the number of airports would impose a restriction where the effect of the number of airports on fares is affected by metropolitan area population, which has no theoretical justification.

Recession.[155] Finally, to partly control for different characteristics across routes, we restrict the sample to include only destinations that are served by both single- and multi-airport metropolitan origins, and we include only round-trip direct flights.[156]

We present parameter estimates from ordinary least squares regressions in table 4. The results for the airport dummy variables based on the conservative market definition, shown in the first column, lack statistical significance, possibly because there are only 14 multi-airport markets (9 two-airport markets and 5 three-airport markets, as shown on the left side of table 1).

The second column shows results for the airport dummy variables based on the liberal market definition, which results in 22 multi-airport markets. The parameter estimates are statistically significant, and the signs are consistent with the descriptive differences in yields shown in table 2. The estimates indicate that the presence of exactly two airports in a market *increases* yields by 8.9% relative to yields in single-airport markets, while the presence of three or more airports in a market *decreases* yields by 10.1% relative to single-airport markets. The estimated decrease in yields is plausible given that multiple airports should facilitate both direct and adjacent competitive pressures on fares. As shown in chapter 7, direct competition from Southwest can reduce fares by as much as 30%; the presence of other carriers in a market has a smaller but not negligible effect on fares. Chapter 7 also found that Southwest can reduce fares by as much as 5% in the long run when it provides adjacent competition.

The finding that two airports do not lower fares, but in fact *increase* fares, relative to single-airport metropolitan areas, is surprising given that we expect that fares should decrease with the

---

[155] We performed tests of whether our parameters of interest—the effect of multi-airport presence on yields—were affected by the September 11, 2001, terrorist attacks and the Great Recession, but we could not reject the null hypothesis that the parameters are equal before and after those shocks to the industry, so the pooled estimation is appropriate.

[156] We also exclude lightly travelled routes with fewer than 300 passengers per quarter.

number of airports serving a market, or at least not increase with the number of airports serving a market. In the next section, we disaggregate the data further to provide possible explanations for that finding.

4. **Explaining the Differential Effects of Airport Competition**

The preceding estimates imply that airport competition has differential effects: two airports in a market do not generate sufficient competition to reduce fares and, in fact, their presence increases fares, while three or more airports in a market do generate sufficient competition to reduce fares. We argue that if airport competition affects fares, it does so by affecting airline competition in the metropolitan area. For example, the "Southwest effect"—the observation that originating air travel tends to increase considerably after Southwest's entry into a market—has been well established (Dresner, Lin, and Windle (1996) and Morrison and Winston (2000)). We therefore explore the source of airport competition's differential effects on yields and the justification for excluding airline competition from the specification by re-estimating our model with measures of airline competition and seeing how the estimates of the airport competition dummy variables are affected.

We measure airline competition as the number of traditional, low-cost, and ultra-low-cost carriers serving the market. Table 5 shows the specific airlines in each carrier classification.[157] Data on the airports served by each airline come from the US Department of Transportation's Air Carrier Statistics database (T-100).

---

[157] As Southwest and JetBlue expand into international service and into more (congested) US airports, the distinction between their operations and traditional carriers' operations is not as stark as it used to be. We do not consider private aviation as a distinct carrier classification because the difference between its cost and commercial carriers' first or business class fares is too great to suggest that private aviation and commercial airline service are plausible substitutes for more than a tiny share of travelers.

Table 6 presents ordinary least squares estimation results showing how the previous estimates of airport competition, reproduced in column (1), are affected by including airline competition in the model. Column (2) shows that the number of low-cost carriers reduces yields, and the effect is statistically significant, although the parameter estimate is likely to be biased because the number of low-cost carriers is endogenous. The important finding is that the coefficient for the two-airport dummy increases and it is still statistically significant while the quantitative effect of the three-airport dummy is virtually eliminated, and it becomes statistically insignificant. Thus, the effect of three or more airports on yields appears to be capturing the effect of the number of low-cost carriers on fares, as direct or adjacent competitors, which indicates a specific mechanism by which airport competition reduces fares. At the same time, the effect of two airports is hardly affected by controlling for low-cost carrier competition, which suggests that a different mechanism explains how that form of airport presence affects fares.

Column (3) shows that the preceding findings are not affected when we add the number of traditional and ultra-low-cost carriers to the specification. So, it appears that airport competition in metropolitan areas with three or more airports reduces fares by facilitating additional low-cost carrier competition. However, we need to probe more deeply to understand why the presence of two airports in a metropolitan area increases fares relative to one airport.

Table 7 presents descriptive evidence on yields, airport size (based on hub classification), and the share of international traffic for each airport in markets with two airports (first panel) and with three or more airports (second and third panels) to assess the extent that airports within those markets compete with each other. For example, consider two airports in the same metropolitan area with only one of the airports offering international flights. Those airports may be less likely to be served by airlines with overlapping domestic route networks that compete intensely with

each other. The reason is that airlines offering international service tend to be major network carriers that align their domestic and international routes accordingly, while airlines that do not offer international service tend to be smaller point-to-point carriers with route networks that have a small overlap with larger carriers' route networks. In contrast, a third airport in a metropolitan area is likely to increase airline competition because it will be served by major network carriers or point-to-point carriers or both.

Recall that air travelers often bypass airports closest to their homes and use airports farther away to obtain lower fares, better service, or both. We presented McWeeny's (2019) evidence for the San Francisco multi-airport market to illustrate that travelers with the option to use three or more airports may drive to a farther airport to obtain lower fares to their metropolitan destination, which may be served by different carriers operating out of different origin airports. In addition, those travelers may drive to a farther airport to fly on their preferred carrier based on frequent flier mileage accumulation. In contrast, travelers in two-airport markets are less likely to have the option to use an alternative airport to obtain lower fares, although they might drive to a farther airport to take international or domestic flights that are not offered at the airport that is closest to their home or workplace.[158]

Evidence from Airlines for America's 2019 "Air Travelers in America" survey is consistent with the conjecture that travelers from metropolitan areas with three or more airports are more likely than travelers from metropolitan areas with two airports to use an airport that is farther from their home or workplace. Specifically, 39.7% of travelers flying from metropolitan origins served by three or more airports used an airport that was not the closest airport to their

---

[158] Canadians living close to the US border often drive to US airports because airfares and airport fees are lower in the US. The main instances of substitute airports along the US-Canadian border are *Toronto*: Niagara Falls International Airport and Buffalo Niagara International Airport; *Montreal*: Plattsburgh International Airport; *Ottawa*: Ogdensburg International Airport; and *Vancouver*: Bellingham International Airport.

home or office, while 28.7% of travelers flying from metropolitan origins served by two airports used an airport that was not the closest airport to their home or office.[159]

The first panel of table 7 shows that in two-airport metropolitan areas, one of the airports is often served by airlines that provide international service and is either a large or medium hub airport, while the other airport is not served by airlines that provide international service and is often a small hub or nonhub airport. For example, in the Detroit area, Detroit Metropolitan Airport (DTW) is a large hub that provides international flights and is served by almost all the major carriers, while Flint Bishop Airport (FNT) is a small hub that provides few, if any, international flights and its only major carrier is Delta Air Lines. Note that yields at FNT are 15% lower than yields at DTW, in all likelihood because international flights tend to be more expensive than domestic flights and because more passengers flying out of DTW are likely to be paying more expensive business and first-class fares on longer flights.[160]

In sum, the airlines providing service in two-airport metropolitan areas appear to segregate and specialize their operations such that they provide different services at each airport in accordance with their different business models and cater to different types of travelers to a significant extent. It is thus plausible that airlines that serve a single-airport metropolitan origin, which includes both network and point-to-point carriers, could generate more competition that reduces yields than airlines that serve a two-airport metropolitan origin.

Regulatory restrictions on international airline competition also may contribute to different airports pursuing different airline business models because burdensome regulations often make it

---

[159] It is more likely that the survey results capture the effect of exogenous characteristics of airports, including airlines with low fares, better service, and the like, on travelers' airport choice behavior rather than those airport characteristics being caused by travelers' airport choice behavior.

[160] FNT's lower fares in recent years also could be explained by service provided by low-cost and ultra-low-cost carriers.

difficult for carriers to provide international service from specific airports. Such restrictions include entry barriers in international markets that are not governed by open skies agreements and entry barriers such as slot and gate restrictions at domestic and foreign airports that comprise US international routes.

The second and third panels of table 7 show that in three-airport markets, at least two of the airports are often served by carriers that provide some international service and are large or medium hubs. For example, the Washington market is served by three large hub airports, all of which provide some international service and one of which, Baltimore/Washington International Airport (BWI), has long been served by Southwest Airlines, which, in addition to providing direct competition, has helped to reduce yields by providing adjacent competition for airlines flying out of Reagan National Airport (DCA) and Dulles International Airport (IAD) (Cho, Windle, and Dresner (2015)).[161]

Table 8 presents descriptive evidence that a greater number of low-cost and ultra-low-cost carriers serve markets with three or more airports compared with markets with only one or two airports. Ultra-low-cost carriers serve a greater number of large hub single-airport markets than they serve two-airport markets, and low-cost carriers serve a comparable number of single-airport markets and two-airport markets. Medium hub single-airport markets are served by fewer low-cost and ultra-low-cost carriers than are served at two-airport markets. In contrast, metropolitan origins with three or more airports are more likely than single- and two-airport metropolitan origins to be served by low-cost and ultra-low-cost carriers, and, on average, have more low-cost, ultra-low-cost, and total carriers offering flights than single- and two-airport metropolitan origins have.

---

[161] BWI was a hub for US Airways, but Southwest Airlines essentially drove them out of BWI. Of course, other low-cost and ultra-low-cost carriers operate in the three Washington metropolitan area airports.

5. **Considering Additional Channels that Enable Airport Competition to Reduce Fares**

We have found evidence that appears to indicate airport competition reduces fares by facilitating direct and adjacent airline competition, especially involving low-cost carriers. However, there are additional channels by which the presence of multiple airports in a metropolitan area could reduce fares, including true airport competition and capacity expansion.

In true airport competition, airports would take certain actions that reduce airlines' operating costs to attract additional airlines to serve their airports. Airlines that are induced to serve airports that provide such incentives would have lower costs and possibly lower fares. We have pointed out that most (subsidized) public airports do not aggressively seek to attract additional airlines. However, if such an incentive existed, airports operating in the same metropolitan area would be more likely to take actions to attract additional airlines if they were operated by different airport authorities instead of by a monopoly authority that has little incentive to encourage competition among the airports that it operates.

Metropolitan areas that are served by three or more airports and where at least two of the airports are operated under the same airport authority include the Chicago (ORD, MDW), New York (LGA, JFK, EWR, SWF), and Washington (DCA, IAD) metropolitan areas. Thus, as an empirical test of the possible effect of true airport competition on fares, we added a new variable to our specification where we specified a same-airport-authority dummy variable, interacted it with the three-or-more-airports dummy variable, and re-estimated the fare model. We found that its coefficient was statistically insignificant and that its inclusion had little effect on the magnitude and statistical significance of the three-or-more-airports dummy variable, which indicates that pure airport competition is not a channel that reduces fares.

The capacity expansion channel suggests that the presence of multiple airports results in more runways and more gate and terminal capacity that could, for example, reduce takeoff and landing delays, and reduce operating costs and fares. We tested for the effect of additional capacity directly by interacting the three-or-more-airports dummy with the number of runways in the metropolitan area.[162] Including this interaction term in our regression did not change the magnitude of the coefficient on the three-or-more-airports dummy, and the coefficient on the interaction term was small and statistically insignificant.

Because capacity expansion also may encompass more gates and terminals, we used hub size as a broader measure of capacity. Generally, larger airports and greater airport capacity are associated with larger hub sizes, which are associated with larger metropolitan areas. We again created a new variable where we interacted the dummy variable for three or more airports with dummy variables that identified metropolitan areas with either two large hub airports or with one large hub airport and two medium hub airports and we re-estimated the fare model. We found that the coefficient for the new (capacity expansion) variable was statistically insignificant and that its inclusion had little effect on the magnitude and statistical significance of the three-or-more-airports dummy variable. In sum, the statistically insignificant effects of the number of runways and hub sizes suggest that capacity expansion is not a channel that reduces fares in multi-airport metropolitan areas.[163]

---

[162] We included only runways that were longer than 5,000 feet, which is an adequate length for most commercial aircraft to take off and land. Data on runway length are from the Federal Aviation Administration's National Flight Data Center. We are grateful to Jan Brueckner and Alberto Gaggero for sharing the data with us.

[163] In theory, there could be a "delay" channel where airports take actions to reduce delays, which increases effective competition and reduces fares, although conversely such action may increase airport demand and raise fares. Regardless, lower delays would increase consumers' surplus and benefit travelers. However, in practice, airports do not set congestion fees to reduce delays and only a few airports limit flights with slot controls, which raise fares. Efforts to reduce delays by building more runways is likely to generate induced demand, like highways, and not have very much effect on delays.

6. **The Potential Gains from Additional Airport Competition**

Given that our analysis suggests airport competition in metropolitan areas with three or more airports can help to reduce fares, it is of interest to calculate the potential benefits of increasing the number of airports in metropolitan areas served by fewer than three airports. The left panel of table 9 lists the markets, as defined by metropolitan area, served by three or more airports, alongside their populations. The right panel identifies markets served by one or two airports that could possibly sustain three airports, given that their populations are comparable to markets currently served by three or more airports. We have simply identified those markets as candidates for a cost–benefit assessment; we do not assert that they could sustain more airports with certainty.

In addition to raw population comparisons, other factors make those markets reasonable candidates for additional airports. For example, Dallas and Houston are large metropolitan areas with millions of passengers flying through their airports annually, while land prices in Texas are relatively low compared with land prices in the rest of the US. The Detroit metropolitan area *did* at one point have three functioning airports—DTW, FNT, and Coleman A. Young International Airport (DET)—but following Detroit's declaration of bankruptcy in 2013, DET was listed as an asset of the city that could be sold to cover debts and it currently does not serve any commercial passengers. Charlotte is currently served as a fortress hub for American Airlines, Charlotte Douglas International Airport (CLT), and could plausibly sustain a third airport in addition to Concord Regional Airport (JQF). St. Louis, which currently has a larger population than Charlotte, also could possibly sustain a third airport in addition to Lambert International Airport (STL) and MidAmerica Airport (BLV). In addition, several large single-airport metropolitan areas could

possibly sustain two additional airports, including Atlanta (home to the busiest airport in the world), Seattle, Minneapolis, San Diego, and Denver.

<div align="center">Benefits of Additional Public Airports</div>

We calculate the annual fare savings from adding additional public airports using our parameter estimates from table 4. For the two-airport markets, we calculate the reduction in yields caused by the addition of a third airport, which is the difference between the coefficient estimates for the three-airport dummy and the two-airport dummy: –0.101 – 0.089 = –0.190. This figure represents the percentage change in fares per passenger-mile. We then multiply –0.190 by the annual passenger-miles for the markets. This calculation assumes that the total volume of passenger-miles is unchanged with the addition of a third airport, which provides a lower bound on total benefits. We estimate that the annual fare savings from adding an additional airport to the five two-airport metropolitan areas shown in Table 9 is roughly $4.1 billion.

Considering the benefits from fare reductions alone is not enough to justify building a new airport, because the construction and capital costs may exceed the benefits. Setting the annual savings on fares equal to the annualized capital cost of building an airport allows us to solve for the maximum airport construction cost that would still result in net benefits.[164] The results of this cost–benefit analysis also are reported in Table 9. We estimate that the addition of a third airport in Dallas, for example, would provide net benefits if the cost to build the airport were less than $31.43 billion. We do not know what the cost of a new airport in Dallas would be, but it would likely be far less than $31.43 billion, given that Denver International Airport (DEN) was built in 1995 at a cost of about $8.2 billion in 2020 dollars.[165] Taking DEN as the baseline, new airports

---

[164] The calculation assumes a 30-year lifetime for the airport and a 3% discount rate.

[165] The costs of building new airports today are likely to be inflated by a modest amount by increasingly higher capital costs and higher costs of environmental review, legal services, and the like.

in Houston, Detroit, and Charlotte would almost certainly provide positive net benefits from fare reductions alone.

Table 9 also presents the results of similar calculations for the five largest single-airport markets. The annual savings from adding two additional airports are calculated analogously to the two-airport case, except that we need only the coefficient estimate for the three-airport dummy from Table 4: –0.101. Multiplying –0.101 by the annual passenger-miles for the one-airport markets yields the annual savings on domestic fares. We report the maximum cost *per airport* that would provide net benefits. Building two new airports in Atlanta, for example, would provide net benefits from fare reductions alone if the construction cost for *each airport* were less than $8.56 billion. Given the construction cost of DEN, it is possible that building two new airports in Atlanta would provide net benefits, while the cost to build two new airports in Seattle, Minneapolis, San Diego, and Denver would have to be much lower than the cost to build DEN to provide net benefits.

As noted, our estimates of the benefits from fare reductions are conservative because we hold annual passenger-miles constant. Importantly, the *total* benefits from building a new airport include much more than the benefits from fare reductions. Additional airports would attract more airline service and more flights and possibly increase a metropolitan area's economic development (Green (2007)), which would generate significant benefits to travelers and residents and strengthen the case for building additional airports.

A possible argument against the potential benefits of additional airports, associated with Delta Air Lines' operations in Atlanta Hartsfield–Jackson Airport (ATL), a monopoly airport, is that Delta's dominant position has enabled it to develop an extensive network that provides a large choice of nonstop destinations and more frequent service. Delta also has been able to consolidate flow-through traffic to achieve economies of density that reduce its operating costs and fares to

some extent. However, there is no reason why Delta's alleged benefits to travelers cannot be subjected to a market test. If an additional airport or airports were built in Atlanta and other airlines entered those airports and provided service, then travelers would reveal their preferred combination of airline fares and service quality and preferred airlines. In the process, Delta's market share would be affected in accordance with travelers' preferences, which would leave travelers at least as well off as before the additional airport or airports were built.[166]

Another consideration when assessing the potential benefits of new airports is whether air traffic control is significantly disrupted by multiple airports in the same region and whether it is appreciably more difficult with multiple airports to create flight paths that would avoid generating noise over heavily populated areas. We are not aware of evidence that travel in metropolitan areas with multiple airports is compromised by air traffic problems that are more common in such areas or that residents in those metropolitan areas experience more noise than residents in single airport metropolitan areas. In any case, those issues should be addressed by improving air traffic control services and government airplane noise policies (Morrison, Winston, and Watson (1999)).

<u>Potential Benefits When Airports are Private</u>

It is important to bear in mind that our estimates of the effects of airport competition are based on an analysis of *public* airports, which have vast pricing, investment, and operating inefficiencies (Winston (2013)). As we have indicated, public airports do not have strong incentives to compete for airlines and travelers, and we have found that those incentives are not increased in the case of pure airport competition even when each airport operates under different

---

[166] If the additional airport or airports are public, the taxpayers would bear the cost if those airports could not overcome Delta's advantage and attracted few carriers and passengers. Airport investors would bear the cost if the additional airport or airports were private, and they would have the incentive to give careful thought to whether a new airport entrant could compete successfully and to vigorously explore alternative strategies following entry to overcome any advantage for Delta that existed.

airport authorities. In practice, public airports may face opposition from incumbent airlines to build additional terminal space and gates to enable new carriers to provide service (Winston (2010)).[167] Morrison and Winston (2000) estimated that the absence of available gates cost travelers $4 billion in higher fares. In addition, Delta is an example of an airline that has made aggressive efforts through lobbying to protect its strong competitive position at ATL by preventing Atlanta-area general aviation airports from adding commercial airline service.

In contrast to US cities, many cities throughout the world, including London, New Delhi, Rome, Sydney, and Tokyo, have privatized their airports subject to varying degrees of regulation. Winston (2013) reports that case studies find that privatization has improved airport efficiency in some countries, and Oum, Yan, and Yu (2008) find in a worldwide comparison of airports that airport privatization has promoted competition and reduced costs.

Suggestive evidence that private airport competition could reduce fares and improve service is provided by experience in the UK and other European countries. Starkie (2009) concludes that private airport competition in the UK has been an effective regulator of what an airport can charge an airline. Of particular interest to the US is that, like contract rate negotiations that have enabled railroad shippers to obtain lower rates by playing one railroad off against another, UK airports have had to compete for airlines by offering beneficial long-term contracts that stipulate the charge and service quality that an airline can expect from an airport.

Another benefit of UK airport privatization is that the local airport at Luton renamed itself as London Luton Airport and allowed the emerging low-cost carrier Ryanair into the Dublin–

---

[167] A few public airports have attempted to overcome that opposition. For example, in 2018, Miami International Airport changed its gate fees policy from per-use to flat monthly fees, which opened the door for Frontier and other low-cost carriers (including Southwest) to enter an airport that had been too costly to use. Passenger Facility Charges (PFCs) have enabled some airports to more easily finance terminal expansions that would have been much harder to do when most airports were locked into long-term leases with legacy carriers that gave them veto power over terminal expansions.

London route, which was at a vital stage in its development.[168] Chapter 7 contained an estimate that Ryanair has reduced fares 20% in the European Union markets that it serves. The implication of the UK's experience for the US is that some existing general aviation airports could expand their operations to provide commercial service and generate airport competition in a privatized environment without requiring substantial capital investments.

Still another potential benefit that has emerged from airport privatization abroad is that larger airports are likely to try to attract low-cost carriers with low airport charges to generate large volumes of passengers for their facilities with spare capacity. Smaller airports have anticipated this competition and have attracted low-cost carriers by building special facilities and providing fast turnaround times.

Because private commercial airports in the United States would seek to maximize the return on their investments, they would have stronger incentives than public airports to compete for all types of airlines and to increase passengers by providing terminal capacity and gates for any airline that is willing to pay the (marginal) cost of the facilities.[169] Yan and Winston (2014) simulated the effect of airport privatization by developing a model where privatized airports in the San Francisco Bay Area with separate owners compete for airline operations by setting profit-maximizing runway charges that reduce delays and airlines compete for passengers. Runway charges are determined through separate negotiations between airlines organized as a bargaining unit and each of the three Bay Area airports, Oakland, San Jose, and San Francisco.

The authors find that by setting different charges for different classifications of airport users (scheduled commercial carriers and general aviation), the Bay Area airports would gain from

---

[168] Luton's local government did not actually sell the assets of the airport, but it granted a long-term operating lease to a private company.

[169] In 1997, Congress established the Airport Investment Partnership Program, but only a handful of small airports have taken part.

privatization, as would commercial travelers and carriers. Commercial air travelers would pay higher fares because airport charges to airlines would increase, but the time savings from less-congested air travel would more than offset that cost. General aviation would face higher charges, but their losses would be softened if policymakers expanded airport privatization to encourage (smaller) private airports to compete for (smaller) aircraft operations.

It also is possible that some private airports would convert their operations and compete for commercial airline traffic by taking advantage of advances in GPS technology that have improved access to smaller airports, by upgrading runways and gates, and by offering van and rental car service to improve travelers' access to the central city and other parts of the metropolitan area. By having more flight alternatives, travelers in low-density markets could especially benefit if private airports nationwide offered commercial service.

In the context of this analysis, private airport competition could reverse the positive sign of the effect of competition on fares in markets with two airports because private airports are likely to encourage all types of airlines to serve them—that is, carriers that serve domestic short-haul and long-haul markets and international markets—and to compete with other airlines serving the two-airport metropolitan area.

Another advantage of private airport competition is that private airports may enhance competition among airlines by continuing to construct additional terminals for low-cost carriers.[170] This would be a less expensive and more feasible alternative than a new entrant incurring the huge expense of building an entirely new airport. Recall that incumbent airlines could possibly block investments such as new terminals by public airports, but private airports' financial arrangement with airlines would not allow an incumbent airline to prevent a private airport from expanding its

---

[170] Some airports, such as Austin–Bergstrom International Airport, have opened new terminals or concourses for ultra-low-cost carriers in recent years.

capacity to accommodate new airline entrants that were willing to pay the cost of the expanded capacity.

Finally, as private airports shed the inefficiencies that have accumulated in public airports for decades and introduce new policies and operations, pure private airport competition in metropolitan areas with less than three airports may become a channel that reduces airfares and challenges our conclusion based on public airports that such competition has no effect on fares. In sum, private airport competition that results in the construction of new airports is likely to generate greater benefits from lower fares and additional service than the estimated benefits that we reported based on public airports.

7. **Final Comments**

Although US airlines have been deregulated for decades, airports are still largely owned and operated by public authorities. We investigated whether public airports compete in the sense of causing fares to decline in metropolitan areas with a greater number of airports. We found that fares are lower when three or more airports serve a metropolitan origin, because those airports facilitate more direct and adjacent competition from low-cost carriers, and they are not segregated by different carriers that have different business models.

We estimate that the construction of additional public airports in metropolitan areas served by two and possibly by one airport would benefit travelers by generating billions of dollars in fare reductions and would produce large benefits from additional service. We argue that those benefits are likely to be even greater if airports were privatized and if international airline markets were fully deregulated because US metropolitan areas would be served by more domestic and international carriers that would face fewer barriers from serving airports in those areas. Airline

competition in such an environment could result in efficiency gains, technological innovations, and benefits to travelers that we have yet to envision.

Table 1. Market Definitions

Conservative Market Definition				Liberal Market Definition			
Market	Airports	Radius from Centroid (miles)	Maximum Driving Distance (miles)	Market	Airports	Radius from Centroid (miles)	Maximum Driving Distance (miles)
Chicago	ORD:MDW	5	25	Buffalo	BUF:IAG	10	19
Cincinnati	CVG:DAY	34	78	Charlotte	CLT:JQF	7	22
Cleveland	CLE:CAK	14	51	Chicago	ORD:MDW:RFD	68	95
Dallas	DFW:DAL	5	18	Cincinnati	CVG:DAY	34	78
Detroit	DTW:FNT	13	74	Cleveland	CLE:CAK	16	51
Houston	IAH:HOU	25	30	Dallas	DFW:DAL	5	18
Los Angeles	LAX:BUR:LGB	10	41	Detroit	DTW:FNT	13	74
Miami	MIA:FLL	9	29	Fort Myers	RSW:PGD	8	37
New England	BOS:MHT:PVD	51	105	Houston	IAH:HOU	25	30
New York	LGA:JFK:EWR	8	29	Los Angeles	LAX:BUR:LGB:SNA:ONT	21	58
San Francisco	SFO:OAK:SJC	21	35	Miami	MIA:FLL:PBI	48	72
Tampa	TPA:PIE	6	14	New England	BOS:MHT:PVD:BDL	46	140
Virginia Beach	RIC:PHF	12	62	New York	LGA:JKF:EWR:HNP:ISP:SWF	22	108
Washington	DCA:IAD:BWI	31	58	Orlando	MCO:SFB:MLB	23	75
Maximum		51	105	Philadelphia	PHL:TTN:ACY	32	80
Mean		17	46	Phoenix	PHX:TUS:AZA	48	117
				Pittsburgh	PIT:LBE	54	63
				San Francisco	SFO:OAK:SJC:STS:SMF	55	119
				St. Louis	STL:BLV	22	38
				Tampa	TPA:PIE:SRQ	19	53
				Virginia Beach	RIC:PHF:ORF	24	90
				Washington	DCA:IAD:BWI	31	58
				Maximum		68	140
				Mean		28	68

Sources: Brueckner, Lee, and Singer (2014); Google Maps; authors' calculations.
Notes: Radius from centroid measures the distance in miles as the crow flies from the centroid of the metro area. Maximum driving distance measures the distance in miles via road between the farthest two airports in the market.

Figure 1. Airport Locations and Market Definitions
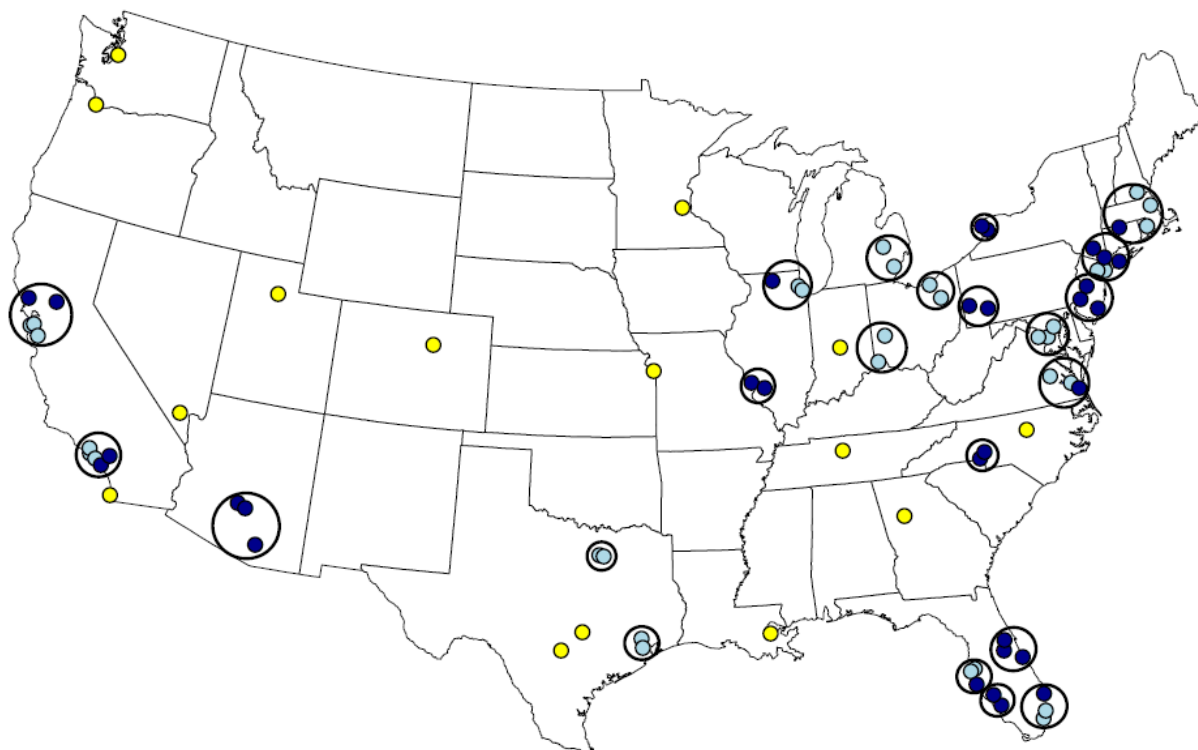
Notes: The yellow dots are single-airport metropolitan areas. The light blue dots correspond to the conservative definition of multi-airport metropolitan areas. The dark blue dots plus the light blue dots correspond to the liberal definition of multi-airport metropolitan areas, which are encompassed by circles.

Table 2. Mean Yields in Dollars by Number of Airports in the Market, 1993–2019

Number of Airports in the Market	Conservative Market Definition	Liberal Market Definition
One	0.187	0.190
Two	0.185	0.209
Three or more	0.156	0.157
All	0.176	0.176

Sources: Bureau of Transportation Statistics; airportcodes.us; authors' calculations.

Table 3. Cross-Tabulations of Mean Yields in Dollars by Number of Airports in the Market and Hub Classification, 1993–2019

Hub Classification	Three or More Airports	Two Airports	One Airport
Large	0.144	0.208	0.176
	(18)	(7)	(10)
Medium	0.196	0.205	0.189
	(10)	(5)	(20)
Small and nonhub	0.189	0.198	0.251
	(17)	(8)	(266)

Sources: Bureau of Transportation Statistics; airportcodes.us; authors' calculations.
Notes: Statistics are for the liberal market definition. The number of airports in each cross-tabulation is in parentheses.

Table 4. The Effect of Multi-Airport Markets on Yields, 1993–2019

Log Yield (dollars)	Conservative Market Definition	Liberal Market Definition
Two airports	–0.004	0.089**
	(0.049)	(0.039)
Three or more airports	–0.062	–0.101***
	(0.049)	(0.038)
No. of observations	149,320	149,600
R^2	0.729	0.741

Sources: Bureau of Transportation Statistics; Bureau of Economic Analysis; Census Bureau; authors' calculations.

Notes: Regressions include controls for log population at the origin, log income at the origin, distance and distance-squared between the origin and destination, destination fixed effects, year fixed effects, and quarter fixed effects. Robust standard errors clustered at the origin metropolitan statistical area are shown in parentheses. Statistical significance is indicated at the ***1 percent, **5 percent, and *10 percent levels.

Table 5. List of Carriers

Name	Notes
Traditional Carriers	
Alaska Airlines	
America West Airlines	Acquired by US Airways in 2005
American Airlines	
Continental Air Lines	Acquired by United in 2010
Delta Air Lines	
Hawaiian Airlines	
Northwest Airlines	Acquired by Delta in 2008
Trans World Airways	Acquired by American in 2001
US Airways	Acquired by American in 2015
United Air Lines	
Low-Cost Carriers	
ATA Airlines	Acquired by Southwest in 2008
AirTran Airways	Acquired by Southwest in 2014
JetBlue Airways	
Southwest Airlines	
Virgin America	Acquired by Alaska in 2018, which remains a traditional carrier
Ultra-Low-Cost Carriers	
Allegiant Air	
Frontier Airlines	
Spirit Air Lines	
Sun Country Airlines	

Source: Bureau of Transportation Statistics.

Table 6. Yields, Multi-Airport Markets, and Number of Carriers by Classification, 1993–2019

Log Yield (dollars)	(1)	(2)	(3)
Two airports	0.089**	0.109***	0.104***
	(0.039)	(0.035)	(0.034)
Three or more airports	–0.101***	–0.035	–0.029
	(0.038)	(0.032)	(0.031)
Number of traditional carriers			–0.009
			(0.006)
Number of low-cost carriers		–0.061***	–0.050***
		(0.010)	(0.010)
Number of ultra-low-cost carriers			–0.026***
			(0.008)
No. of observations	149,600	149,600	149,600
R^2	0.741	0.746	0.748

Sources: Bureau of Transportation Statistics; Bureau of Economic Analysis; Census Bureau; authors' calculations.

Notes: Results are for the liberal market definition. Regressions include controls for log population at the origin, log income at the origin, log of distance and distance-squared between the origin and destination, destination fixed effects, year fixed effects, and quarter fixed effects. Robust standard errors clustered at the origin metropolitan statistical area are shown in parentheses. Statistical significance is indicated at the ***1 percent, **5 percent, and *10 percent levels.

Table 7. Yields in Dollars, Hub Classification, and International Market Penetration, 1993–2019

Market	Hub Class	Yield	% Int'l	Market	Hub Class	Yield	% Int'l	Market	Hub Class	Yield	% Int'l
Buffalo				Chicago				Philadelphia			
BUF	Medium	0.222	0.0	ORD	Large	0.181	9.8	PHL	Large	0.178	9.0
IAG	Non-Hub	0.082	0.3	MDW	Large	0.151	0.3	TTN	Non-Hub	0.110	0.0
Charlotte				RFD	Non-Hub	0.100	0.0	ACY	Small	0.097	0.1
CLT	Large	0.284	4.5	Los Angeles				Phoenix			
JQF	Non-Hub	0.111	0.0	LAX	Large	0.120	19.2	PHX	Large	0.155	0.5
Cincinnati				BUR	Medium	0.260	0.0	TUS	Medium	0.200	0.0
CVG	Large	0.251	3.2	LGB	Small	0.153	0.0	AZA	Non-Hub	0.092	1.6
DAY	Small	0.304	0.0	SNA	Medium	0.201	0.0	San Francisco			
Cleveland				ONT	Non-Hub	0.198	0.1	SFO	Large	0.136	16.8
CLE	Large	0.191	0.6	Miami				OAK	Large	0.172	0.7
CAK	Small	0.193	0.0	MIA	Large	0.146	43.7	SJC	Medium	0.177	1.1
Dallas				FLL	Large	0.117	10.3	STS	Non-Hub	0.274	0.0
DFW	Large	0.203	4.3	PBI	Medium	0.148	0.7	SMF	Medium	0.197	0.0
DAL	Medium	0.228	0.0	New England				Tampa			
Detroit				BOS	Large	0.173	12.5	TPA	Large	0.159	1.7
DTW	Large	0.190	7.5	MHT	Medium	0.237	0.0	PIE	Small	0.095	0.1
FNT	Small	0.163	0.0	PVD	Medium	0.206	0.4	SRQ	Small	0.191	0.0
Fort Myers				BDL	Medium	0.190	0.5	Virginia Beach			
RSW	Medium	0.141	1.0	New York				RIC	Small	0.370	0.5
PGD	Non-Hub	0.086	0.0	LGA	Large	0.219	0.5	PHF	Small	0.224	0.0
Houston				JFK	Large	0.110	49.1	ORF	Medium	0.302	0.0
IAH	Large	0.196	10.7	EWR	Large	0.162	22.8	Washington			
HOU	Medium	0.216	0.5	HPN	Small	0.303	0.3	DCA	Large	0.259	0.3
Pittsburgh				ISP	Small	0.145	0.0	IAD	Large	0.170	25.0
PIT	Large	0.223	1.3	SWF	Non-Hub	0.193	3.6	BWI	Large	0.166	2.4
LBE	Non-Hub	0.075	0.1	Orlando							
St. Louis				MCO	Large	0.139	7.1				
STL	Large	0.201	0.6	SFB	Small	0.093	36.2				
BLV	Non-Hub	0.091	0.3	MLB	Non-Hub	0.242	0.2				

Source: Bureau of Transportation Statistics. *Note:* Percent international is the percentage of passengers with destinations outside of North America.

Table 8. Share of Markets Served by and Number of Low-Cost Carriers, 1993–2019

| Number of Airports in the Market | Share of Markets Served | | Number of Carriers Serving | | | |
	Low-Cost Carriers	Ultra-Low-Cost Carriers	All Carriers	Traditional Carriers	Low-Cost Carriers	Ultra-Low-Cost Carriers
One						
Large hubs	0.870	0.814	10.93	7.00	1.85	2.08
	(0.336)	(0.390)	(2.57)	(1.72)	(1.17)	(1.41)
Medium hubs	0.899	0.643	8.46	5.64	1.45	1.37
	(0.301)	(0.479)	(1.83)	(1.63)	(0.84)	(1.26)
Two	0.920	0.749	9.55	5.78	1.92	1.85
	(0.271)	(0.434)	(1.89)	(1.62)	(0.97)	(1.42)
Three or more	0.973	0.906	12.25	6.87	2.79	2.59
	(0.162)	(0.292)	(2.31)	(1.76)	(1.07)	(1.41)

Sources: Bureau of Transportation Statistics; authors' calculations.
Note: Statistics are for the liberal market definition. Standard deviations are shown in parentheses.

Table 9. Aggregate Consumer Savings from Expanding to Three Airports

Three-Airport Markets		Two-Airport Markets			
Market	Population (million)	Market	Population (million)	Annual Savings on Fares ($ bil.)	Breakeven Airport Construction Cost ($ bil.)
New York	19.22	Dallas	7.57	1.604	31.43
Los Angeles	13.21	Houston	7.07	0.837	16.41
Chicago	9.46	Detroit	4.32	0.605	11.85
New England	7.70	St. Louis	2.80	0.262	5.13
Washington	6.28	Charlotte	2.64	0.790	15.49
Miami	6.17	Total	24.40	4.098	
Philadelphia	6.10	One-Airport Markets			
Phoenix	6.00	Atlanta	6.02	0.874	8.56
San Francisco	4.73	Seattle	3.98	0.594	5.82
Tampa	3.19	Minneapolis	3.64	0.393	3.86
Virginia Beach	3.06	San Diego	3.34	0.276	2.70
Orlando	2.61	Denver	2.97	0.708	6.94
Total	87.73	Total	19.95	2.844	

Sources: Bureau of Transportation Statistics; Census Bureau; authors' calculations.
Notes: Results are for the liberal market definition. Annual savings on fares measures the total annual reduction in fares for all domestic flights from the market. Breakeven airport construction cost is per airport and is found by setting the annualized capital cost of building an airport equal to the annual savings, assuming a 30-year lifetime and a 3 percent discount rate. Populations are for 2019 metropolitan statistical areas. The New England market includes Boston, Manchester, Providence, and Hartford. The Phoenix market includes Tucson. The Virginia Beach market includes Richmond.

Chapter 9. Private Airport Competition and Innovation: The Case of Heated Runways

Jukwan Lee, Clifford Winston, and Jia Yan

1. Introduction

Winter storms that produce snow and ice on runways and taxiways reduce the ability of aircraft to take off and land safely, which can delay travelers for hours or even days if a flight is cancelled because the airport closes a runway, which significantly slows operations, or stops operating. Closing a runway is not an arbitrary decision. Although the limits vary for types of aircraft, most are prohibited from operating on runways that are covered by untreated ice or that have accumulated more than one-half inch of snow or slush.

Currently, commercial airports in the United States remove snow manually with plows, brushes, blowers, and shovels, in combination with chemical deicers. An alternative and potentially more effective and efficient method for removing snow and ice is to melt it by heating runway and taxiway pavement before it can accumulate and disrupt aircraft operations. However, this innovation is not used on a large scale by any US airport.[171]

This chapter examines whether competition among private airports could spur innovation that leads to the adoption of heated airport runways that reduce air travelers' delays due to snow and ice. We first assess the case for US commercial airports adopting heated runways to improve airline operations and benefit travelers by estimating the benefits to travelers and airlines from heated runways and taxiways, as reflected in reduced aircraft flight delays and passengers' value

---

[171] Some airports use heated surfaces on a small scale. For example, at Chicago O'Hare International Airport, two taxiway bridges over roadways have embedded pipes that carry heated oil, keeping the surfaces free of snow and ice for planes without having to plow it onto passing cars below. Boston's Logan International Airport places giant snow-melters in terminal areas because snow cannot be pushed aside without blocking gates and roads. Homeowners purchase electric systems to melt snow and ice on their driveways and walkways and Green Bay's Lambeau Field was the first in the National Football League to install a heating system to keep the turf from freezing.

of reduced waiting time, and compare those benefits with the costs of installing and operating heated runways.

Consistent with the observed behavior of public commercial airports, we find that a strong case does not exist on cost–benefit grounds for those airports to adopt heated runways. This finding is intuitive because given that public airports receive government funding and do not make investments to maximize profits, they have little economic incentive to invest in innovations like heated runways regardless of whether travelers benefit from less snow-related delay. Installing and operating heated runways would adversely affect airports' budget by increasing their costs, which they could not recover in higher takeoff and landing fees to airlines, because current weight-based landing fees are regulated following Federal Aviation Administration guidelines. Heated runways may be politically attractive to the extent that they reduce fatalities and injuries caused by aircraft accidents, but commercial air travel—even during winter months—has become extremely safe.

We then extend a model of private airport competition developed by Yan and Winston (2014) to investigate whether a private airport would adopt heated runways that could benefit air travelers. In contrast to public airports, a private airport would have an economic incentive to install and operate heated runways because it could increase its profits, after accounting for the higher runway charges that it would be free to charge airlines to cover installation and operating costs. Travelers would benefit from heated runways that reduce snow-related delays if they value the time savings by at least as much as the higher fares they would have to pay to cover airlines' higher (heated) runway charges. Based on our simulations, we find that a private airport can simultaneously increase their profits and benefit travelers by installing heated runways.

Generally, public production is justified when a facility or service is socially desirable but is not privately profitable (Winston (2021b)). However, in this case, public airports would not adopt an innovation, heated runways, which could benefit travelers, while private airports would be willing in certain situations to adopt heated runways. The analysis provides another example of how an environment of competition and innovation improves the efficiency of the US transportation system and the welfare of its users. In contrast, an environment in which the public sector is responsible for enhancing travelers' welfare may not improve the system's efficiency and travelers' welfare because policymakers have little incentive to take actions that could do so.

2. Airports and Data Sources

We conduct our analysis using data on airport operations, weather, and weather-related travel delays for the three commercial airports in the greater Boston area: Boston Logan International Airport (BOS), T. F. Green International Airport (PVD) in Rhode Island, and Manchester–Boston Regional Airport (MHT) in New Hampshire. We attempted to include Worcester Regional Airport (ORH) in Massachusetts, but we could not collect the complete set of variables for that airport that we were able to collect for the other airports in the analysis. We chose the three greater Boston area airports because although they are currently publicly owned and operated, they could explicitly compete as private airports with separate owners because they are located within a 60-mile radius and are used by passengers from the same catchment area.[172]

BOS, PVD, and MHT normally experience winter snowstorms that cause flight delays and cancellations. In addition, because airlines operate over a network, their flights also can be delayed by time-varying network-wide shocks on airport operations that, for example, originate at a flight's

---

[172] Using a conservative definition, as discussed in chapter 8, to identify airports in the same market, Brueckner, Lee, and Singer (2014) argue empirically that BOS, PVD, and MHT belong in the same market.

origin and cause the flight to arrive late to its destination. We choose a control group for those shocks that consists of the four commercial airports in the greater Los Angeles area: Los Angeles International Airport (LAX), Hollywood Burbank Airport (BUR), Long Beach Airport (LGB), and John Wayne Airport (SNA). Delays at both the Boston area and LA area airports can arise from network-wide shocks and are therefore correlated. However, unlike Boston area airports, LA area airports do not experience snow that causes delays; thus, by using LA area airports as the control group, we are able to separate the effects of snow on flight delays from the effects of time-varying network-wide shocks on airport operations.

The first part of our empirical analysis consists of estimating the impact of hourly snow precipitation on airline flight delays. We conduct it using data from April 2004 to March 2015, during which time there were several major snowstorms in the northeast of the United States. Data on airlines' on-time performance are from the Bureau of Transportation Statistics (BTS), and contain information on flight cancellations and scheduled and actual arrival times of domestic flights for the major carriers. Departure and arrival flight delays at an airport are defined as the difference between the scheduled and actual departure or arrival time. Each airport's hourly weather information is from the National Oceanographic and Atmospheric Administration (NOAA) based on its nearest weather observation station.

3. Research Design

Identification of our model of the determinants of hourly flight delays is achieved by using both a parametric difference-in-difference-in-difference (DDD) approach and a non-parametric matching approach.

<u>DDD Identification</u>

Figure 1 illustrates the intuition behind DDD identification of the effect of snow on flight delays. Winter in the greater Boston area runs from December through March. We use a binary indicator of snow precipitation in an hour as the treatment status of a Boston airport. However, given that there are unobserved weather-related factors that also affect delays, we must control for those. Finally, we use LA airports as a control group to remove the effect of network-wide shocks on airport operations.

Controlling for observed weather factors, such as temperature and wind speed, the DDD approach first employs difference-in-differences (DID) during the winter to identify the effects of unobserved weather factors on flight delays. Conditional on the effects of winter weather on flight delays, the DDD approach employs a further difference between snow and non-snow periods of the year on observations from the treatment group to identify the pure effect of snow events on flight delays.

We implement DDD identification by estimating the following regression equation by OLS:

$$\text{Delay}_{ita} = \beta_0 + \beta_1 \times \text{Winter}_t + \beta_2 \times \text{Boston}_a + \beta_3 \times \text{Boston}_a \times \text{Winter}_t$$
$$+ \beta_4 \times \text{Boston}_a \times \text{Winter}_t \times \text{Snow}_{ta} + \beta_5 \times \mathbf{X}_{ita} + \eta_a + \eta_m + \varepsilon_{ita} \qquad (1)$$

where Delay_{ita} is the delay (in minutes) of flight i in hour t at airport a; Winter_t is a dummy variable defined as 1 for the period from December to March, 0 otherwise; Boston_a is a dummy variable for the Boston area airports (with the LA area airports as the reference group); Snow_{at} is a dummy variable indicating the occurrence of snow precipitation in hour t at airport a; $\mathbf{X}_{ita}$ is a vector of observed information about the flight, airport, and weather; η_a controls for airport fixed effects; and η_m controls for month (indexed by m) fixed effects. The βs are estimable parameters

with $\delta \equiv \beta_4 - \beta_3$, which measures the pure effects of snow on flight delay, the main parameter of interest.

Matching Identification

As an additional identification approach, we take advantage of the large amount of flight data to identify the effect of snow on flight delays non-parametrically based on matching observations with similar weather and traffic conditions. The matching procedure is as follows:

— Step 1: We partition the flight observations into two groups: a treatment group including flights experiencing snow during the hour of their scheduled departure or arrival time, and a control group including flights that do not experience snow during their scheduled departure or arrival time.

— Step 2: We match each flight observation in a treatment or control group to the K most similar ones in the other group, where similarity is based on the Mahalanobis distance between two vectors of observables containing a measure of airport traffic, the total number of scheduled flights at the endpoint airports in an hour, and observed weather information including temperature, visibility, precipitation, and wind speed.[173]

The matching algorithm generates a set of N one-to-K matched pairs. We then estimate the effect of snow on flight delays for the N matched pairs using the bias-corrected matching estimator (Adadie and Imbens (2011)) given by:

$$\tau = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{Y}_i^1 - \hat{Y}_i^0 \right) \tag{2.1}$$

---

[173] The Mahalanobis distance is often used to identify outliers in a sample.

$$\hat{Y}_i^1 = D_i Y_i + (1 - D_i)\frac{1}{K}\sum_{j \in T_i}\left(Y_j + \hat{\mu}_1(X_i) - \hat{\mu}_1(X_j)\right) \tag{2.2}$$

$$\hat{Y}_i^0 = (1 - D_i)Y_i + D_i\frac{1}{K}\sum_{j \in C_i}\left(Y_j + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_j)\right) \tag{2.3}$$

In equation (2.1), $\hat{Y}_i^j$, $j = 0, 1$, denotes the imputed delay of a flight for observation i under treatment status j (1 if there is snow; 0 otherwise). In equations (2.2) and (2.3), the imputed delay under a treatment status equals the actual observation if the observation is observed under the treatment status; otherwise, it is computed from the K matched observations from the other group (which have a different treatment status); D_i is an indicator of the treatment status of observation i, taking a value of 1 if the observation is treated; 0 otherwise; and T_i and C_i denote the set of treated and control flights that are matched to observation i and that belong to the control and treated group, respectively. Finally, the conditional mean, $\mu_j(X_i) \equiv E\left(Y_i^j \mid X_i, D_i = j\right), j = 0, 1$, is estimated by OLS.

4. Descriptive Summary of Delays at Boston and LA Airports

Figure 2 decomposes the components of flight delay from the beginning to the end of a flight. Gate departure delay is the difference between a flight's scheduled and actual departure time from the gate. The flight can incur taxi-out delay after it leaves the gate and must wait until its wheels are off the ground. Once the flight is airborne, it can incur delay that increases the time that its wheels are off and back on the ground if it must reduce its speed or take a circuitous route to its destination. After the flight lands, it can incur taxi delay from the time its wheels are on the ground until it arrives at the gate. Finally, gate arrival delay is the difference between a flight's scheduled and actual arrival time at the gate. We measure the effect of snow on gate departure

and arrival delay, as well as its effect on runway-related delays associated with taxing to and from the gate.

Table 1 presents the means and standard deviations for delay and weather-related variables for the Boston and LA airports during the winter and non-winter seasons for comparative purposes. Two clarifications about the data are in order. First, extreme weather delays are departure or arrival delays caused by extreme weather conditions, but the airlines' on-time performance data do not indicate the specific weather conditions that cause extreme delay, which could help us understand why extreme delay is much lower than other types of delay. Second, the data do not include cancellation delays—that is, the time that travelers must wait for the next available departing flight to their destination. We account for cancellation delays in our analysis of private airports' decision to invest in heated runways based on an analysis conducted by Yan et al. (2016).

The table shows, in general, that the means for the delay variables are considerably less than the standard deviations of those variables, indicating the importance of the effect of variations in weather on delays. Departure and especially arrival delays at Boston airports are greater during the winter compared with those delays during non-winter. Extreme weather delays also are greater in the winter than in non-winter as are taxi-out and taxi-in times. In contrast, arrival, departure, and extreme delays and taxi-out and taxi-in times at LA airports do not change very much from winter to non-winter. As expected, Boston airports operate in colder temperatures, with more snow, less visibility, and greater wind speeds but with less precipitation during the winter than in non-winter. LA airports generally operate in mild weather conditions all year around. Finally, both Boston and LA airports handle fewer flights during the winter than they do in non-winter.

5. Estimation Results for Delays

We present in table 2 the DDD OLS regression results of the determinants of delays at the arrival and departure gates and for taxi-out and taxi-in times. The parameter estimates indicate that a snow event increases flights' arrival and departure delays by roughly 5 minutes and increases taxi times by 2.5 minutes. Passengers would incur a much longer delay from a missed connection or a canceled flight because they would be forced to rebook the next available flight to their destination that may depart much later in the day or on the following day. Flight delays also are longer in colder temperatures and in less visibility. Holding other influences constant, additional airport operations increase flight delays, while flight delays at Boston airports are less than flight delays at LA airports and gate delays are less in winter than in other seasons. The positive coefficient for the interaction between Boston airports and the winter season indicates that flight delays also can be attributed to unobserved weather-related factors.

The findings based on the matching estimator, presented in table 3, are robust with respect to the choice of K, the number of the most similar flights to a given flight, and are broadly consistent with those obtained by DDD estimation. The two matching models' estimates of the effect of snow on gate delay closely bound the DDD estimate of the effect of snow on gate delay and are only slightly greater than the DDD estimate of the effect of snow on taxi time.

6. Runway Deicing and Heated Pavement Technologies

Heated runways represent a significant technological improvement over traditional deicing approaches because they have the potential to virtually eliminate snow-related delays by rapidly melting snow and ice.

Traditional Deicing Approaches

Airports use several different pavement deicing and anti-icing agents, including ethylene glycol, propylene glycol, urea, an ethylene glycol-based fluid known as UCAR (containing approximately 50% ethylene glycol, 25% urea, and 25% water by weight), potassium acetate, sodium acetate, sodium formate, and calcium magnesium acetate (CMA). Many airports apply those chemical agents after using mechanical equipment to remove heavy accumulations of snow on runways and taxiways, which increases delay times. Airports also apply anti-icing pavement agents based on predicted weather conditions and pavement temperature. Deicing and anti-icing solutions are applied using either truck-mounted spray equipment or manual methods. Table 4 shows the approximate unit cost and application rate of the chemical deicers.

Heated Pavement Systems

Heated pavement technologies include electric systems, which use electricity to heat fluids, and hydronic systems, which rely on heated fluids from a variety of sources. The most efficient system is geothermal waters, but those fluids may be limited only to runways that are sufficiently close to tectonic plate boundaries. Airports in other locations can consider ground source heat pumps, heat exchangers, or boilers to boost efficiency and reduce operational costs. Alternative heat sources, such as waste heat, may be used if a reliable supply over the design lifetime of a runway is guaranteed. Table 5 summarizes the cost estimates of installing, operating, and maintaining different types of heated pavement systems.

7. Costs and Benefits of Public Boston Area Airports Installing Heated Runways

Heated runway technologies exist that could benefit travelers by reducing snow-related travel delays that they incur at airports that are not prevented by current deicing technology.

However, just because heated runway technology exists does not imply that public airports should adopt that technology or that private airports would find it in their interest to do so.

To shed light on those issues, we conduct a cost-benefit analysis of each Boston area airport's decision to install a heated runway pavement system to eliminate snow-related delays. We assume as public airports that their observed decision to *not* install such a system was consistent with a calculation of the present value of heated runways' net benefits that was not positive. We then assess whether any of the Boston area airports would install a heated runway pavement system if they were privatized and based their decision on whether the system increased their profits.

We assume public airports take actions independently without accounting for the interactions among them. The net present value of installing a heated runway pavement system at an airport is given by:

$$NPV = \sum_{t=1}^{T} \frac{R_t}{(1 + i)^t} - \text{Installation Cost} \qquad (3)$$

where T is the planning horizon, which is assumed to be 20 years; R_t is the one-period net benefit, which is the difference between the social benefits and operating costs associated with the heated pavement system; and i is the discount factor, which is assumed to be 2.15%.[174] The initial installation cost of a heated runway depends on the airport's runway capacity. To account for changes in air traffic over time, we use the US Federal Aviation Administration's assumed annual growth rate of 2.8% for the next 20 years.

---

[174] This is roughly the average 20-year Treasury rate during the past 5 years. In chapter 8, we assumed a slightly higher discount rate of 3% for constructing an airport with a 30-year lifetime.

When an airport installs a heated runway, the investment benefits airlines by reducing their operating costs due to snow-related delays and it benefits travelers by reducing their travel time costs. We calculate the benefits to airlines as:

$$\text{Airlines' Benefits (\$)} = \alpha \times \text{average delay due to snow (minutes)}$$

$$\times \text{ total flight operations in winter}$$

$$\times \text{ average flight delay cost (\$ per flight per minute)} \quad (4)$$

where α is the share of delay that is reduced by a heated runway, which we assume to be one. We estimated previously that the average gate and taxi delay due to snow is 8 minutes per flight. Total flight operations in winter are the aggregated number of arrival and departure flights that are affected by snow precipitation. Table 6 reports that in 2015, 7,549 flights at BOS, 389 flights at MHT, and 768 flights at PVD operated during snowy weather.[175] We estimate that the average cost to an aircraft of one-minute of flight delay in 2015 is \$62.55, based on cost data from American Airlines.[176]

The last column of table 6 shows the number of passengers per flight at each of the airports in the Boston area, which we use to estimate the total number of passengers that experienced flight delay at those airports. According to the US Department of Transportation (2016), the average value of travel time per hour for a traveler was \$47.10. The estimated value of time is much lower than the estimate of \$104 per hour of air travel delay reported in Yan and Winston (2014) based on data for travelers using San Francisco Bay Area airports. We assume the lower estimated value here because our purpose is to use the conventional NPV analysis to model the decision making of public airports, which would be more inclined to rely on DOT's estimate than an independent

---

[175] It is possible that airlines would operate more flights if there were fewer delays; thus, the estimated benefits to airlines may be conservative.

[176] The cost includes additional operating and maintenance costs, such as crew, pilot salary, extra fuel consumption, and other costs.

estimate in a scholarly journal and to make a decision that led them to not install heated runways. We also do not include the higher travel time costs that result from snowy weather that causes flight cancellations and passengers to miss their connections because that information is unlikely to be incorporated by public airports into their decision making.

The benefits of a heated runway at an airport to travelers are then calculated as:

$$\text{Travelers' Benefits (\$)} = \alpha \times \text{average delay due to snow (minutes)}$$

$$\times \text{ total flight operations in winter}$$

$$\times \text{ average number of travelers per flight}$$

$$\times \text{ average value of time} \tag{5}$$

Table 7 presents the benefits of installing heated runways at each of the three Boston-area airports, assuming those runways eliminate snow-related delays. The total annual benefits mainly accrue to airlines and travelers at BOS, and are nearly $10 million, which amounts to a present value of benefits of roughly $165 million. As noted, benefits would be higher if heated runways also led to more flights during winter.

Table 8 shows the construction costs of heated runways at the three airports. Because those costs can vary, we consider low-cost and high-cost scenarios, which correspond to the lower and upper bounds of unit costs of heated runways presented in table 5. We include operating costs, but, based on the data in table 5, those costs are small and would not affect any of the conclusions we reach here.

Table 9 combines the benefits and costs in tables 7 and 8 to present the NPV of installing heated runways at each of the three airports in the Boston area, using a 20-year time horizon. We find that installing heated runways would not yield a positive net present value of benefits for any of the airports, except for the low-cost scenario at Boston Logan.

It is possible to make alternative assumptions such that installing heated airport runways would yield a positive net value of benefits under any cost scenario. For example, improvements in airline service quality could improve passenger demand and increase airport concession purchases and revenues, which are not included here.[177] It also is possible that given lower travel demand during the winter, Boston Logan could reduce the cost of heated runways by installing fewer of them, which encompass the most highly used runways, and still eliminate snow-related delays and benefit travelers. However, given that Boston Logan has, to the best of our knowledge, not shown any interest in installing heated runways or in conducting a preliminary assessment of their potential benefits, it is likely that any net present value calculation they perform would assume a high-cost scenario and would not include the potential benefits we have noted.

The behavior we ascribe to Boston Logan is consistent with public enterprises conducting assessments, especially for potential urban transportation projects, by making assumptions that fit their preferred decision. Often those assumptions overstate benefits and understate costs, with the outcome that the project is completed and incurs large deficits (Winston (2013)). In the case of heated runways, public airports are not making assumptions about benefits and costs with that bias, which fits with their preferred decision.

10. Private Airports' Adoption of Heated Runways in a Competitive Environment

Using plausible assumptions, the NPV analysis is consistent with the observed decision of Boston area airports not to install heated runways. Although we could make assumptions that would justify Boston Logan installing at least some heated runways, naysayers could make a

---

[177] Some of those purchases are likely to be spontaneous and would not have been made outside of the airport.

plausible case against those assumptions. As noted, our finding is intuitively plausible because public commercial airports have little economic incentive to install heated runways.

Privatized airports would have an economic incentive to install one or more heated runways if the investment increased their profits. We assess this possibility and, to account for political considerations, assess whether private airports could find it profitable to install heated runways without reducing airlines' profits and travelers' welfare.

The Yan–Winston Model of Private Airport Competition

We assess the decision of Boston area private airports to install heated runways by extending Yan and Winston's (2014) model of private airport competition. The model is based on the following three-step sequential moves game. We apply two versions that make different assumptions about airports' runway charges. In the first version:

— Stage 1: The three Boston area airports simultaneously make decisions on adopting heated runways, but they do *not* change their landing fees.

— Stage 2: Airlines choose capacity (number of seats) on their spoke routes that are connected to the three airports.

— Stage 3: Airlines set profit maximizing fares for their flights, which allocates passengers' demands among airlines.

In the second version of the game, the three Boston area airports simultaneously make decisions on adopting heating runways and set profit-maximizing runway charges.

Airlines take current runway charges as given in the first stage and engage in price and service competition, where flight frequency, which determines schedule delays, is the key service variable.[178] Travelers choose among possible airport-airline routing combinations for domestic

---

[178] Schedule delay is the difference between the traveler's desired flight time and their actual flight time.

trips that originate or terminate in one of the three Boston area airports. A traveler's choice is a function of a vector of prices, schedule delays, and other exogenous influences. In the second stage, given aircraft size, airlines choose spoke capacities thereby determining schedule delays. Finally, given spoke capacities, airlines engage in Bertrand competition, as each airline simultaneously chooses prices for the products it offers in Boston area markets by maximizing profits subject to the seat capacity it provides to satisfy demand.

New Estimates of Demand and Costs

We update our original 2014 analysis by using data from the US Department of Transportation's DB1B for the fourth quarter of 2015 and the first quarter of 2016 to estimate an aggregate discrete choice model of travelers' airline choices to and from airports in the Boston area, instead of to and from airports in the San Francisco Bay Area. The parameter estimates are used to solve airlines' profit maximization problem. We summarize the data for that model in table 10. As in the original analysis, the data include fares, service attributes for the alternative carriers and airports, and airport and airline dummy variables; we also include various sources of travel delay, including weather delay and flights affected by snow.

The estimates of travelers' demand parameters, presented in table 11, have plausible signs and most are estimated with statistical precision. We treat the fare coefficient as a random parameter, assuming it is normally distributed, to incorporate travelers' preference heterogeneity.[179]   Higher fares, more connections, and longer delays reduce the likelihood of a traveler choosing a given alternative, while travelers prefer flights with longer distances but at a decreasing rate. All else constant, travelers prefer the low-cost airlines, United is the reference

---

[179] We estimated the random coefficient aggregate discrete choice BLP model assuming a normally distributed and a log-normally distributed random parameter for the fare coefficient, but we could not obtain convergence with the log-normal distribution.

airline, and prefer Providence Airport, Boston Logan is the reference airport. Based on the coefficients, the implied values of travel delay, $85 per hour, and flight frequency, $11 per hour, are somewhat lower than the values we found for San Francisco Bay Area travelers, but they are still plausible. The estimated elasticity of demand of -1.71 is slightly higher than the elasticity that we estimated for Bay Area travelers.

Finally, we present the estimates of the airlines' cost coefficients in appendix table A1 and we present the airlines' and airports' marginal costs, which can be derived from the model, in table 12. Although several of the coefficients are not precisely estimated, the absolute and relative values of the estimated marginal costs are plausible. For example, the legacy carriers American, United, and Delta, have higher marginal costs than Southwest, a low-cost carrier.

Simulation Results

As noted, we assume two different airport privatization scenarios in our simulations. In the first, each airport is privatized but still maintains carriers' weight-based landing fees, following Federal Aviation Administration guidelines, and concession charges. Given the estimated demand and cost functions and airlines' profit-maximizing behavior, we determine whether any of the airports could increase their profits by installing heated runways.

As shown in Table 13, we find that installing heated runways at the three airports would increase passenger demand and reduce airlines' operating costs, which increases airlines' profit and travelers' welfare. Airports increase their revenue because the additional passengers generate additional landing fee revenues of $2 per seat and additional concession revenues of $10 per passenger. However, the increase in airport revenue for the winter season cannot cover the annualized cost of installing heated runways at any of the airports, so Boston area private airports would not adopt heated runways.

In the second version of the simulations, we assume private airports make decisions on both the level of runway charges and whether to install heated runways to maximize profits. Because the two smaller airports, MHT and PVD, do not have large enough passenger demand even if they adjust their runway prices to support installing heated runways, we consider a feasible privatization scenario in which only Boston Logan airport is privatized, and the other two airports do not change their runway charges and runways.

We also include additional benefits from adopting heating runways by including the time savings from flights that would be cancelled. Based on research by Yan et al. (2016), cancellations and missed connections could cause passengers to incur delays of more than 10 hours, depending on the load factor for their next alternative flights. As a conservative assumption, we assume passengers are delayed 5 hours and value that airport delay time at $85 per hour, based on the findings of our demand estimation. The value of delay time assumption is conservative because travelers might value the delays caused by cancellations and missed connections higher than they value conventional airport delays given the uncertainty and anxiety associated with the effects of such delay on travelers' trips and activity at the destination.

Airport privatization that has occurred throughout the world has been marked by tension between airports and airlines because airports feel they should be able to pass their costs on to airlines, which can pass those costs on to passengers through higher ticket prices. However, this behavior may spark protests by groups representing travelers and airlines and possibly discourage government transportation authorities from privatizing Boston Logan. Thus, to account for political considerations, we consider profit maximization behavior by Boston Logan in its choice of runway installation and airport takeoff and landing charges subject to runway price caps that still allow airlines to raise their fares but that prevent air travelers from being worse off.

Table 14 shows that under this scenario it is possible for Boston Logan to find it profitable to install heated runways; airlines could increase their profits; and accounting for the reduction in travel delays and the increase in airport charges that would be passed on in higher ticket prices, travelers would not be worse off.

Because we have made conservative assumptions about the benefits to travelers from eliminating delays from cancellations and missed connections, it is conceivable their benefits could be higher and airports could charge higher runway fees and be more profitable, or that travelers also could gain along with airports and airlines. In addition, we have assumed that Boston Logan would install heated runways for all its runways, which could lead to excessive costs. As noted, it is possible that Boston Logan could install fewer heated runways while still achieving most of the reductions in snow-related travel delays. The lower costs of runway installation could enable all participants to increase their welfare from this innovation. And like all innovations, technological advance that reduces the cost of runway installation is likely to occur as heated runways are increasingly adopted.

Finally, heated runways could provide environmental benefits. Shen et al. (2016) find that heated runways require less energy and produce less greenhouse gas emissions compared with traditional snow and ice removal systems. Heated runways also would reduce the adverse effect that airport deicing operations can have on the aquatic ecosystem and the quality of finished drinking water (Swietlik (2010)).[180]

---

[180] The effects of climate change on our analysis are likely to be ambiguous. On the one hand, the PDV of benefits of heated runways would decrease if snowfalls in the Boston metropolitan area occurred less common. However, the PDV of benefits would increase if extreme weather events, such as heavy snowfalls, occurred more often.

11. Conclusions

Deregulation of intercity transportation enabled the private sector to show how government policy suppressed competition and innovation in transportation. The development of ridesharing showed how innovations can spur competition that had been suppressed in regulated urban transportation markets. Unfortunately, the absence of private sector involvement may enable government intervention in other transportation services and facilities to continue to suppress the potential gains from innovation.

This chapter has shown that public airports may be neglecting the potential benefits to travelers, carriers, and airports from heated runways. Public airports have no incentive to adopt heated runways because they incur a cost that they could not fully recover, and although travelers would benefit, public airports could not make travelers pay for that benefit. In practice, the Federal Aviation Administration, not public airports, would probably have to take the lead if public airports were to install heated runways, which makes it even more likely that the public sector would neglect their benefits because the FAA has even less incentive than public airports to allocate its funds to cover the expense.

In contrast, private airports would have an incentive to adopt heated runways because they could recover their costs and increase profits by adjusting runway charges. Privatizing airports also could be attractive from a political perspective to the extent that airport authorities could set caps on runway charges that do not disincentivize airports from adopting innovations that make travelers better off.

Typically, government intervention is justified when a service or facility is socially desirable but privately unprofitable. We have revealed an important case in transportation where a government-run airport facility in Boston would not implement a socially desirable innovation,

heated runways, but a privatized facility in the Boston metropolitan area would have an incentive to implement the socially desirable innovation. Importantly, in addition to Boston Logan, it is likely that other major airports in the Northeast, Midwest, and Canada also could benefit air travelers by installing heated runways.
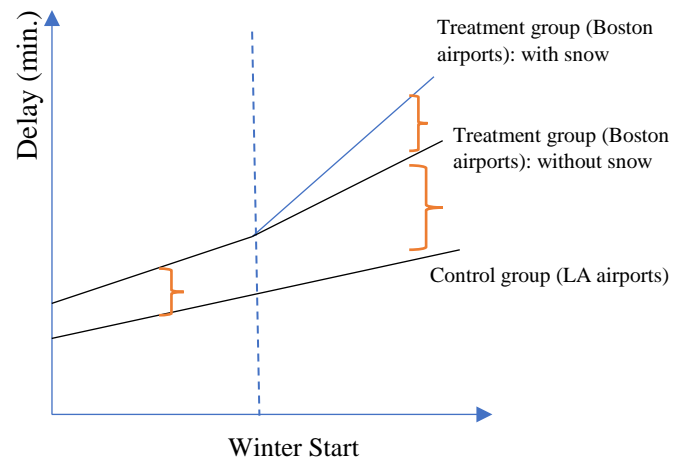
Figure 1. DDD Identification

Figure 2 Decomposing Flight Delays

Table 1. Means (Standard Deviations) for the Delay and Weather Data

Variables	Winter Boston airports	Winter LA airports	Non-Winter Boston airports	Non-Winter LA airports
Departure delay (minutes)	12.82	9.16	8.438	9.04
	(38.96)	(34.31)	(32.80)	(32.70)
Arrival delay (minutes)	10.13	4.88	3.799	5.75
	(43.32)	(36.71)	(35.68)	(34.65)
Extreme weather delay (minutes)	3.71	1.99	2.031	1.17
	(24.18)	(15.20)	(14.43)	(12.28)
Taxi-out time (minutes)	19.76	15.74	17.462	15.47
	(12.18)	(8.15)	(9.05)	(7.75)
Taxi-in time (minutes)	7.19	8.28	6.878	8.02
	(5.15)	(6.13)	(4.33)	(5.93)
Snow (share of days snowing)	0.12	0	0.007	0
	(0.33)	(0)	(0.09)	(0)
Precipitation (inches per season)	0.002	0.001	0.003	0.00
	(0.01)	(0.01)	(0.02)	(0.01)
Temperature (°F)	29.07	63.95	59.87	69.40
	(11.60)	(7.64)	(14.45)	(7.52)
Visibility (miles)	8.76	8.80	9.28	9.52
	(2.79)	(2.32)	(2.10)	(1.36)
Wind speed (miles per hour)	11.08	5.98	10.35	7.89
	(6.00)	(4.41)	(5.13)	(4.86)
Flight distance (miles)	976	1,034	1,015	1,045
	(741)	(816)	(776)	(825)
Flights per day	530	882	558	967
	(242)	(422)	(242)	(456)
Number of observations	60,412	131,475	205,177	426,331

Table 2. DDD Estimation Results

Variable	Model 1 Gate delay	Model 2 Taxi time
Boston × winter × snow dummy	5.108 (0.501)	2.414 (0.101)
Boston × winter	3.432 (0.239)	0.504 (0.048)
Boston	–2.782 (0.382)	–1.019 (0.077)
Winter	–2.013 (0.145)	0.958 (0.029)
Temperature (°F)	–0.043 (0.005)	0.000 (0.001)
Visibility	–0.588 (0.022)	–0.176 (0.004)
Distance (miles)	–0.001 (0.000)	0.000 (0.0001)
Number of daily operated flights at airports	0.013 (0.001)	0.005 (0.0001)
Departure dummy	3.126 (0.076)	7.957 (0.015)
Constant	8.339 (0.545)	7.455 (0.110)
Month fixed effects	Yes	Yes
Airport fixed effects	Yes	Yes
Number of observations	823,546	823,546

Table 3. Matching Estimation Results

	Model 1 $K = 1$	Model 2 $K = 4$
Gate Delay	5.6570 (1.6521)	3.7677 (1.1550)
Taxi Time	3.1735 (0.6696)	3.8365 (0.5103)
Number of observations	60,412	60,412

Table 4. Unit Costs of Common Deicers

Deicer	Approximate Cost	Reference	Application Rate
Sodium chloride (NaCl)	$26/ton $36/ton $26–42/ton $66–79/ton	Zang et al. (2009) Levelton Consultants Limited (2007) Shi et al. (2009b) Rubin et al. (2001)	170–890 lbs/12-ft lane mile (13–68 g/m^2), $0.0003/$m^2$
Magnesium chloride (MgCl$_2$)	$95/ton	Zang et al. (2009) Levelton Consultants Limited (2007)	100–150 lbs/12-ft lane mile (8–11 g/m^2), $0.0002/$m^2$
Calcium chloride (CaCl$_2$)	$294/ton $120/ton $267/ton	Zang et al. (2009) Levelton Consultants Rubin et al. (2001)	Used along with NaCl in the US, $0.03/$m^2$
Calcium magnesium acetate (CMA)	$670/ton $1,280/ton	Zang et al. (2009) Levelton Consultants Limited (2007)	200–500 lbs/12-ft lane mile (15–39 g/m^2), $0.004/$m^2$
Potassium acetate (KAc)	NA	Zang et al. (2009)	0.9–9.1 gal/1,000 ft^2
Salt mixed with Calcium Chloride (NaCl and CaCl$_2$)	$98/ton	Zang et al. (2009)	5–12 gal CaCl$_2$/ton of NaCl, $0.01/$m^2$

Source: Nevada Department of Transportation (2015), p.25.

Table 5. Cost Estimates of Heated Pavement Systems

Source	Installation	Operation	Maintenance	Heat Source
Lund (1999)	$20/ft^2	$3,000	$500	Geothermal system
Minsk (1999)	$48–$70/ft^2	$1.48–$1.54/ft^2		Geothermal (bridge)
	$22–$26/ft^2	$0.98/ft^2		Electric heating(bridge)
Hoppe (2001)	$30/ft^2	$18/h (gas)	$1.74/ft^2, $500/year	Geothermal (bridge)
Anand (2015)	$15–$65/ft^2	$1.12/ft^2		Natural gas

Table 6. Flights and Passengers Affected by Snow at Airports in Boston Area in 2015

	Operations During Snow	Total Operations	Snow Affected Cases	Passengers per Flight
BOS	7,549	218,605	3.5%	89.7
MHT	289	13,078	2.2%	42.2
PVD	768	23,718	3.2%	54.8

Source: Data presented in this table are based on summaries of data collected by the Bureau of Transportation Statistics and the Federal Aviation Administration.

Table 7. Annual Benefits of Installing Heated Runways at Boston Airports

	Benefits to Airlines	Benefits to Travelers	Total Benefits
BOS	$4,542,467	$5,113,598	$9,656,065
MHT	$173,900	$92,099	$265,999
PVD	$462,129	$317,824	$779,953

Note: Annual benefits are calculated based on data from 2015.

Table 8. Installation Costs of Heated Runways at Boston Airports

Airport (# of runways)	Area of Runways (ft^2)	Cost of Construction	
		$15 per ft^2	$45 per ft^2
BOS (6 runways)	7,861 × 150	$17,687,250	$53,061,750
	10,006 × 150	$22,513,500	$67,540,500
	7,001 × 150	$15,752,250	$47,256,750
	5,000 × 100	$7,500,000	$22,500,000
	2,557 × 100	$3,835,500	$11,506,500
	10,083 × 150	$22,686,750	$68,060,250
	Total	$89,975,250	$269,925,750
MHT (3 runways)	9,250 × 150	$20,812,500	$62,437,500
	7,650 × 150	$17,212,500	$51,637,500
	Total	$38,025,000	$114,075,000
PVD (2 runways)	8,700 × 150	$19,575,000	$58,725,000
	6,081 × 150	$13,682,250	$41,046,750
	Total	$33,257,250	$99,771,750

Table 9. Net Present Value of Installing Heated Runways at Boston Airports

	Scenario	Total Net Benefit
BOS	Low cost	$110,958,239
	High cost	–$68,992,260
MHT	Low cost	–$32,489,815
	High cost	–$108,539,815
PVD	Low cost	–$17,027,173
	High cost	–$83,541,673

Table 10. Summary Statistics for the Dataset

Variable	Source	Mean	Std. Dev.
Fare ($)	DB1B	521	349
Passengers per quarter	DB1B	88.882	350.351
No. of connections	DB1B	3.506	0.751
Distance (miles) [1]	DB1B	3,415	1,438
American Airlines	DB1B	0.059	0.236
Alaska Airlines	DB1B	0.007	0.085
JetBlue Airways	DB1B	0.119	0.324
Delta Air Lines	DB1B	0.230	0.421
Spirit Airlines	DB1B	0.006	0.076
Sun Country Airlines	DB1B	0.002	0.047
United Airlines	DB1B	0.099	0.299
US Airways	DB1B	0.100	0.300
Virgin America	DB1B	0.011	0.105
Southwest Airlines	DB1B	0.366	0.482
Departure delay (in minutes)	On Time	11.111	2.239
Arrival delay (in minutes)	On Time	7.521	3.601
Weather delay (in minutes)	On Time	3.265	1.739
Schedule delay (in minutes)	On Time	8.422	40.642
Daily departure (in minutes)	On Time	244	166
Snow affected flights	On Time	0.105	0.070
Passenger share of BOS	DB1B	0.768	
Passenger share of MHT	DB1B	0.092	
Passenger share PVD	DB1B	0.140	
Passengers per market (in thousands)	DB1B	26.7	31.1
Market size (population in million)	Census	3.2	2.1
Aircraft size	T-100	147	17.964

[1] Includes all segments of connecting flights.

Table 11. Random-Coefficient Multinomial Logit Demand Estimates

$\ln S_j - \ln S_0$	Coefficient Estimate	Standard Error
Constant	–13.034	0.562
Fare ($100)	–0.413	0.117
Standard Deviation of Fare Coefficient[a]	0.271	0.259
No. connection	–1.481	0.173
Distance	0.715	0.166
Distance squared	–0.093	0.002
Airport delay	–0.350	0.014
Schedule delay	–0.011	0.011
American Airlines	–1.712	2.67
Alaska Airlines	–1.839	7.976
JetBlue Airways	0.073	0.391
Delta Air Lines	–0.104	0.485
Spirit Airlines	4.915	4.753
Sun Country Airlines	3.183	3.558
US Airways	–3.563	0.478
Virgin America	9.95	2.645
PVD	10.119	1.874
MHT	–7.664	3.08
Implied mean value of delay	$85 per hour	
Implied mean value of flight frequency	$11 per hour	
Implied mean price elasticity	1.71	
Number of observations	17,287	

[a] We assume that the standard deviation for the random fare coefficient is normally distributed.

Table 12. Marginal Cost Estimates ($ per Passenger Mile)

Overall	0.15
By Airline	
American	0.16
United	0.17
Delta	0.14
Alaska	0.18
US Airways	0.13
Southwest	0.13
JetBlue	0.14
Virgin America	0.15
Other	0.16
By Airport	
BOS	0.15
MHT	0.16
PVD	0.14
Number of observations	17,287

Table 13. Benefits of Installing Heated Runways at Boston Airports in Winter Season

	BOS	MHT	PVD
Annualized installation cost (million $ per year)	7.18	3.17	2.77
Change in airport revenue (million $)	4.87	1.02	1.08
Change in airline profit (million $)	22.12	0.08	0.23
Change in demand (seats)	405,223	84,623	89,598
Change in consumer surplus (million $)		16.57	

Table 14. Equilibrium Outcome of Privatizing Boston Logan Airport

	BOS	MHT	PVD
Airport charge ($ per seat)	21	2	2
Heated runway adoption	Yes	No	No
Change in airport profits (million $ per winter season)	16.3	0.00	0.00
Change in airlines' profits (million $ per winter season)	0.8	0.00	0.00
Change in consumer surplus (million $ per winter season)		0.00	

Appendix

Table A1. Estimates of Cost Coefficients

Cost Variables	Coefficient estimate	Standard error	t	$P > t$	[95% confidence interval]	
No. connection	16.926	10.423	1.624	0.104	–3.505	37.357
Distance	75.041	25.138	2.985	0.003	25.768	124.32
Distance squared	–9.49	3.474	–2.732	0.006	–16.3	–2.681
American Airlines	–1.944	30.588	–0.064	0.949	–61.9	58.011
Alaska Airlines	22.8	78.893	0.289	0.773	–131.8	177.44
JetBlue Airways	27.965	24.397	1.146	0.252	–19.86	75.785
Delta Air Lines	–6.869	18.375	–0.374	0.709	–42.89	29.148
Spirit Airlines	33.334	86.576	0.385	0.7	–136.4	203.03
Sun Country Airlines	49.024	137.03	0.358	0.721	–219.6	317.61
United Airlines	5.9626	25.763	0.231	0.817	–44.54	56.46
US Airways	–61.53	24.23	–2.539	0.011	–109	–14.04
Virgin America	8.5303	65.392	0.13	0.896	–119.6	136.71
PVD	13.961	20.295	0.688	0.492	–25.82	53.742
MHT	3.0639	24.819	0.123	0.902	–45.58	51.711
Slot control dummy	46.561	32.941	1.413	0.158	–18.01	111.13
Intercept	–192.9	49.843	–3.871	0	–290.6	–95.25

Note: The omitted category is Southwest Airlines at BOS airport.

Chapter 10. Competition and Innovation to Further Improve
Intercity Passenger Transportation

Clifford Winston and Jia Yan

1. Introduction

If firms continue to compete and innovate, the development of the US transportation system will be a never-ending process. Potential sources of new competition and innovation currently exist that could enable intercity passenger transportation to continue to improve. Of course, the specific providers of transportation are likely to evolve over time as some fail and others become the leaders of a new transportation technology.

As in the past, doubts are likely to be raised about the viability of a new transportation technology and claims will be made that it is over-hyped. For example, supersonic air transportation failed during the early 2000s, but investors believe that it may reemerge in the future. As discussed in the final chapter, it is very difficult to predict when innovations occur and their ultimate success. Current signs point to the development of new intercity transportation technologies that could be commercially viable in the long run.

New Modal Sources of Competition and Innovation

We noted in chapter 6 that autonomous cars could compete with air transportation for certain intercity passenger trips. Autonomous buses and passenger rail service also could compete for passengers. Explorations have begun to automate air transportation and even to create autonomous flying cars! Still another potential improvement in air transportation is commercial supersonic service.

We have identified the policymakers who are responsible for guiding the nation's adoption of autonomous vehicles. In addition, the Federal Aviation Administration (FAA), which is

responsible for managing and operating the nation's air traffic control (ATC) system, could be faced with additional air traffic—from flying cars to supersonic planes—that could add more stress to an already stressed ATC system. Thus, government agencies also will face the challenges of facilitating the operations of new innovative services to enhance the intercity passenger transportation system.

Autonomous Surface Passenger Modes. People who dislike flying or who are traveling with several family members or friends may find autonomous cars and buses attractive for long-distance intercity trips. Autonomous cars could be designed with sleeping compartments and entertainment centers for such trips, while autonomous buses could improve the cost and service of bus transportation.[181] Automation would eliminate a major advantage that the intercity bus has over the automobile because automobile travelers would no longer have to drive. So, intercity bus would be challenged to find a niche in the autonomous intercity passenger transportation market.

Similarly, intercity passenger rail would be severely challenged by automation. President Biden's infrastructure bill includes some $66 billion for Amtrak to repair its aging infrastructure and to expand its routes across the country.[182] However, like bus, automation would cause Amtrak to lose a major advantage it has over the automobile. Commuter rail systems, even if automated, also would face the same problem in retaining passengers.

---

[181] In 2006, the deregulated intercity bus sector began to expand service nationwide at a fast rate with the emergence and growth of Megabus, a new low-cost operator owned by the successful British company Stagecoach Ltd., and new East Coast and West Coast operators. Other services emerged including Greyhound's Bolt Bus and the so-called Chinatown buses that connect Washington, D.C., and New York City and other origin-destination pairs.

[182] Amtrak's efforts to expand its route coverage conflict with rail freight operations when Amtrak must access track that is owned by private freight railroads. For example, Amtrak seeks to provide new service along the Gulf Coast, but negotiations with private railroads have stalled over a disagreement about the cost of the capital investment in the rail infrastructure that is necessary to restore Amtrak's passenger service. Because the railroads have not agreed to give Amtrak access, the Surface Transportation Board (STB) must intervene to determine whether additional passenger service will significantly impair rail freight service. The STB recently ordered Amtrak, CSX, and Norfolk Southern railroads to engage in mediation over the matter.

Flying Cars. Flying cars have evolved into a potentially viable mode as indicated by the FAA now referring to them as Advanced Air Mobility (AAM). Recently, Klein Vision's "AirCar" completed a 35-minute test flight with a fixed propeller aircraft transforming from aircraft to car in less than 3 minutes. It was awarded an official Certificate of Airworthiness by the Slovak Transport Authority after completing 70 hours of "rigorous flight testing." Flying cars are attracting investors in various parts of the world to help develop and test the technology.

eVTOL Craft. Several companies are attempting to develop a new generation of electric-powered vertical take-off and landing (eVTOL) craft for use as air taxis in highly populated corridors, which could save people considerable travel time compared with surface passenger transportation modes. Joby Aviation, a leading industry participant, recently announced the completion of a 150-mile flight on a single electric charging, new plans to operate an eVTOL in 2024 as an air taxi that is cost competitive with taxis on the street, and an agreement with the Japanese airline ANA to establish air taxi service between Japanese cities and Japanese airports. Lilium and Archer also are leading industry participants; other eVTOL companies include AeroMobil, GM's Cadillac Vertile, Kitt Terrafugia, Kitty Hawk, PAL-V Liberty, Pop.Up Next, and Vertical Aerospace.

US airlines also have shown interest in eVTOLs as American Airlines has confirmed delivery slots for the first 50 eVTOL aircraft of an initial pre-order of up to 250 aircraft, with an option for up to 100 more. United Airlines has put a deposit on its order with Archer Aviation for a fleet of 200 eVTOL aircraft to be used in partnership with Mesa Airlines, a regional carrier, and has invested $15 million in Eve Air Mobility, a start-up manufacturer, and signed a conditional agreement to buy 200 of its aircraft with an option to buy 200 more.

The private sector's active interest in a new mode and in overcoming the challenges of designing and building air taxis and flying cars that work is encouraging. However, one must not lose sight of the infrastructure—namely, safe and convenient places for takeoff and landing—that are essential for the mode's commercial success. In addition, those operations must be safely and effectively integrated into the current Federal Aviation Administration air-traffic control systems, which as pointed out below, have not been modernized to reduce ongoing air traffic congestion and delays.

The FAA also is responsible for certifying eVTOL aircraft to meet safety and noise standards and to provide the eVTOL industry with a clear certification process. The FAA has provided Joby Aviation with a certification that would allow it to begin its air taxi operations with a conventional airplane. However, Joby still has regulatory hurdles to clear before its five-seater eVTOL aircraft can legally fly passengers. In May 2022, the FAA announced it would certify eVTOL aircraft that will not need runways to operate under a special "powered-lift" category, rather than as an airplane. The change will require a rulemaking and a new set of guidelines for operators to move forward, meaning that pilots will have to obtain certification as a powered lift operator because they are not flying a traditional plane or helicopter. Thus, like other modal innovations, eVTOLs are facing regulatory hurdles that will delay their widespread adoption in the United States.

SST Air Travel. Supersonic (SST) air travel came to a halt in 2003 a few years after the Air France Concorde flight from Paris to New York burst into flames shortly after takeoff. However, the aviation sector is currently attempting to revive commercial supersonic aircraft that does not cause sonic booms when flying over land and can meet environmental and safety standards

throughout the world. Such aircraft could enable, for example, travelers on the East Coast to make transatlantic day trips and to fly to Tokyo in 6 hours.

The Canadian business jet manufacturer Bombardier announced that it has successfully tested a smaller private jet at supersonic speeds. Boom Supersonic hopes to have a supersonic jet that can provide business class service in just over three hours from New York to London by 2029. United Airlines has a firm order for 15 of Boom's supersonic airliners once the model meets United's safety, operating, and sustainability requirements, and the deal includes options for 35 additional aircraft. UA plans to configure the aircraft to seat roughly 80 business class passengers and charge fares that are comparable to current business class fares. Japan Airlines may purchase up to 20 aircraft. Economically viable and safe supersonic airline service that meets noise and environmental regulatory standards would be a breakthrough for air travel.

The incremental automation of commercial aviation can reduce the cost of air travel by, for example, enabling aircraft operators to relax the minimum crew requirement and to operate with one pilot instead of two (Zakharenko and Luttmann (2022)). Additional improvements could occur by modernizing air traffic control.

Modernizing Air Traffic Control

A modernized air traffic control system is essential to enable the current and potentially new users of air space, from flying cars to supersonic planes, to operate safely.[183] The US Federal Aviation Administration has long been criticized for its failure to develop and operationalize a technologically modern ATC system, which could reduce traffic delays and enhance safety.

Nav Canada, Canada's corporatized ATC system, is seen as a model for the United States to follow to improve its efficiency, technology, and effectiveness. Nav Canada and NATS, the

---

[183] Some of this material synthesizes aviation newsletters on the US aviation system and comments by Robert Poole of the Reason Foundation.

UK's air navigation system, are responsible for the operational benefits from faster flying speeds and greater safety facilitated by Automatic Dependent Surveillance—Broadcast, ADS-B, in the North Atlantic airspace.

ADS-B was originally designed as an alternative to radar to keep better track of aircraft in flight. It updates the position, altitude, and velocity of aircraft much faster than radar. Given the system's constellation of 66 satellites, an aircraft can be under surveillance by as many as seven satellites at a time, providing considerable redundancy to ensure safety. However, because ADS-B relies on GPS, FAA and counterparts overseas no longer consider that it will completely replace radar, which, along with some ground beacons, will remain in place for the foreseeable future

Aireon made a breakthrough by developing ADS-B, which provides global space-based coverage for the 70% of global airspace where there is no radar—the oceans, polar regions, and many mountainous areas. Nav Canada was the lead investor in Aireon and was subsequently joined by several other air navigation service providers (ANSPs), including UK's NATS. Many other ANSPs now subscribe to Aireon's services. However, FAA does not, and it has not implemented ADS-B over the Pacific Ocean airspace, depriving travelers from the United States to Hawaii, Asia, Australia, and other destinations the benefits of a modern ATC system.

In another case of the FAA's poor performance, it has been in an interagency battle with the Federal Communications Commission about the roll out of new 5G cell service near airports that could interfere with low-altitude aircraft operations such as take-offs and landings. Newer radar altimeters (in most large airliners) have adequate shielding against stray signals from the 5G antennas to prevent interference. But older Airbus A320 series planes and most of the regional jets produced by Canadair (now Airbus) and Embraer do not have such filters. FAA is now on a crash program to get those planes equipped, which is causing havoc among the regional carriers

that serve smaller airports. However, FAA waited about a year, knowing of the likely problem before raising the issue with FCC on the eve of its spectrum auction. It is currently estimated that it will take at least five years for all altimeters to be upgraded to allow 5G cell service near airports.

Summary

Thanks to airline and bus deregulation, the intercity passenger transportation system has significantly improved in recent decades and, as we have discussed, new sources of competition and innovation offer additional opportunities for the intercity passenger system to provide faster, more reliable, and less costly service.

Government will again play a critical role to enable the new sources of competition and innovation to move forward. Unfortunately, policymakers continue to impede further competition in air transportation with regulations that fail to grant cabotage rights to foreign airlines and that maintain public ownership of airports. At the same time, policymakers are slowing the adoption of innovations in autonomous surface passenger modes, auto, bus, and rail, which can provide intercity trips, and the FAA's failure to modernize ATC facilities could jeopardize the introduction and performance of new aviation services.

It is encouraging that the private sector continues to engage in competition that is leading to new innovations in autonomy and aircraft, some of which may be technologically and commercially successful and may greatly advance intercity passenger transportation. However, government must provide effective policy guidance and manage transportation infrastructure efficiently to give the innovators a chance to succeed.

PART III

FREIGHT TRANSPORTATION

Deregulation of the railroad and trucking industries in 1980 significantly improved the efficiency of the US intercity freight transportation system and benefited both rail carriers and shippers. Although the trucking industry adjusted to the more competitive environment with the emergence of advanced truckload carriers, more efficient less-than-truckload carriers, and less restricted private trucking operations, it has recently had difficulty attracting drivers and its disrupted operations have contributed to supply chain shipment delays. Following deregulation, railroad mergers left the rail industry with two major carriers in the east, CSX and Norfolk Southern, and two major carriers in the west, Union Pacific and BNSF, raising ongoing concerns that rail freight transportation is insufficiently competitive, especially for so-called captive shippers of bulk commodities, such as coal and grain.

This part of the book provides perspective on competition in the railroad industry as it has evolved since deregulation to achieve a contract equilibrium, where a large fraction of freight traffic, especially coal and grain, move under contract rates. We analyze duopoly competition in coal markets with long-term contract rates and summarize recent evidence on rail competition and policymakers' efforts to spur more competition by requiring railroads to allow reciprocal switching, where a shipper could potentially receive lower rates from a carrier because another carrier has easier access to its network to offer a competitive alternative. We conclude that government intervention to increase competition in the railroad industry is unlikely to be effective and is not justified.

We then look to the future and argue that freight transportation like other transportation services has the potential to be greatly improved by autonomous operations in all modes that would

increase competition and improve service. We discuss autonomous rail freight services and quantify the potential benefits to shippers from autonomous trucking services, which would help address future supply chain disruptions and strengthen trucking as a competitive threat to rail for bulk commodities. We conclude this part of the book by pointing out that more competition and innovation also could improve water transportation and could lead to a new service, drones, that could further improve the speed and reliability of freight delivery services.

Chapter 11. Railroad Competition and Innovation

Vikram Maheshri, Clifford Winston, Jia Yan, and Scott Dennis

1. Introduction

Following deregulation in 1980 and various carrier consolidations thereafter, most shippers in the United States have only two rail carriers competing for their business, and some have only one. So-called "captive shippers" and the various organizations that represent them complain that rail rates are not always reasonable and that the Surface Transportation Board (STB)—the successor to the Interstate Commerce Commission, which has the authority to determine the legality of rates in accordance with maximum rate regulation—does little to protect them.[184] In response to such charges, Congress has periodically considered legislation to increase rail competition and the STB has considered mandating reciprocal switching to ensure that an alleged captive shipper has access to an additional railroad.[185]

In this chapter, we assess the evolution of competition in the railroad industry since deregulation. A key feature of such competition is that prices are often determined by long-term contract rates, especially for shipments of coal and grain, which are negotiated between a shipper and a railroad. We determine empirically the competitive equilibrium that develops in the specific case of duopoly competition between Burlington Northern (BN) and Union Pacific (UP) for coal shipped from the Powder River Basin in the western United States to a power generating utility, which negotiates long-term contract rates with one of the two carriers. We

---

[184] Under maximum rate guidelines (49USC 10707 (d) (2)), shippers can challenge a rate if it exceeds 180% of variable costs and if the railroad in question has no effective competition. However, the process is lengthy and costly.
[185] In addition, following a public hearing on *Urgent Issues in Freight Rail Service* held on April 26 and 27, 2022, the STB announced that freight railroads will be required to provide additional data and regular progress reports on its service. For their part, the railroads reported that they are making considerable progress to hire more train crew members to improve service. The Federal Railroad Administration has concerns with rail safety and is seeking to re-impose a rule that would require two-person crews to operate freight trains.

find that a rail carrier's entry into a monopoly market causes the time path of coal transport rates to approach a plausible estimate of the long-run marginal cost of rail service, suggesting that a duopoly in Powder River Basin markets provides sufficient competitive pressure to substantially eliminate the price markups that one might expect in this competitive setting.

Given that we must restrict the sample in our empirical analysis to 1998 because by 1999 electric utilities began to win a handful of maximum rate cases before the STB, which could indicate that reductions in coal rates could be attributable to residual regulation, we extend our assessment of rail competition by providing descriptive evidence. Again, we conclude that the evidence suggests that the railroad industry is sufficiently competitive, and that government intervention is not justified.

Finally, we provide an overview of autonomous railroad operations, which will reduce rail costs and rates to shippers and improve rail transit time and its reliability. Autonomous rail operations could lead to a seamless North American railroad network, with service to the US, Canada, and Mexico, and will be essential to enable rail to compete effectively with autonomous trucks.

2. An Overview of Rail Competition in Duopoly Coal Markets

Coal from the Powder River Basin (PRB) in Wyoming and southern Montana burns cleaner than most coal mined in the United States because of its lower sulfur and ash composition. It is extracted by roughly 20 mines from a single, vast, 70-foot-thick deposit at constant returns to scale, has homogeneous characteristics, and generally at a uniform mine mouth price. Demand for PRB coal increased substantially between 1988 and 1997 (figure 1) because the 1990 amendments to the Clean Air Act required electricity-generating plants to reduce their emissions. By switching

to PRB coal, a plant can remove one ton of sulfur dioxide for $113, whereas a plant burning coal from the eastern United States must spend $322 to remove one ton of the pollutant by installing scrubbers.[186]

In our analysis, a rail transportation market is a route consisting of a coal producing region with multiple mines at the origin and an electric utility plant at the destination. Utilities purchase coal from one or more mines but they do not co-locate, so only one generating plant is at each destination. Because virtually all PRB coal shipped to electric utility plants moves by rail for most of or the entire journey, railroads do not compete directly with trucks or barge transportation in those markets.

In the late 1970s, Burlington Northern (BN) began transporting substantial amounts of coal from the Powder River Basin and enjoyed a monopoly in many but not all markets. Union Pacific (UP) had a monopoly in a smaller set of markets, and the remaining markets were solely served by Kansas City Southern Railway or another carrier. Competition between BN and UP in particular began to intensify in 1985 and gave a growing number of plants alternative access to PRB coal, as UP and its partner, the Chicago and North Western (CNW) Transportation Company, received authorization from the Interstate Commerce Commission (ICC) to build into the southern portion of the region. Those carriers also engaged in a joint venture to acquire a share in the main north-south track through the PRB, which had been formerly owned solely by BN. UP eventually acquired CNW in 1995 and Southern Pacific Railroad (SP) the following year. Burlington Northern began to enter UP's monopoly markets during the 1980s and was aided in the following decade in a few cases by trackage rights it received as part of the UPSP merger. BN acquired Atchison, Topeka, and Santa Fe Railroad in 1995.

---

[186] Those figures are from *Coal Age*, volume 104, August 1999.

Given that PRB coal is very similar throughout the basin, with only slight mining cost differences, utility plants that were served by UP and BN had two alternative sources of PRB coal. However, because almost all PRB coal is transported under long-term contracts, the full impact of UP's and BN's entry on prices was not immediate; instead, it developed as power plants' contracts with both carriers expired and plants renegotiated their contracts in a duopoly market, where UP and BN provided *direct* rail transportation of PRB coal from a mine to the plant.[187] To provide that service, those carriers operated a joint line that enabled them to ship coal from PRB mines to their network; each carrier then transported coal on one of its mainline routes and delivered it directly to a plant on a section of track known as a *buildout*, which connected the plant to the carrier's network.

In this study, we define entry into a PRB market as occurring when a carrier can fully provide direct service. Because they operated a joint line from the mines and had extensive networks developed partly through mergers, UP and BN generally consummated their entry into specific markets when they completed a buildout that enabled them to provide direct service to a plant.[188] However, as discussed later, the decision to enter a market was determined by the overall profitability of that market.

Railroad technology is often thought of as characterized by high fixed costs, but entry into additional markets is not unusual or particularly costly in the situation we are studying. Electric utilities often provide their own coal cars (but not locomotives), while railroads are able to redeploy or lease their cars and locomotives for use in other markets. Because the incumbent railroad, BN

---

[187] In 1999, the Dakota, Minnesota & Eastern (DME) Railroad Corporation, indicated an interest in connecting its network to the Powder River Basin. Local landowners, however, have opposed the proposed rail line, and the carrier has been unable to secure private financing and has been turned down by the Federal Railroad Administration for a loan guarantee. The DME was acquired by the Canadian Pacific Railway in 2008, but Canadian Pacific announced in 2012 that it would not try to connect its network to the Powder River Basin.

[188] In a handful of cases, entry was consummated because a carrier secured access to another carrier's track on a mainline route.

or UP in almost all the cases, owned the existing track into an electric utility's plant and could extract most, if not all, monopoly rents, effective entry by another carrier required it to build additional track to connect an electric utility's plant to its rail network. However, the additional track represents a tiny fraction of the distance from mine to plant (in our case, less than 1%). Hence, the cost of a buildout is relatively modest and sunk costs are minimal (roughly 2 to 3% of the expected revenues over the life of the asset) because small sections of track are often sold to short-line railroads, and rails are frequently reused on other parts of a carrier's system or sold for scrap. Finally, given that mining and rail service in the Powder River Basin are relatively new, UP and BN have been able to employ the most efficient operations possible, unencumbered by the older rail infrastructure and outdated technology that railroads have been shedding since deregulation.

In sum, Powder River Basin coal transportation markets provide an appropriate setting for determining the quantitative difference in pricing behavior between monopoly and duopoly railroad competition and to assess the competitiveness of duopoly markets for several reasons. First, the physical characteristics of PRB coal distinguish its supply and demand from other domestic coal markets. Second, utilities that receive PRB coal by rail transportation face either a monopoly supplier or a duopoly of suppliers. Third, an entrant into those markets does not incur notable sunk costs. Fourth, the railroads in question provide a homogeneous service. Namely, the primary carriers, BN and UP, use the same technology to transport coal from the same source over similar routes; carry 10,000 to 15,000 tons of coal in unit trains that do not need to be assembled, disassembled, or switched en route; often use cars that are supplied by the shippers themselves; and offer very similar mean transit times and transit time reliability, while the low value of coal ensures that shippers place little weight on non-transport logistics costs or loss and damage. Fifth,

given that coal power plants tend to have the common technological characteristics of operating at constant returns to scale (Sarica and Or (2007)) and exhausting their potential for technical efficiency (Olatubi and Dismukes (2000)), it seems plausible that shippers' demands for coal are likely to be characterized by the same functional form (up to a scaling factor). Lastly, although rail freight charges are determined through long-term contract negotiations between utilities and rail carriers, this feature does not detract from our analysis of duopoly competition. Utilities negotiate several contracts during any given year, often in response to market fluctuations. Each contract represents a small fraction of the coal transported from the PRB, and the contracts expire at different times.

3. An Econometric Model of Rail Transportation Markets for Coal

We develop a model of demand and supply for coal transportation by rail, where entry affects supply. We then specify a model of carrier entry and jointly estimate the central influences on market prices, tons of coal shipped, and the decision by a second rail carrier to enter a market. Finally, we isolate the effect of a carrier's entry on equilibrium rail transportation price markups by simulating a price path and comparing it with a plausible external estimate of the long-run marginal costs of transporting coal by rail.

Our modeling approach is appropriate because direct estimation of market supply and demand, when possible, remains the surest method to characterize market equilibrium with minimal assumptions on competitive structure. Moreover, we can enrich this framework by incorporating a model of endogenous entry behavior that explicitly accounts for the effect of equilibrium prices and quantities on entry decisions.

Demand

Our empirical analysis is conducted on a panel of electric utility plants. Because rail transportation is derived demand (in this case, it is an input into the final production of electricity), we can follow, say, Friedlaender and Spady (1981), and specify power plant i's cost-minimizing input demand for rail transportation, Q_{it}^D, at time t as:

$$Q_{it}^D = D(p_{it}^D, X_{it}^D; \varepsilon_{it}^D),\tag{1}$$

where p_{it}^D is the price of rail transportation (in dollars per ton-mile), X_{it}^D contains exogenous influences on demand, and ε_{it}^D is an error term.

We include the length of haul from the mine mouth to the plant among the exogenous influences on coal shippers' demand for rail transportation of PRB coal. An electric utility presumably has an inventory policy in which it seeks to maintain a desired stock of coal, and it must receive reliable deliveries by rail to do so. Data on the standard deviation of rail transit time, a common measure of service reliability, are not publicly available at the route level; hence, there is a long tradition in railroad economics of using the length of haul, which is highly correlated with rail service time reliability, to control for this effect (Grimm and Winston (2000)). Greater lengths of haul may result in less reliable deliveries of coal and should therefore reduce the demand for rail shipments of PRB coal.

We also include dummy variables that indicate whether the plant can receive non-PRB coal by rail or water transportation to control for alternative coal sources. Those sources can be observed because major coal-fired utility plants report the modes of transportation, including barge and the terminating rail carriers, which they are able to use for their coal deliveries. Effective competition could be provided by rail or water carriers that can ship non-PRB coal to utilities without using carriers operating from the PRB to complete the shipment. We specify separate rail

and water dummy variables to measure access to non-PRB coal, thereby allowing for the possibility that the modes may have different effects on demand. The availability of an alternative source of coal should reduce the demand for rail shipments of PRB coal.

Because rail transportation of coal is an input into the final production of electricity, we specify the maximum theoretical output (nameplate capacity) of a given plant and the average price of natural gas nationally to capture substitution with an alternative source of energy. We use nameplate capacity because unlike the actual observed levels of electricity generated, capacity is unaffected by the type of coal that a plant chooses to burn.[189]

We include an index to capture the sulfur dioxide emission caps imposed in 1995 to implement the standards set by the 1990 amendments to the Clean Air Act. Those caps were imposed on some but not every plant in our sample, so our index takes on values between and including zero and one, where a value of zero indicates that the plant is entirely restricted from producing SO_2 (i.e., it must reduce its SO_2 emissions in a given year 100% from its 1985 level). The index can be entered directly into the specification because it increases smoothly and continuously, with a value of one indicating that the plant has no pollution cap. It is reasonable to treat the caps as exogenous to the extent that most of the plants in our sample did not engage in SO_2 emissions trading, which began in 1994 (Burtraw (1996)). We examined the transactions data in the Environmental Protection Agency's Clean Air Markets Division business system to check whether any of the 48 electric utility plants in our sample discussed later purchased emissions and we found that only one bought allowances in 1994 and only seven bought allowances in 1998, the

---

[189] Nameplate capacity is a better measure of derived demand than is the price of coal because in our sample it varies widely between plants each year by up to a factor of ten. In contrast, the price of coal varies very little across mines, so its effect on rail demand was statistically insignificant.

last year of our sample. Thus, most plants were effectively bound by their initial allocation of allowances.

Nameplate capacity and natural gas prices should have a positive effect on the demand for PRB coal, thereby increasing the demand for rail transport. The sign of the emissions caps is indeterminate because in response to the caps, plants might reduce their demand for all sources of coal and produce less or substitute cleaner PRB coal for other coal sources and maintain output. The first adjustment would cause their demand for rail transportation to fall while the second would cause their demand for rail transportation from the PRB to rise. A plant also could install scrubbers to reduce emissions; but we indicated that this is likely to be a relatively costly option for plants that use low-sulfur coal. Plant level fixed effects should capture the variation among plants in the use of this option.

The passage of the 1990 amendments to the Clean Air Act increased demand for PRB coal by encouraging all electric utilities to seek lower-cost ways to reduce pollution. Our demand specification captures this effect with a dummy variable indicating the years since the act's passage. This dummy should have a positive sign because the demand for electricity grew while the emissions caps were constant. Finally, we specify a time trend, as well as fixed effects at the regional, utility (some utilities own multiple plants), and plant level to capture any unmeasured influences on rail demand in those dimensions.

<div align="center">Supply</div>

Our specification of supply is based on Porter's (1983) model. In a market for a homogeneous good that is produced by firms facing a demand with elasticity η, profit maximization implies that firm k's pricing behavior can be characterized as:

$$p(1 + \theta_k/\eta) = MC_k(q_k),\qquad(2)$$

where θ_k is firm k's conduct parameter, MC_k is its marginal cost function (which may be the same as or different from other firms' marginal cost functions), and q_k is its output. This relationship holds under both cooperative and noncooperative assumptions about firms' strategic behavior.

Following Porter, we note that equation (2) can be aggregated across firms so that the relationship between market supply price, output, and entry effectively characterizes an industry supply curve. Because we have plausible estimates of marginal costs, we specify the supply price of rail transportation to power plant i at time t as:

$$p_{it}^S = S\big(Q_{it}^S, E_{it}, X_{it}^S; \ \varepsilon_{it}^S\big) \tag{3}$$

where p_{it}^S is the price as a function of the quantity of coal transported, Q_{it}^S, the entry of the second carrier, E_{it}, supply characteristics (including marginal cost), X_{it}^S, and an error term, ε_{it}^S. The supply curve is estimated along with the demand curve and a model of entry discussed below, and the resulting parameter estimates are used to simulate the behavior of prices over time to characterize firm conduct as duopoly competition evolves.

We measure the effects of entry with a dummy variable, treated as endogenous, that indicates whether a plant can receive PRB coal from two rail carriers at time t (it is defined for each time period as 1 if two carriers can transport PRB coal to the plant; 0 otherwise). We can measure this variable because coal-fired utilities identify the rail carriers that provide direct service, as noted previously, and that have physical access to their plants.

We recognize that a new entrant's impact on rail prices may intensify over time because a shipper's coal shipments are generally governed by several distinct contracts with the incumbent railroad that may last for several years. Because we do not know precisely how many contracts a shipper had with an incumbent railroad and when each contract expired, we explored alternative ways to capture the dynamic effects of rail competition. We obtained the best statistical fit of our

model by interacting the rail entry dummy variable with a variable indicating the number of years the second railroad served the market for those plants experiencing entry. The value of this variable was held constant after three years to capture the fact that most contracts will have expired by that time, thereby making it unlikely that the entry effect continues to intensify. Holding this variable constant after five years produced a slightly worse statistical fit. Specifying time dummies to indicate each year since the entry of a second carrier produced an even worse fit, because more endogenous variables were effectively used to capture the same effect. The entry variable should have a negative effect on rail prices unless carriers engage in some form of collusive behavior.[190]

In contrast to some markets (e.g., airlines), potential rail competition is not likely to be a relevant factor in PRB markets. A common definition of a potential competitor in network industries is a carrier that provides service at the origin and the destination but does not offer service on the route (Morrison and Winston (1995)). Because negotiations over contract rates occur only when actual entry is assured by a carrier's commitment to construct a buildout to connect a utility with its network, a carrier is effectively an actual competitor without ever being a potential competitor.

Supply characteristics consist of other sources of competition and rail costs. We include two dummy variables to indicate whether plants can receive coal from outside the Powder River Basin by an alternative railroad or by water transportation. Both measures of source competition should have a negative effect on rail prices.

Route specific marginal costs are generally unavailable. Hence, we estimate those costs using Bitzan and Keeler's (2003) translog railroad cost function, which covers virtually the same years as our study and includes ton-miles of unit train traffic at the firm level as a distinct output,

---

[190] Schmidt (2001) and Grimm, Winston, and Evans (1992) have estimated reduced form rail rate equations and found that an increase in the number of rail carriers in a market decreases rail rates.

factor prices, and operating characteristics. Their cost function results in an expression for marginal cost that yields estimates that vary both across plants through the length of haul variable—they find economies of length of haul meaning that longer hauls reduce the marginal cost of a ton-mile—and over time through an estimated time trend, which captures technical change (productivity improvements) in rail freight transportation.[191] We expect higher marginal costs to increase prices.

Finally, we include a time trend and fixed effects at the regional, utility, and plant level to capture unmeasured influences on prices in those dimensions. Utility level fixed effects differ from plant level fixed effects because they account for potentially greater bargaining power that utilities with multiple generating plants may enjoy when negotiating rates.

<div align="center">Entry</div>

Although the primary providers of rail service in the Powder River Basin, Union Pacific and Burlington Northern, have expanded their networks to serve more utilities, they have not elected to serve all markets where a utility receives PRB coal, suggesting that their entry decisions are endogenous in our framework because they are undoubtedly based on specific market conditions for coal transportation. Generally, entry represents a strategic decision consistent with profit maximizing behavior. We first discuss the relevance of dynamic considerations in modeling rail entry and then specify a model of entry based on a firm's profit function.

A dynamic model of entry may be appropriate in two situations. First, because shippers and railroads negotiate contracts prior to a carrier actually providing service, a dynamic specification of the key determinants of entry (such as prices and quantities) may be necessary

---

[191] This approach may introduce measurement error because it holds variables other than unit train output and length of haul constant at industry means. But this error is likely to be small because output and length of haul account for most of the variation in marginal costs (Bitzan and Keeler (2003)).

unless future values of those variables can be determined from their current values. If current prices and quantities are sufficiently strong predictors of future prices and quantities, then a static model of entry is appropriate, although the effects of entry on prices are likely to be dynamic.

In the case of rail transportation of coal, contracts reflect current and future supply and demand conditions. For example, contracts specifying negotiated rates contain standard cost escalation provisions based on independent regulatory findings, thereby reducing the uncertainty associated with the supply price. Uncertainty is further reduced for a carrier considering entry because the supply price is bounded from above by the charge that an electric utility is currently paying the incumbent rail carrier for service, and it is bounded from below by the charge that will enable the entrant to earn positive profits on the route. Little uncertainty is associated with transportation demand because the large base-load plants in this study run at full capacity. Accordingly, exploratory estimations showed that influences at time t are sufficient statistics for predictions of future influences. For example, the quantity of coal that was shipped in each market tended to be stationary throughout the period covered by our sample unless a second railroad entered the market.

Second, a dynamic specification might be appropriate if railroads enter and exit markets and change their prices frequently. However, as noted, contract rates are negotiated for several years and have standard annual provisions that govern rate changes, and no railroads in our sample exit any of the PRB markets; that is, they never relinquish the ability to serve a particular electric utility.

Finally, for exploratory purposes, we attempted to use our data to estimate a dynamic model of entry, supply, and demand, but we concluded that we needed a much larger sample of plants, with a longer time series of market behavior, and more instances of entry to capture possible

dynamic influences in a reasonable manner. We specified entry in terms of a Bellman equation as a function of the buildout, and dynamic demand and supply functions assuming an AR(1) process. We found that the entry of a second railroad in a market had a modest effect on reducing rail rates, but that the supply elasticity had the incorrect negative sign, and the demand elasticity was considerably less than 1.0, which raises a red flag because a monopoly rail carrier serves many markets in the sample. In contrast, as we report later, when we used our data to estimate a static model allowing entry to have dynamic effects on rail prices, we did not encounter those empirical problems.

We therefore develop a model of entry behavior based on a static profit function. As pointed out by Jia (2008), a convention (by convenience) in the empirical industrial organization literature on entry is to omit price, quantity, and sales variables from the specification because the relevant data are unavailable for most industries, and to assume that firms' profits decline in the presence of rivals. Because we do have data on prices and quantities, we explicitly include those variables in the specification of profit that determines firm entry.

The profit available to a given firm from entering market i at time t is simply the total revenue less the variable costs of providing service and the fixed costs of entry and can be specified generically as $\Pi(p_{it}, q_{it}, X_{it}^E; \varepsilon_{it}^E)$, where X_{it}^E contains observable determinants of service and entry costs and ε_{it}^E captures the unobservable determinants of profits.

Assuming entry occurs only in profitable markets (where $\Pi_{it} > 0$) and that the unobservable term ε_{it}^E is an additively separable determinant of profit, the probability of observing a second competitor in a market can be written simply in terms of the variables that comprise revenue and cost. Namely,

$$\Pr(E_{it} = 1) = \Pr\left(\varepsilon_{it}^E > -\Pi(p_{it}, q_{it}, X_{it}^E)\right). \tag{4}$$

Equilibrium prices and quantities are endogenously determined by our supply and demand equations, so the error terms in those equations indirectly enter equation (4). Because those error terms are unobservable determinants of prices and quantities, they also will be unobservable determinants of profits; hence, ε_{it}^{E} contains components of both endogenously determined revenues and other unobservable exogenous determinants of entry.

Consistent estimation of the carrier's entry decision and its effect on prices requires that at least one element of X_{it}^{E} can be excluded from both the vector of supply covariates (X_{it}^{S}) and the vector of demand covariates (X_{it}^{D}). We meet this requirement by using an observed component of carriers' fixed costs of entry that is uncorrelated with the supply and demand error terms. As noted, a carrier must connect an electric utility plant with its physical network to enter a market that is defined by a coal mine at the origin and the plant at the destination. We capture the fixed cost of this connection by specifying the distance of the buildout from the plant to the closest non-incumbent rail line, which is measured using a publicly available railroad atlas and commercial atlas that depicts rail networks and plant locations. Those costs are a combination of one time set up costs and ongoing annual fixed costs. We expect that as the required buildout increases, the likelihood of a second entrant decreases. But crucially, the required buildout does not affect the supply and demand for rail service, all else constant. Moreover, shocks in supply and demand do not causally affect the buildout because it is determined years in advance by utilities' choices of plant locations and non-incumbent rail carriers' choices of mainline track locations.

4. Estimation

Estimation of our model of demand, (inverse) supply, and entry is complicated by the fact that entry must be specified as an endogenous function of quantity (demanded) and (supply) price.

Because of the nonlinear relationship between the unobservable determinants of supply and demand that affect entry, we cannot assume a specific distribution for the error term in the entry equation. We must therefore take a semiparametric approach to obtain consistent estimates of the effect of entry on prices (Powell (1994), Rothe (2009)). We do so by following Matzkin (1992), using the exogenous measure of the fixed costs of entry, the distance of the buildout, to identify the entry decision of the second carrier. The three equations are then estimated jointly using a generalized method of moments (GMM) approach for systems estimation, which also involves estimating an efficient and heteroskedasticity-robust asymptotic covariance matrix of all the estimated parameters. We present our procedure for obtaining consistent and efficient GMM estimates in the appendix.

We assume that supply and demand have a logarithmic functional form, which is plausible (see, for example, Porter (1983)) and fits the data better than a linear functional form does. We make no assumption on the functional form of the entry condition beyond the fact that the error term ε_{it}^E is additively separable. The system of supply, demand, and entry equations to be estimated can therefore be written as:

$$\ln p_{it}^S = \beta_1^S \ln Q_{it}^S + \beta_2^S E_{it} + \beta_3^S X_{it}^S + \varepsilon_{it}^S \tag{5}$$
$$\ln Q_{it}^D = \beta_1^D \ln P_{it}^D + \beta_2^D X_{it}^D + \varepsilon_{it}^D \tag{6}$$
$$E_{it} = \mathbf{1}(\Pi(p_{it}, q_{it}, X_{it}^E) > -\varepsilon_{it}^E) \tag{7}$$

where all variables are as defined previously.

Identification of price, quantity, and entry is achieved by instrumental variables. Quantity in the supply equation is identified by exogenous demand variables, such as nameplate capacity and SO_2 emissions caps that shift demand, but otherwise do not have a direct effect on price or entry. Price in the demand equation is identified by exogenous cost variables, such as route-level marginal cost.

We identify the effect of entry in the supply equation with the entry specific variable buildout distance, which does not directly affect market equilibrium prices and quantities. Although computational complexities discussed in the appendix prevented us from including other variables that might influence entry, such as measures of variable costs, two points should be stressed. First, our estimates are still consistent. Second, such measures would inevitably involve the quantity of coal traffic shipped (e.g., variable cost per ton-mile multiplied by the number of ton-miles of coal), which we have controlled for in our analysis.[192] Finally, the entry equation is identified by the exogenous supply and demand shifters that are accounted for by the control variables discussed in the appendix.

5. Sample

The data sources for the variables used in our analysis and their sample means and standard deviations are presented in table 1. The data are annual observations. Our empirical analysis is based on the shipping activity of the 48 electric utility plants in operation from 1984 to 1998 that burned an average of at least one million tons of Powder River Basin coal per year. Those plants account for 75% of all Powder River Basin coal shipped by rail during the period. Plants burning less than one million tons were too small to have received unit train service on a regular basis or were engaged in small "test burns" of PRB coal. In either case, such plants did not receive rates that reflected regular unit train service. Almost all plants in the sample burned more than one million tons of PRB coal in all years. The table shows that only 5.7% of the plants receive non-PRB coal by rail and 9.5% receive non-PRB coal by water; thus, most plants rely on PRB coal for

---

[192] A measure of variable cost that would account for the difference between the incumbent carrier's and the potential entrant's length of haul does not appear to be useful because the differences are small, roughly 30 miles or 3% of the average length of haul.

a fuel source. This does not pose a problem here because our interest is in rail competition for PRB coal transportation, the large plants in our sample account for a large fraction of PRB coal shipments, and it is not clear that the smaller plants that are not included in the sample receive a much larger share of non-PRB coal as a fuel source.

We began the sample in 1984 because it was the year before the Interstate Commerce Commission authorized entry into the Powder River Basin. We ended the sample in 1998 because by 1999 electric utilities began to win a handful of maximum rate cases before the Surface Transportation Board, so recent reductions in coal rates could potentially be attributable to residual regulation. However, no utilities in our sample received a lower rail rate during 1984-98 by winning a maximum rate proceeding. Of the 48 plants, 31 were served by a single railroad throughout the sample period and 17 experienced entry between 1985 and 1998. A few of the single-served plants came on line after 1984, thus our final sample consists of 711 observations.

It is important to point out that virtually all railroad coal traffic is transported under private contracts that do not reveal the shipper's rate. Thus, we collected publicly available data from the U.S. Department of Energy, Energy Information Administration, and estimated rail rates for electric utilities as the difference between the delivered price per ton of PRB coal that is consumed at the plant and the price per ton of coal at the mine mouth. The delivered price of coal reported by the utility includes all the costs incurred by the utility in the purchase and delivery of the fuel to the plant (FERC (1995)). The mine mouth price reported by the mine is the total revenue received using the actual F.O.B. rail sales price (EIA (1995)). The price per ton-mile in each market is obtained by dividing the difference between the delivered price of PRB coal to the utility and the price at the mine mouth by the length of haul. Subsequent discussions that we had with

railroad personnel confirmed that our estimates of rail rates in each PRB market were quite close to actual rail rates.

Rail transportation contracts for PRB coal might typically cover three million tons of coal per year for a period of five or more years. However, annual price and quantity data are appropriate to use in our model because rail contract rates have annual variation, minimum annual quantity provisions, and annual quantity discounts. Rail contracts specify a price per ton, which is adjusted annually based on railroad input costs and productivity. Minimum annual quantity provisions limit the impact of a new entrant on rates until the contract expires. Utilities receive annual quantity discounts that are reflected in the annual data on the delivered price of coal. Finally, rail contracts for PRB coal are *not* two-part tariffs.[193] As noted, utilities may negotiate several contracts with a railroad during a year. Thus, data on the market price for a given year consists of a ton-mile weighted average of all existing PRB coal transportation contracts, each of which includes the preceding adjustments to reflect current shipping activity.

The presence of a second rail competitor that can deliver PRB coal and the existence of rail and water source competition that can provide non-PRB coal were obtained from the annual Fieldston *Coal Transportation Manual*, which maintains transportation service by specific rail carriers and barge transportation and coal delivery records for major coal-fired utility plants. As noted, entry of a second rail competitor occurred in a market only when that carrier could provide direct service, a condition that in almost every case was met when the carrier constructed a buildout to the plant.

---

[193] Two-part tariffs are not considered rates under Section 6 of the Interstate Commerce Act, and are therefore illegal in the railroad industry. (See *Grain by Rent-a-Train, IFA Territory to Gulf Ports*, 335 ICC 111 (1969); 339 ICC 579 (1971)). Hence ICC's maximum rail rate guidelines (49 USC 10707 (d) (2)) are defined in terms of markups over average variable costs and not as two-part tariffs that require an entry fee to secure transportation.

Route-specific marginal costs were derived from Bitzan and Keeler's (2003) translog cost function, specified as a second-degree polynomial in logs of the variables with all variables divided by their sample mean (the base point of approximation). Because Bitzan and Keeler's coefficients are estimated at the firm level not at the route level, substituting route-specific output would cause the output variable to be very far from the industry mean used in the translog specification and result in poor estimates of marginal costs. We therefore calculated route-specific marginal costs for each market in our sample by evaluating the marginal cost function at Union Pacific's mean unit train output divided by the industry mean, the actual length of haul in each market divided by the industry mean, and the year. The resulting estimates of route-specific marginal costs were consistent with industry operating expenses per ton-mile, cost estimates from other studies, and evidence in railroad coal rate cases pertaining to the relationship between price and marginal cost.

Figure 2 presents a comparison of average annual real rail rates over time for the 31 plants in our sample that were served by a single railroad throughout the period and that did not experience entry (No Entry) and the 17 plants in our sample that experienced entry sometime between 1985 and 1998 (Entry). Because of improvements in rail productivity in growing PRB markets, real rates declined throughout the sample period for both types of plants;[194] but by the early 1990s rates were lower for plants that experienced entry of an additional railroad. Based on our data, in 1984, price markups above marginal cost were 1.3 cents per tonmile for plants that at some point during our sample experienced entry of another railroad, compared with price markups above marginal cost that were 1.1 cents per tonmile for plants that were served by a single (monopoly) carrier during our sample. By 1998, those margins decreased to 0.3 cents per tonmile and 0.5 cents per tonmile, respectively. Thus, entry tended to occur in the most profitable markets;

---

[194] Bitzan and Keeler (2003) find that productivity improvements reduced railroad costs nearly 45 percent during virtually the same period covered by our sample.

but by the end of the sample, market power was eroded in those markets such that they were less profitable than the remaining monopoly markets.

Figure 3 presents the average annual real rail rate at the time of entry for plants that experience entry of another rail carrier. Plants that experience entry obtain initial reductions of 24% in the year entry occurred (year 0) and obtain much larger reductions of 44% by the third year after entry occurred. This finding suggests that rail entry reduces rates, but that its full impact takes time. Of course, our summary of rail rates in the sample for different market structures and over time does not hold other influences on rates constant, so it cannot isolate the effect of entry. We do so by using our model of rail transportation supply, demand, and entry in Powder River Basin markets to simulate the behavior of rates as market structure changes.

6. Estimation Results

For purposes of comparison, we present in the first two columns of table 2 estimates of the model by a simpler estimation procedure, two-stage least squares (2SLS), which treats entry as exogenous but accounts for the endogeneity of price and quantity as explanatory variables in the demand and supply equations. As noted, we specified a time trend to control for unobserved temporal effects and fixed utility and plant effects, but they were statistically insignificant in the supply and demand equations and their exclusion had little effect on the other parameter estimates so they are not included in the specification presented here.[195]

Generally, the coefficients have their expected signs and are statistically significant. The demand for rail transportation of coal is elastic ($\eta = -1.15$), which is consistent with actual price-cost markups for PRB coal traffic and the high percentage of monopoly routes in our sample.

---

[195] As discussed, route level marginal cost includes a time trend to account for technical change.

Given that rail freight transportation is derived demand, the elasticity also reflects the utilities'
underlying demand elasticity for coal and the (high) share of rail transport costs in the total
production costs of coal.[196] The positive elasticity of rail prices with respect to tons shipped in the
inverse supply equation, 0.13, indicates a modestly upward sloping supply curve at the route level
that is likely to reflect the effect of congestion/capacity constraints on carrier costs.

The estimates derived from the rail competition variable in the (inverse) supply equation
are of central importance for our purposes. We find that the initial entry of a second carrier into a
Powder River Basin market reduces rail rates 15% and that this effect becomes stronger over time,
albeit at a diminishing rate.[197] For example, after three or more years of entry, during which time
many contracts are likely to have expired, entry of a second carrier will have reduced rail rates
30%.[198]

As stressed throughout the chapter, coal shippers negotiate contract rates with railroads
that generally last for several years. According to our findings, contract rates dampen the initial
impact that a new rail entrant has on observed prices. But as shippers' contacts expire, they can
obtain lower rates when they negotiate new contracts, presumably by playing one railroad off
against another. Apparently, carriers have not been successful in reaching a tacit understanding to
prevent such competition from developing.[199] As discussed previously, we did not expect potential

---

[196] During our sample period, the average price paid by a utility for a ton of coal was $42 while the cost of shipping one ton of coal 1000 miles, roughly the average length of haul, at an average rate of 1.7 cents per mile amounted to $17 for the movement. Thus, transport costs account for 40% of the price and presumably the total production costs of coal.

[197] Based on our coefficient, this estimate is obtained by calculating: $\ln(1 + 1\ \text{year}) \times 0.215 = 15\%$. We expressed the persistent effect of rail entry as $\ln(1 + \text{years of entry})$ for up to three years because $\ln(1) = 0$ and $\ln(0)$ is undefined. This specification captures diminishing marginal reductions in rail prices caused by the entry of a second carrier. We explored other functional forms for this variable including a Box-Cox transformation, but this functional specification produced the best statistical fit. We also estimated a model that lagged the initial entry variable, but this did not lead to a better statistical fit.

[198] This estimate is obtained by calculating $\ln(1 + 3\ \text{years}) \times 0.215 = 30\%$.

[199] Scherer and Ross (1990) provide examples in other industries where large buyers play one seller off against another to elicit price concessions.

competition to have an influence on rail rates. If it did, its effect would be partly captured in the time trend or year dummies as UP and BN entered more markets, but those variables were statistically insignificant.[200]

The remaining parameter estimates reflect the workings of standard economic forces. Rail prices reflect marginal costs of production and other sources of competition.[201] We find that an increase in rail marginal cost is slightly more than fully passed on in higher rates, which is consistent with rates that are set as a markup over variable costs. (Recall that rates cannot be challenged unless they exceed 180% of variable costs.) Rail source competition lowers rail rates by 24%, but the effect of water source competition is statistically insignificant.[202]

Utilities demand more coal shipped by rail from the Powder River Basin as their nameplate capacity increases, as natural gas prices rise, and after the passage of the Clean Air Act of 1990. We also find that electric utilities located in the South have a greater demand than other utilities in the country have for coal shipped by rail from the Powder River Basin. Southern power plants tend to face more rapidly growing demand than other plants in the country face, so they may prefer large shipments of coal that can be sent by unit coal trains. The Powder River Basin can accommodate this preference more easily than other coal-producing regions in the country are able to because its mines generate more coal than other single mine mouths. Utilities demand less coal as their distance from a coal mine in the Powder River Basin increases, if they can receive coal from another source by water transportation, and as their sulfur dioxide emission caps become

---

[200] As noted, UP and BN were involved in mergers during the period of our sample; but we did not find that those mergers lead to cost savings or rate reductions that were realized in PRB markets because such effects would be captured in time trends or year dummies, which turned out to be statistically insignificant.

[201] It has been argued that shippers who provide their own rail cars reduce rail costs and thus receive a lower rate. We specified the percentage of a utility's coal traffic that is shipped in its private cars in the inverse supply equation, but it had a statistically insignificant effect on rail prices and is not included here. We suspect that this finding is because most of the plants in our sample ship large shares of their coal in private cars.

[202] Water competition may have a negligible effect on rail rates because almost all plants with river access also have rail source competition and because rail must be used for part of any shipment of PRB coal by water.

tighter (i.e., the index becomes smaller), leaving them to choose between reducing output or purchasing emissions permits.[203] Apparently, maintaining output and emissions by substituting PRB coal for non-PRB coal is a less efficient option than reducing output.

Before turning to the GMM parameter estimates of our model that treats entry as endogenous, it is useful to provide empirical support for using the buildout distance as an instrument for entry in the supply equation. Although there is no test to satisfactorily determine whether buildout distance affects or is affected by supply and demand, we previously argued that current shocks in supply and demand would not affect the buildout because it was determined years in advance by a rail carrier's and utility's plant location decisions. We also did not find that the buildout had a statistically significant effect on demand or inverse supply. Intuitively, that is not surprising because table 1 shows that compared with the length of haul on a route, which influences utilities' demand for coal because it captures the effect of rail reliability, the buildout distance is small and should not affect the reliability of a shipment.

Turning to inverse supply, on routes that do not experience entry, marginal changes in the buildout distance and the associated costs incurred by a potential entrant, especially when annualized, are apparently not large enough to affect the incumbent's price, and on routes that do experience entry, marginal changes in (annualized) buildout costs are apparently not large enough to affect the entrant's price. As noted, entry is influenced by the profitability of the entire route. We provide evidence below that the buildout distance plays a role in that decision, and the entry equation is identified by the exogenous supply and demand shifters that are accounted for by the control variables.

---

[203] Holding rail prices constant, we did not find that the presence of rail source competition had a statistically significant effect on the quantity of coal shipped by rail. The presence of water source competition affects the quantity of coal shipped by rail because it is less expensive to ship coal by water than by rail.

The GMM parameter estimates are presented in the third and fourth columns of table 2, many of which are very similar to the 2SLS estimates. We also found that the time trend and the fixed utility and plant effects were statistically insignificant. Robust estimation of the covariance matrix accounts for error correlation across plants. The most important change in the estimates is that we now find that the initial entry by a second rail carrier reduces coal rates 9% and after 3 or more years rates have fallen 18%, which are still sizable effects but they are statistically significantly different from the 2SLS estimates and suggest that treating entry as exogenous may cause an upward bias in those estimates.[204] That bias goes in the expected direction because a second railroad is more likely to enter a route with more favorable unobservable characteristics, such as lower costs, which means that an entry dummy that is assumed to be exogenous would be picking up the effect of those unobservables on rates.

In figure 4, we indicate how the buildout distance affects the probability that a second railroad enters a route by graphing the findings of a nonparametric regression of the predicted probability of entry (which contains the actual values of the control variables) on the buildout distance. Consistent with our expectations, markets that require a greater buildout for a rail carrier to provide direct service are less likely to experience entry by a second competitor; twenty miles appears to be the threshold distance that virtually eliminates the likelihood of entry.

7. Characterizing Duopoly Railroad Competition

We have argued that rail competition in the Powder River Basin provides an appropriate setting for analyzing duopoly competition because carriers that had monopolies in the region,

---

[204] Using a specification test for nonlinear regressions in Davidson and MacKinnon (1981), we rejected the hypothesis that the coefficients on entry in the 2SLS and GMM specifications are equal at the 99% significance level.

primarily Burlington Northern and to a lesser extent Union Pacific, gradually experienced new entry from each other to supply a homogeneous service. We use the parameter estimates to simulate the effect of a carrier's entry on rail prices for coal transportation. This exercise is complicated by the fact that several other variables besides a carrier's entry could affect rail rates. For example, rates may fall in PRB markets because of exogenous declines in the marginal cost of rail transportation or productivity growth. To isolate the effect of entry, we hold *all* variables except the entry variable at their 1984 levels—that is, before carriers began to gradually enter PRB monopoly markets. We then use the inverse supply and demand equations to simulate *market equilibrium* prices in response to the changes in entry behavior over time.

We provide perspective on the quantitative effect of entry by comparing the simulated price path with a simulated long-run marginal cost of transporting coal by rail. It is reasonable to make this comparison because in PRB markets rail is characterized by constant returns to scale—which allows marginal cost pricing to be financially viable. In addition, shippers and carriers enter rate negotiations with a view toward the long run because contracts typically last for several years. To maintain consistency with the price path, the long-run marginal cost does not reflect changes in rail markets after 1984 that may have affected costs.

The simulated marginal cost is derived from the same Bitzan and Keeler (2003) cost function we used to calculate route-specific marginal costs. However, in the simulation we set the cost function's explanatory variables at their 1984 values, thereby assuming that marginal costs are not affected by entry. This is a reasonable assumption because, as we pointed out, PRB coal transportation markets are relatively new, and the primary rail service providers have been able to employ the same technology and most efficient operations during the sample period.

The price path and long-run marginal cost were converted to 1998 dollars using appropriate indices; the simulated marginal cost of transporting coal in 1984 (in 1998 dollars) is 1.9 cents per ton-mile. Because all variables except entry are held constant in our simulation, marginal cost is not reduced by exogenous technical change that may have occurred after 1984. The simulated marginal cost estimate is plausible because it is consistent with previous estimates obtained by Bereskin (2001) and Ivaldi and McCullough (2001) for unit train operations as well as one obtained by Winston, Grimm, Corsi, and Evans (1990) based on data generated before 1984.

To summarize, the price path captures the effect of a second entrant only; the simulated marginal cost is used for purposes of comparison; and all other influences on rates and costs are held constant at their 1984 values. Accordingly, the simulated price path and marginal costs do *not* represent predictions of actual rail prices and costs; it is appropriate for marginal cost to remain constant during the simulation because no other influences except entry are changing; and one should not expect the simulated rail price path to be aligned with the actual behavior of rail prices in figure 2 because the latter reflects changes in all influences on prices.

Figure 5 presents price paths based on the 2SLS and GMM parameter estimates and shows that in the markets that a second carrier has entered at some point between 1984 and 1998, duopoly railroad pricing behavior has evolved slowly, but it appears that rail prices approach marginal cost, and as indicated, PRB market characteristics lend themselves to this competitive outcome. We have suggested that the 2SLS estimates may overstate the effect of entry by a second carrier on prices and, accordingly, its price path is steeper than the price path based on the GMM estimates, which may depict more accurately the time it has taken to generate intense rail competition. From 1985 to 1994, rail prices declined modestly from their monopoly level. But from 1994 to 1998, they have fallen faster—and by more than in the preceding ten years—as UP in particular has

expanded service to a sufficiently large cohort of plants and as BN has been forced to compete more extensively with UP for traffic because shippers' contracts with the incumbent carrier have expired.

Although our estimate of long-run marginal cost is consistent with other estimates in the literature, one could argue that if we (and others) have overestimated marginal costs, then rail prices may not be declining to the competitive level. However, based on the GMM coefficients, we estimate that 38% of the 54% decline in actual rail prices from 1984 to 1998 could be attributed to the additional competition supplied by a second entrant and, as shown in figure 5, simulated prices were still declining in duopoly markets in the final years of our sample. As time continues to pass following the entry of a second carrier and as contracts continue to expire after 1998, the GMM-based price path would undoubtedly draw closer to marginal cost before stabilizing, while the 2SLS-based price path may unrealistically fall below marginal cost, even if marginal cost were somewhat lower than we estimated.[205]

Indeed, the pervasive use of contract rates in rail freight transportation is probably the most important reason why the transition from monopoly to duopoly resulted in a significant reduction in rail rates. That is, each carrier faces the prospect of getting none of a utility's business for several years unless it lowers its rate in response to a competitor's bid. Given that a typical contract might call for three million tons of coal to be shipped annually for five or more years, a railroad has a lot to lose if it does not compete fiercely for a utility's business and allows the utility to take

---

[205] Both price paths would stabilize shortly after 1998 because the effect of the entry variable was held constant after three years. We also calculated the GMM price path based on a model that specified an entry variable whose effect on price was held constant after five years. The price path based on this model showed a later but steeper drop that nonetheless appeared likely to stabilize close to marginal cost.

its traffic elsewhere. Scherer and Ross (1990) point out that, in general, lumpiness and infrequency of orders limit oligopolistic coordination.

Another factor that facilitates—but does not ensure—intense rail competition is that coal transportation is a homogeneous service. As discussed previously, rail carriers use the same technology to transport coal from the same source over similar routes and offer similar transit time and reliability. A railroad is unlikely to convince shippers that it is providing a different, let alone superior, service. Product differentiation, primarily through advertising, might explain why some oligopolies have been able to maintain high price-cost margins (e.g., Baker and Bresnahan (1985)). Our findings, however, are not consistent with a successful strategy of product differentiation.

Finally, given that BN initially provided rail service for many utilities that demanded PRB coal, UP had less (existing) profits to lose by supplying additional capacity in this market. Moreover, UP could primarily gain revenue from additional traffic by cutting into BN's traffic. In such a "market stealing" environment, the two were likely to end up as intense competitors because it is more difficult to compete in quantities in zero-sum situations.

8. A Current Overview of Competition in the Rail Industry

As noted, we are unable to update our analysis of duopoly behavior in the PRB coal markets because rates in those markets may have been affected by STB maximum rate cases. Thus, we turn to the available descriptive evidence on rail competition to update our discussion of the competitiveness of the railroad industry.

Freight railroads were deregulated in large part because they had failed to earn their cost of capital for several decades and policymakers were concerned that the government might have to take over freight operations like they took over passenger operations by creating Amtrak

in 1970. Policymakers hoped that with more freedom to set rates, cut costs, and abandon unprofitable markets that the rail industry would be able to improve its financial health.

Deregulation led to the greater use of contract rates and as discussed by Grimm and Winston (2000), shippers were able to develop ways to generate competition that helped them negotiate for lower rates, including geographic competition (shipping outputs to different destinations served by different railroads) and product competition (receiving different inputs or products from different origins served by different railroads or modes).

It took many years for rail to shed its regulatory-bequeathed inefficient capital structure, but as shown in figure 6, some 25 years after deregulation, rail's return on investment has covered its cost of capital for more than a decade. As shown in figure 7, rail's return on equity has been broadly comparable in the last decade to the median return on equity of *Fortune* 500 companies.

Rail deregulation accomplished its objective of enabling the railroad industry to increase its earnings, achieve profitability, and attract investment, and little evidence exists that rail is exploiting market power. Although shippers of bulk commodities, coal and grain, have complained that they pay excessive rates in some instances, rail has been subject to product competition in those markets because energy shippers' demand has strongly shifted from coal to natural gas and agricultural shippers' demand has shifted to some extent from grain to other agricultural products. Coal's share of rail traffic originated declined from 30% in 1990 to 13% in 2021; grain's share declined from 6% to roughly 4% during the same period.

In addition, the STB Carload Waybill Sample indicates, in general, that a smaller share of traffic is originated or terminated at stations served by a single Class I carrier now than prior to consolidation during the 1990s. Specifically, 59.4% of Class I originations and terminations

of non-exempt carload traffic occurred at single-served stations between 1992 and 1996, compared with 53.7% between 2015 and 2019 (Baranowski and Zebrowski (2022)). Although it is difficult to obtain confidential contract rate information for specific products, as reported by Baranowski and Zebrowski (2022) based on the STB's recent rate study, average rail freight rates measured in terms of real revenues per ton-mile have decreased 7% from 1992 to 2019 (from 4.7 cents to 4.4 cents per ton-mile).

In sum, the available evidence indicates that competition is sufficiently vigorous in the rail industry to enable carriers to earn a competitive rate of return but not to earn excessive profits and to significantly exploit market power in particular markets. Accordingly, expanding government intervention with the hope of protecting shippers' welfare is unnecessary and may be counterproductive.

Unfortunately, Congress is considering the Freight Rail Shipping Fair Market Act[206], which in the alleged interest of promoting competition and reasonable rates, could be used to, for example, institute a policy that would mandate reciprocal switching where a railroad's competitors would have easier access to its network. Like many well-intentioned policies, reciprocal switching could result in unanticipated costs to shippers and railroads, including a railroad abandoning a market if it must provide reciprocal switching; deterioration in regular and intermodal rail service because of the considerable time it may take to provide reciprocal switching; and possible threats to safety while the process of reciprocal switching occurs. Instead of micromanaging railroad operations and competition, policymakers should have a forward-looking vision toward the railroad industry where it is fully deregulated and preparing for competition in an autonomous freight transportation system.

---

[206]file:///C:/Users/cwinston/AppData/Local/Microsoft/Windows/INetCache/Content.Outlook/UCBMQ0B4/STB%20Reauthorization%20Discussion%20DRAFT%206_29_22.docx%20(1).pdf

9. Autonomous Railroads

Like other modes, railroads are beginning the process of adopting autonomous operations, which would improve service time and reliability, reduce rail costs because carriers could operate with a smaller crew and move continuously, and enhance safety. In 2019, Rio Tinto, a mining firm, put its AutoHaul fully automated train operations into service in Western Australia. AutoHaul shows how rail can exploit its economies of train size even more as it involves the simultaneous operation of up to 50 unmanned trains, each 1.5 miles long and carrying 240 cars of iron ore on a 500-mile journey (Smith (2019)). In the US, an autonomous freight train system built by New York Air Brake was tested in the Colorado desert in 2019.

In 2020, three former SpaceX employees founded Parallel Systems in the United States to provide autonomous electric rail vehicles for shipping freight. Because the railcars are individually powered, they can be joined together to create a larger fleet or split off to multiple destinations with the cargo in hand. It also means that they are more flexible because the "platoons" will not need to have a large amount of cargo before they can start moving and can avoid time-consuming switching yards. In the future, it is conceivable that autonomous trains could lead to integrated North American rail systems with US, Canadian, and Mexican railroads competing and providing service throughout the continent, much like airlines, and eventually like trucking.

Scribner (2021b) points out that autonomous rail operations are not currently prohibited in the United States. However, organized labor has prompted federal and state regulators to pursue minimum crew-size rules (for example, at least two crewmembers) for railroads operating in their jurisdictions. Those rules would weaken an important cost advantage of

autonomous rail operations without being based on any evidence that multi-person crews were safer than a single-person crew.

The Federal Railroad Administration (FRA) has compromised safety by impeding the rail industry's automated track inspection (ATI) technology. The FRA refused to extend Norfolk Southern's pilot program after its original time period lapsed and refused to extend the geographic scope of BNSF's program (Marsh (2022)). Several senators have complained to the FRA, arguing that the data-driven fusion of ATI and visual inspections is producing a superior safety outcome, with track employees' hours being reallocated to verifying and remedying the greater number of defects detected by ATI rather than conducting redundant visual inspections. Like the crew-size issue, labor lobbying appears to be behind slowing progress on the use of ATI.

Like other autonomous transportation networks, autonomous rail operations will have to be coordinated and communications between cars and the infrastructure will have to be effective. In addition, they will have to address concerns about cybersecurity and liability. However, because they own their infrastructure, railroads can begin work on developing a national autonomous rail network without delay.[207]

10. Conclusions

Views about the competitiveness of the railroad industry have changed dramatically over time. In the 1880s railroads were thought to be monopolies that required strict regulation. Starting in the 1950s, rail lost considerable traffic to trucking and earned low rates of returns.

---

[207] Switzerland is moving ahead with underground autonomous cargo delivery, Cargo Sous Terrain, between cities and logistics centers, where automated delivery carts on wheels travel through tunnels and pick up and deliver cargo automatically from designated ramps and lifts. See https://spectrum.ieee.org/cargo-sous-terrain

By 1980, rail was deregulated so it could regain its financial health. Since deregulation, concerns have been raised periodically that rail is using its market power to charge excessive rates to certain shippers, although the rail industry has not been raising its rates over time and has not been earning excessive profits. Throughout this period, rail's operations and technology have greatly improved, which has enabled the industry to achieve financial health and benefit shippers.

Autonomous operations will fundamentally change all the transportation modes for the better. Because rail owns its infrastructure, it will encounter few constraints by government as it shifts to autonomous operations. Rail must exploit that advantage because, as we discuss in the next chapter, trucking's adoption of autonomous operations is well underway and will enable trucks to be an even more formidable competitor to rail.

Appendix

This appendix describes the procedure to estimate our model of supply, demand, and entry by the generalized method of moments (GMM). It consists of obtaining consistent estimates of the demand and supply parameters, followed by initial consistent estimates of the entry equation, and then estimation of the asymptotic covariance matrix. Finally, we simultaneously compute robust estimates of the parameters of all the equations.

Consistent Estimates of the Demand and Supply Parameters

Because the demand and supply equations contain endogenous independent variables, we require instruments to obtain consistent parameter estimates. Let us represent each endogenous variable by $N \times 1$ column vectors, and the exogenous variables (the X^js) as $N \times q^j$ matrices for $j = D, S, E$, where N is the number of observations in the panel data set (that is, N is equal to the number of plants multiplied by the number of time periods T). We construct a matrix of instruments, Z, by concatenating the three matrices of exogenous variables; i.e., $Z \equiv [X^D, X^S, X^E]$. Repeated columns of variables are omitted. Thus, Z is an $N \times q$ matrix, where q is the total number of unique variables in all the Xs. Based on the orthogonality conditions $E[\varepsilon^{D\prime}Z] = 0$ and $E[\varepsilon^{S\prime}Z] = 0$, consistent estimates of the demand and supply parameters are defined as

$$\tilde{\beta}^D = \underset{\beta}{\operatorname{argmin}} \|(\ln Q^D - [\ln p^D, X^D]\beta)'Z\| \tag{A1}$$

$$\tilde{\beta}^S = \underset{\beta}{\operatorname{argmin}} \|(\ln p^S - [\ln Q^S, X^S]\beta)'Z\| \tag{A2}$$

where $\| \cdot \|$ is the Euclidean norm.[208] Initial predicted values of supply price and quantity demanded are likewise denoted $\tilde{p}_{it}$ and $\tilde{q}_{it}$, respectively, and predicted values of the first stage supply and demand residuals are denoted $\tilde{\varepsilon}_{it}^S$ and $\tilde{\varepsilon}_{it}^D$.

<div align="center">Initial Consistent Estimates of the Entry Equation</div>

Because the likelihood of entry is a function of revenue, variable costs, and fixed costs—the first two of which are composed of endogenous variables—we take a control function approach in the spirit of Blundell and Powell (2004). Given consistent predictions of price and quantity, we can use the first stage supply and demand residuals as control variables for endogenous price and quantity. Those variables can control for the endogeneity of price and quantity on entry under the identifying assumption:

$$E(\varepsilon^E \mid Z) = E(\varepsilon^E \mid \tilde{\varepsilon}^S, \tilde{\varepsilon}^D) \tag{A3}$$

The assumption states that deviations of observed entry probabilities from their conditional mean must be the same whether the mean is conditional on the vector of instruments or conditional on the control variables. The assumption is satisfied because the control variables include the full portion of the covariates X^E that is correlated with the instruments Z. Accordingly, entry behavior is identified if we estimate the nonparametric entry equation

$$E_{it} = \mathbf{1}\big(\Pi\big(q_{it}, q_{it}, X_{it}^E, \tilde{\varepsilon}_{it}^S, \tilde{\varepsilon}_{it}^D\big) > \varepsilon_{it}^E\big) \tag{A4}$$

Because ε^E is a residual from a fundamentally nonlinear regression, the relationship between the observed determinants of entry and the control variables must be specified

---

[208] The parameters, $\tilde{\beta}^D$ and $\tilde{\beta}^S$, obtained in this manner are asymptotically equivalent to three stage least squares (3SLS) estimates. Although we do not estimate them efficiently because we do not introduce a weighting matrix in the minimizations, we require only consistent parameter estimates at this stage.

nonparametrically.[209] Estimating equation (A4) is difficult because both the functional form of profits and the distribution of the unobserved ε^E are unspecified. In comparison, a standard parametric discrete choice estimation such as probit would require assumptions that enable Π to be fully specified and an assumption that the distribution of the error is normal. Such assumptions are not sufficient to satisfy the identifying condition (A3); informally, under such assumptions the control variables may no longer be able to "control" fully for the endogeneity of price and quantity (see, for example, Blundell and Powell (2004)).

Matzkin (1992) provides a two-step recursive algorithm to estimate the entry equation with minimal parametric and distributional assumptions. To implement Matzkin's approach, we need only assume that the function Π is continuous, monotone, and weakly concave in the buildout distance and that the error term ε^E is drawn from a density with full support and finite variance. Thus, we identify entry probabilities by mildly restricting the set of profit functions to which Π can belong. In the first step, the relationship between the observable and unobservable determinants of entry choices is taken as given, and a vector of predicted entry probabilities that maximizes the likelihood of entry occurring in markets where entry occurs is computed. This enables us to avoid specifying the distribution of ε^E. In the second step, the relationship between observable and unobservable determinants of entry that maximizes the likelihood function is estimated. This enables us to avoid specifying the functional form of Π. The algorithm is recursive because the first step is embedded in the second step; that is, estimating the relationship between observable and unobservable determinants of entry requires re-computation of the vector of

---

[209] In contrast, if ε^E were a residual from a linear regression, then the relationship between X^E, $\tilde{\varepsilon}^S$, and $\tilde{\varepsilon}^D$ could be specified linearly, and the estimation procedure would be equivalent to two-stage least squares.

predicted entry probabilities at each step, which significantly increases the computational demands of the estimation process.

Formally, estimating the entry equation is equivalent to maximizing the log likelihood function with respect to a vector of entry probabilities, φ, and a vector of parameters, π, that relates the observable and unobservable determinants of entry through a series of constraints that ensure that profits and the observed entry choices satisfy certain basic conditions and that enable our estimated entry equation to be consistent with a profit function that satisfies the assumptions we made above. The log likelihood function is given by

$$\max_{\pi,\varphi} \sum_{i,t} [E_{it} \ln \varphi_{it} + (1 - E_{it}) \ln(1 - \varphi_{it})] \tag{A5}$$

Define the matrix X^{E+}, which is formed by horizontally concatenating p, q, and the two control variables $\tilde{\varepsilon}^S$ and $\tilde{\varepsilon}^D$ to X^E. The first set of constraints

$$
\begin{array}{lll}
0 \leq \varphi_{it} \leq 1 & & \text{for all } i, t \\[2mm]
\varphi_{it} \leq \varphi_{js} & \text{if } \pi_{it} \cdot X_{it}^{E+} \leq \pi_{js} \cdot X_{js}^{E+} & \text{for all } i, t, j, s \\[2mm]
\varphi_{it} = \varphi_{js} & \text{if } \pi_{it} \cdot X_{it}^{E+} = \pi_{js} \cdot X_{js}^{E+} & \text{for all } i, t, j, s
\end{array}
\tag{A6}
$$

ensures that entry probabilities fall between 0 and 1 and that the probability of entry must be higher in markets with higher expected profits, conditional on X^{E+}. The second set of constraints

$$
\begin{array}{lll}
\pi_{it} \geq 0 & \text{for all } i, t & \tag{A7} \\[2mm]
\pi_{it} \cdot X_{it}^{E+} \leq \pi_{js} \cdot X_{js}^{E+} & \text{for all } i, t, j, s &
\end{array}
$$

ensures a consistent relationship between the observable and unobservable determinants of entry assumed in the first stage of the estimation procedure. The two sets of constraints are implied by the assumptions that admit a class of permissible profit functions and they ensure that the entry equation is identified in the solution to equation (A5).

We accomplish the first step of the estimation procedure, maximizing the objective in (A5) over the vector φ taking the vector π as given subject to constraints (A6), using an algorithm described in Cosslett (1983). We express the solution to the maximization as $\varphi^*(\pi)$. We accomplish the second step of the estimation procedure, maximizing the objective in (A5) over the vector π subject to the constraints (A7) and $\varphi = \varphi^*(\pi)$, using an algorithm described in Matzkin (1992). The solution to this step gives us $\widetilde{\Pi}_{it}$, the scaled predicted profits in a market, given the observable variables.[210]

Generally, nonparametric estimation suffers from the "curse of dimensionality"—that is, the inclusion of additional explanatory variables increases the complexity of the estimation process exponentially. Because of the recursive nature of the estimation procedure, the curse strikes twice. Each additional element complicates the optimization with respect to φ in the first step (through the constraints (A6)) and complicates the optimization with respect to π in the second step (through the constraints (A7)). Because the first step of the optimization is embedded in the second step of the optimization through construction of the intermediate function $\varphi^*(\pi)$, the computational complexity is scaled exponentially by the *square* of the number of elements of X^{E+}. In exploratory estimations, we found that the twin curses of dimensionality enabled us to estimate the entry equation only with the variables that ensured consistent estimation: price and quantity, with their associated control variables, and the instrument, specifically, the buildout distance, which is excludable from the supply and demand equations.[211]

---

[210] Predicted profits in a market are scaled according to the empirical distribution of the predicted unobservable determinants of entry. In comparison, the predicted parameter estimates in a probit model are scaled according to the distribution of the normally distributed error term.

[211] Horowitz (2011) describes the computational limitation for nonparametric estimation with endogenous regressors in greater generality.

<div align="center">Asymptotic Covariance Matrix</div>

Given the initial consistent estimates $\tilde{\beta}^D$, $\tilde{\beta}^S$, and $\tilde{\Pi}$, we now turn to robust covariance matrix estimation. Define predicted residuals for the three equations as:

$$\tilde{\varepsilon}_{it}^D = \ln Q_{it}^D - [\ln p_{it}^D, X_{it}^D]\tilde{\beta}^D \tag{A8}$$

$$\tilde{\varepsilon}_{it}^S = \ln p_{it}^S - [\ln Q_{it}^S, X_{it}^S]\tilde{\beta}^S \tag{A9}$$

$$\tilde{\varepsilon}_{it}^E = E_{it}\left[\frac{\hat{f}(\tilde{\Pi}_{it})}{\hat{F}(\tilde{\Pi}_{it})}\right] + (1 - E_{it})\left[\frac{1 - \hat{f}(\tilde{\Pi}_{it})}{1 - \hat{F}(\tilde{\Pi}_{it})}\right] \tag{A10}$$

for all $i = 1, \dots, N$. The functions $\hat{f}$ and $\hat{F}$ refer to kernel density estimates of the PDF and CDF of the distribution of predicted residuals in the entry equation.

Given those residuals, we compute, block-by-block, the familiar White heteroskedastic-consistent systems estimator

$$\Phi = \begin{bmatrix} \Phi_{SS} & \Phi_{SD} & \Phi_{SE} \\ \Phi_{DS} & \Phi_{DD} & \Phi_{DE} \\ \Phi_{ES} & \Phi_{ED} & \Phi_{EE} \end{bmatrix} \tag{A11}$$

Each block of this matrix can be estimated with $\widehat{\Phi}_{jk} = \frac{1}{N}\sum_{i,t} z_{it}z_{it}' \varepsilon_{it}^j \varepsilon_{it}^k$, where the z_{it} refer to the $1 \times q$ rows of the matrix of instruments Z. Denote the corresponding blocks of the inverse matrix as $\widehat{\Phi}^{jk} = \left(\widehat{\Phi}^{-1}\right)_{jk}$, and denote $\tilde{X}^E = -\hat{f}(\tilde{\Pi})$. As with the White estimator for three-stage least squares, the asymptotic covariance matrix of the parameter estimates is given by: [212]

$$\hat{\Sigma} = \frac{1}{N}\begin{bmatrix} X^{S\prime}Z\widehat{\Phi}^{SS}Z'X^S & X^{S\prime}Z\widehat{\Phi}^{SD}Z'X^D & X^{S\prime}Z\widehat{\Phi}^{SE}Z'\tilde{X}^E \\ X^{D\prime}Z\widehat{\Phi}^{DS}Z'X^S & X^{D\prime}Z\widehat{\Phi}^{DD}Z'X^D & X^{D\prime}Z\widehat{\Phi}^{DE}Z'\tilde{X}^E \\ \tilde{X}^{E\prime}Z\widehat{\Phi}^{ES}Z'X^S & \tilde{X}^{E\prime}Z\widehat{\Phi}^{ED}Z'X^D & \tilde{X}^{E\prime}Z\widehat{\Phi}^{EE}Z'\tilde{X}^E \end{bmatrix} \tag{A12}$$

<div align="center">Final Estimates</div>

With the asymptotic covariance matrix in hand, we can now compute robust estimates of the supply and demand equation parameters and the predicted probability of entry, $\hat{E}$. The familiar

---

[212] The construction of the asymptotic covariance matrix for a simultaneous system of equations estimated by GMM is explained in more generality in Greene (2003).

White heteroskedasticity-robust estimator, which accounts for error correlation across plants, for this system of simultaneous equations is given by:

$$\begin{bmatrix} \hat{\beta}_p \\ \hat{\beta}_Q \\ \hat{E} \end{bmatrix} = \frac{1}{N} \hat{\Sigma}^{-1} \begin{bmatrix} X^{S\prime}Z\hat{\Phi}^{SS}Z'P^S & X^{S\prime}Z\hat{\Phi}^{SD}Z'Q^D & X^{S\prime}Z\hat{\Phi}^{SE}Z'E \\ X^{D\prime}Z\hat{\Phi}^{DS}Z'P^S & X^{D\prime}Z\hat{\Phi}^{DD}Z'Q^D & X^{D\prime}Z\hat{\Phi}^{DE}Z'E \\ \tilde{X}^{E\prime}Z\hat{\Phi}^{ES}Z'P^S & \tilde{X}^{E\prime}Z\hat{\Phi}^{ED}Z'Q^D & \tilde{X}^{E\prime}Z\hat{\Phi}^{EE}Z'E \end{bmatrix} \qquad (A13)$$

Figure 1. Share of Coal Used in the United States by Origin
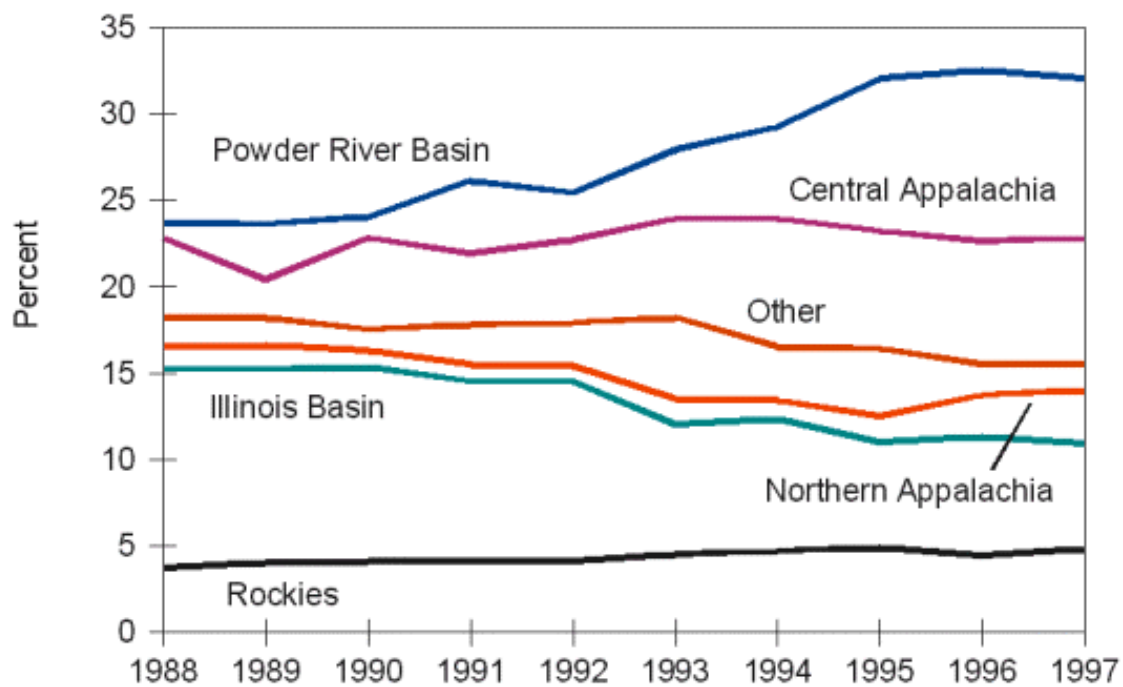
Source: Energy Information Administration, *Coal Industry Annual*, 1994-2000.

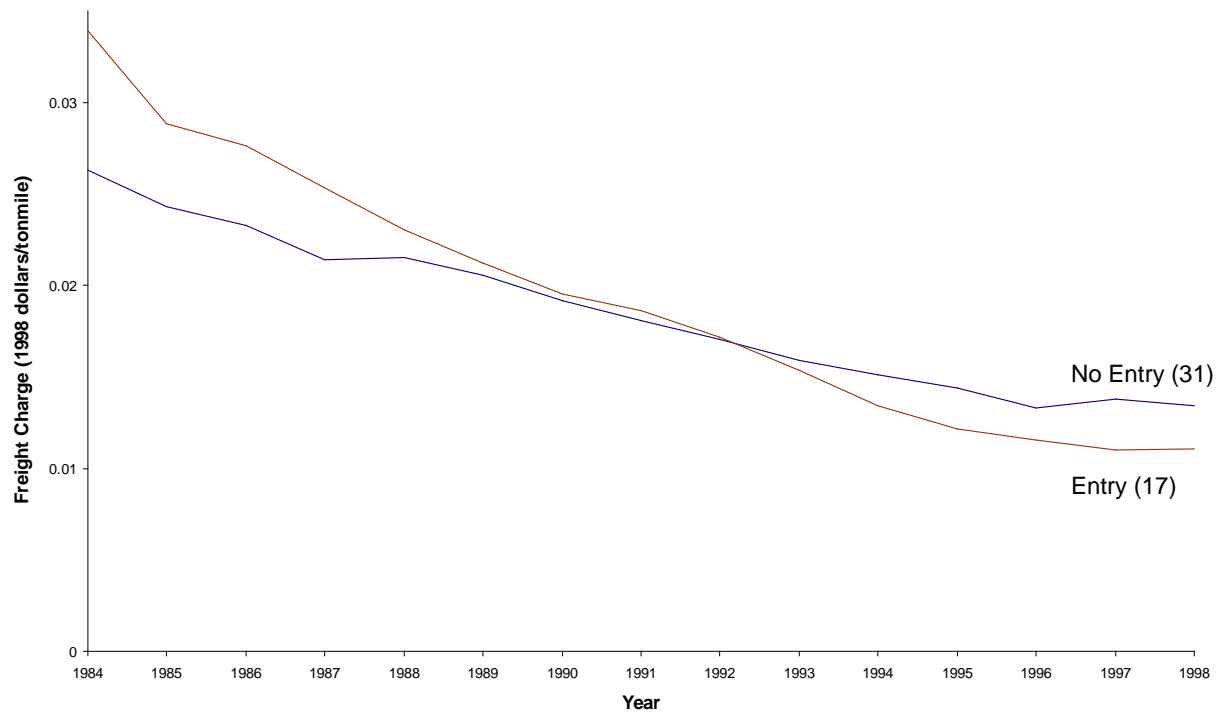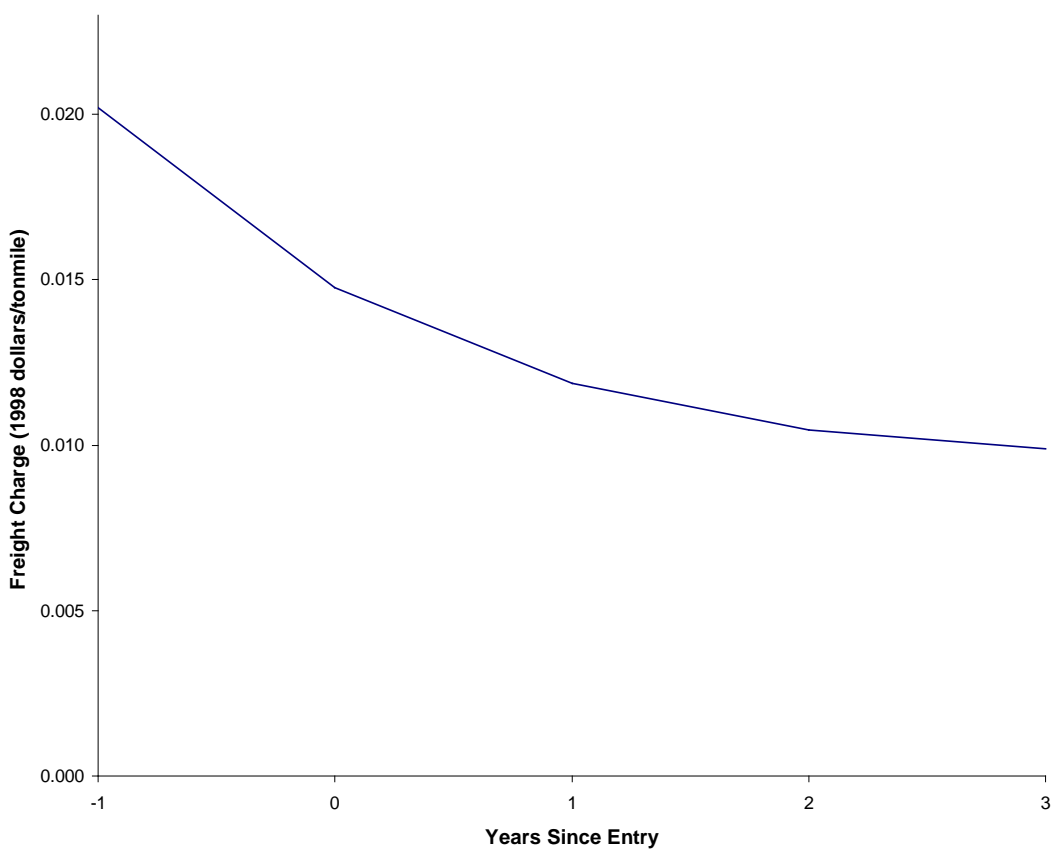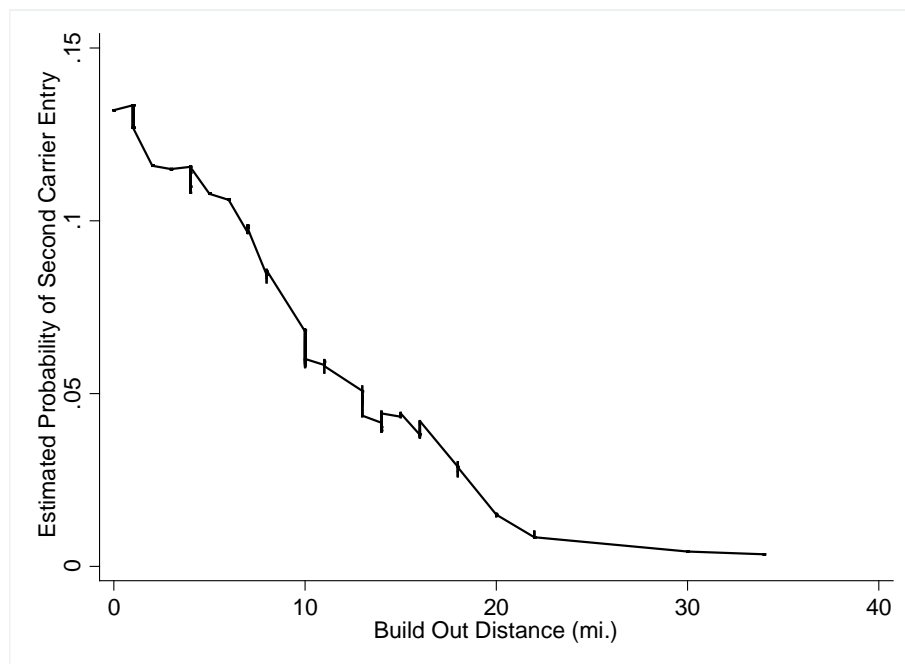Figure 2. A Comparison of Plants' Rail Freight Charges over Time, 1984-1998

Figure 3. Rail Freight Charge Reductions and Years Since Entry

Notes: Entry occurs at time 0 on the *x*-axis; hence, the initial drop from −1 to 0 represents the instantaneous effect of entry on prices.

Figure 4. Nonparametric Estimate of Entry Probability, 1984–98


Notes: Locally weighted least squares regression of the predicted probability of carrier entry on

Figure 5. Simulated Rail Freight Charges and Marginal Cost in Markets Experiencing Entry, 1984–98
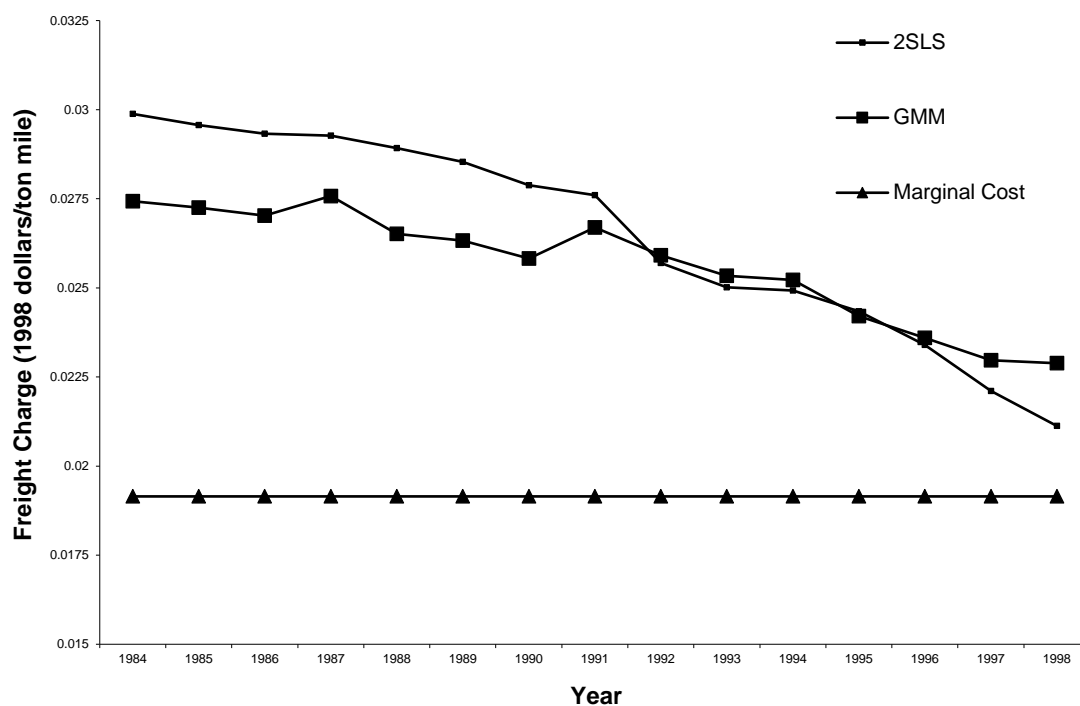

Notes: Rail freight charges are simulated by holding all non-entry variables constant at 1984 values and solving for market equilibrium prices using indicated specifications for supply and demand. We include all markets that experience entry during the period 1984–98.

Figure 6. Rail's Return on Investment Compared with Its Cost of Capital
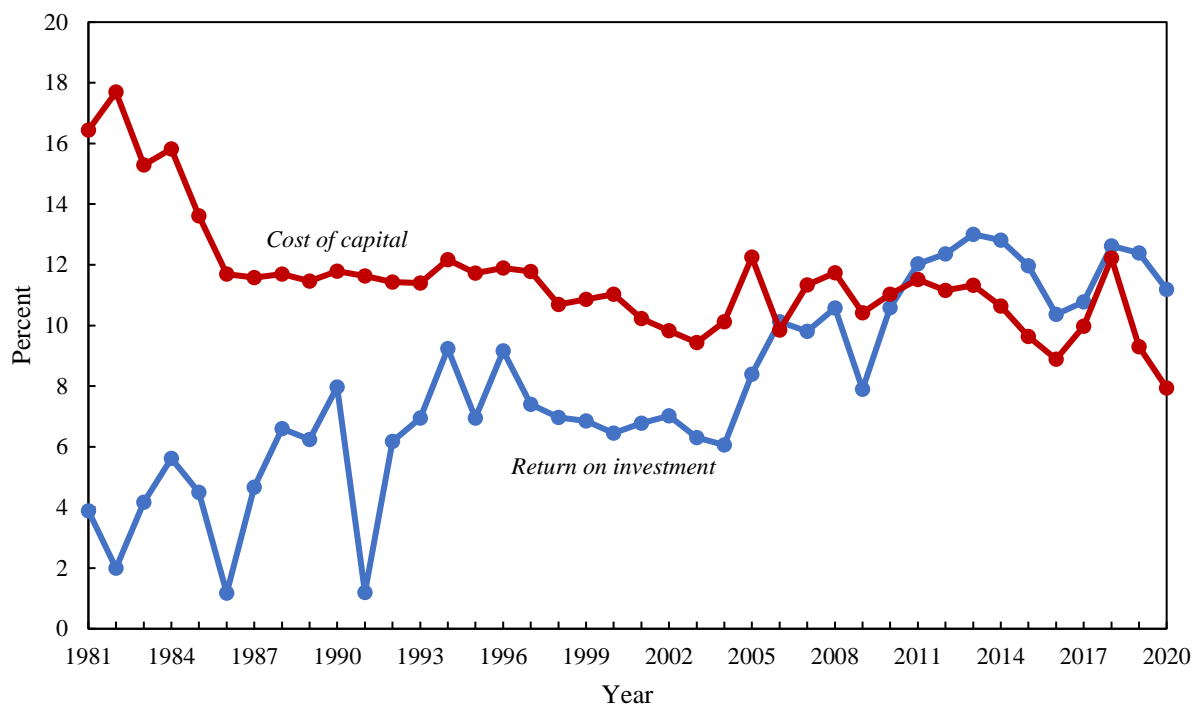
Source: Surface Transportation Board.
Notes: In 2006, the STB significantly changed the method by which it calculates the rail industry cost of capital.

Figure 7. Rail's Return on Equity Compared with *Fortune* 500 Firms' Return on Equity
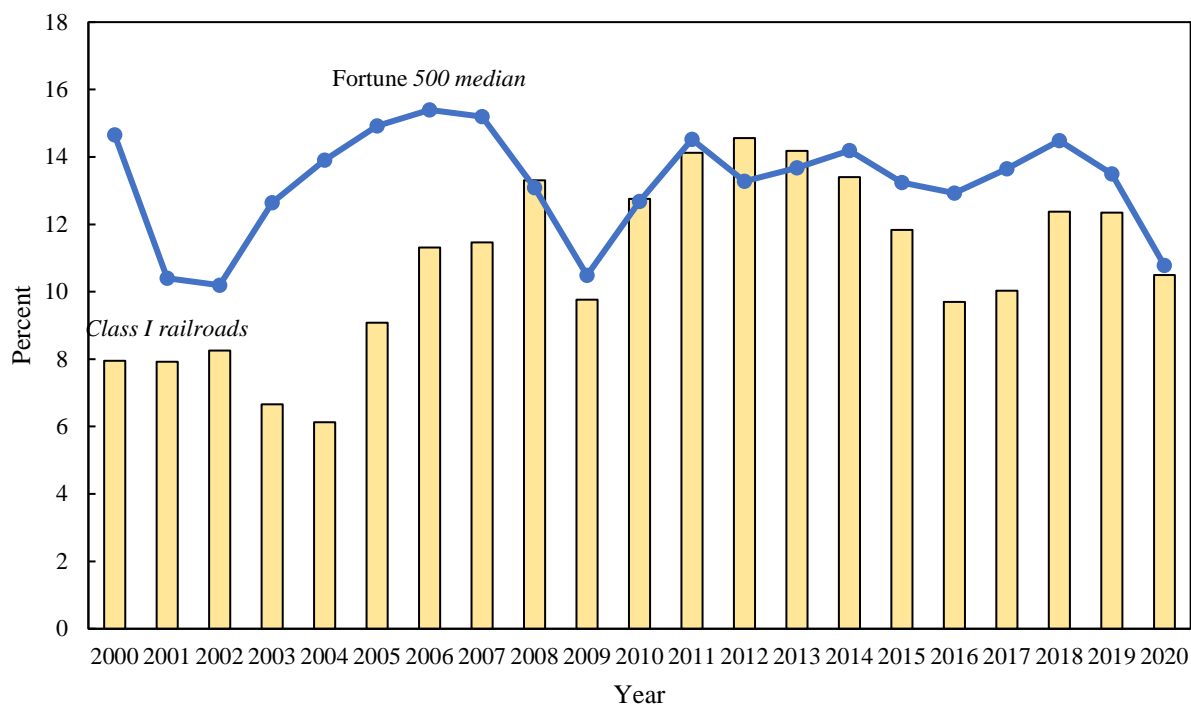
Sources: Association of American Railroads, *Fortune*.
Notes: Return on equity is the ratio of net profit to shareholders' equity.

Table 1. Sample Means and Data Sources of the Variables

Variable	Units	Mean	S.D.	Data Source
Freight charge	$/ton-mile	0.017	0.008	Energy Information Administration, *Cost and Quality of Fuels* and *Coal Industry,* Annual
Tons of coal shipped	1,000 tons	2757	2058	EIA, *Cost and Quality of Fuels*, Annual
Presence of second rail competitor	Dummy	0.068	0.256	Fieldston, *Coal Transportation Manual*, Annual
Rail source competition	Dummy	0.057	0.191	Fieldston, *Coal Transportation Manual*, Annual
Water source competition	Dummy	0.095	0.246	Fieldston, *Coal Transportation Manual*, Annual
Nameplate Capacity	Million KWH	1.172	0.755	EIA, *Annual Electric Generator Report, form EIA-860*, Annual
SO_2 emissions caps	1,000 tons	2.110	13.0	EPA, *Clean Air Act, Title IV*
Natural gas price	$/1,000ft^3	2.449	0.650	EIA, *Historical Gas Annual*, 2000
Route-level marginal cost	$/ton-mile	0.012	0.003	Authors' calculations based on Bitzan and Keeler (2003)
Length of haul from mine to plant	Miles	1024	339	RDI, *Coal Rate Database*, 1997
Length of "Build-Out" for entrant in 1984	Miles	9.204	7.34	Rand McNally, *Railroad Atlas* (1982) and *Commercial Atlas and Marketing Guide* (1991)

Notes: All values are in 1998 dollars where appropriate.

Table 2. Supply and Demand Parameter Estimates, 1984–98

	2SLS		GMM	
Variable	Supply	Demand	Supply	Demand
Freight charge ($/ton mile)*	Dep. Var.	–1.147 (0.327)	Dep. Var.	–1.415 (0.105)
Tons of coal shipped to plant (thousands)*	0.129 (0.020)	Dep. Var.	0.119 (0.007)	Dep. Var.
Supply Characteristics				
Rail competition dummy (1 if two carriers serve the route, 0 otherwise) interacted with number of years a second rail competitor offered service from the Powder River Basin. Three years after entry, the variable stays unchanged.[a]*	–0.215 (0.042)	--	–0.129 (0.013)	--
Route level marginal cost ($/ton mile)*	1.087 (0.055)	--	1.078 (0.022)	--
Demand Characteristics				
Plant Nameplate Capacity (millions of KWH)*	--	0.136 (0.011)	--	0.134 (0.004)
Average national price of natural gas ($/1000ft^3)*	--	0.309 (0.101)	--	0.414 (0.040)
Clean Air Act (1990) dummy (1 if the Clean Air Act has been passed; 0 otherwise)	--	0.299 (0.151)	--	0.367 (0.049)
South regional dummy (1 if plant is located in the South; 0 otherwise)[b]	--	0.944 (0.160)	--	0.959 (0.049)
Plant SO_2 emissions cap index (defined as 1 – (emissions cap)$^{-1}$ for plants subject to emissions caps; 1 for plants not subject to caps)	--	0.381 (0.231)	--	0.439 (0.080)
Shipment Characteristics				
Length of haul from mine mouth to plant (miles)*	--	–0.751 (0.160)	--	–0.959 (0.350)
Source Competition				
Rail source competition dummy (1 if a plant can receive coal from a non Powder River Basin source by a competing railroad; 0 otherwise)	–0.219 (0.079)	--	–0.280 (0.027)	--
Water source competition dummy (1 if a plant can receive coal from a non Powder River Basin source by water transportation; 0 otherwise)	0.025 (0.070)	-1.542 (0.185)	0.050 (0.022)	–1.674 (0.114)
Constant	–5.186 (0.157)	5.658 (1.123)	–5.119 (0.057)	4.394 (1.059)
Number of observations	711		711	

Notes: * denote that the variable has been transformed by natural logarithm. The note [a] indicates we added one to the number of years of entry, ensuring that when we take the natural logarithm, the variable takes on a value of zero for routes without a second competitor. The note [b] indicates that South region includes Alabama, Arkansas, Louisiana , Mississippi, Oklahoma, and Texas. Huber–White robust standard errors are in parentheses.

Chapter 12. The Potential Benefits of Autonomous Trucking

Chad Shirley and Clifford Winston

1. Introduction

Trucking carries roughly two-thirds of the value of freight that is shipped in the United States, and it also accounts for about 80% of freight transportation revenues for primary shipments, according to the American Trucking Association (ATA). Improving the efficiency of trucking operations through automation has the potential to generate large benefits to shippers and the overall economy in several ways.

<u>Characterizing the Benefits</u>

First, autonomous trucks could reduce operating costs by eliminating the driver and greatly improve service times and safety. In recent years, driver wages and benefits have accounted for 44% of the average marginal operating costs of for-hire trucking (Leslie and Murray (2022)). Autonomous trucks also would allow deliveries to be completed more quickly and reliably and would reduce the frequency of accidents and cost of insurance.

Federal hours of service regulations mandate that a driver is allowed to drive a maximum of 11 hours during a 14-hour workday after 10 consecutive hours of off-duty time. An autonomous truck could move a load of cargo nonstop with little risk of an accident. For example, an autonomous truck could travel from Los Angeles to Dallas in 24 hours, compared with the two to three days that it would take a human driver to move the load. The cost per mile of the Los Angeles to Dallas trip is estimated to drop from $1.76 to $0.96 (Picard (2022)). The reduced operating costs and improved service quality would translate into significant price and nonprice benefits for shippers and into lower prices for final consumers.

Nonstop operations and reductions in incident delays caused by accidents also would improve the speed and reliability of freight deliveries. The cost reductions and service improvements would enable businesses to improve their supply chains by holding a smaller share of their raw materials and final products in inventory and by facilitating more precise just-in-time inventory operations. Even small service improvements can generate large annual logistics and inventory cost savings for firms because those costs in the aggregate run in the trillions of dollars (Zimmerman et al. (2021)).

Second, autonomous trucks could enable firms to improve efficiency and expand output by re-optimizing their supply and distribution networks. For example, given faster and more reliable deliveries of certain inputs, which they can purchase at lower cost, some firms might receive inputs from longer distances to produce their products more efficiently and at a higher quality standard. Other firms might expand their sales of perishable products whose value would be preserved because those products would reach their destinations much faster.

Third, the trucking industry would be less vulnerable to labor shortages. The ATA reports that the trucking industry employed 7% fewer truck drivers in 2020 compared with 2019, and in 2021 the industry was 80,000 drivers short to respond in a timely manner to shippers' demand for trucking services—a number that is expected to double by 2030. The ATA also reports that the trucking industry is responding to the shortage of drivers by raising the annual pay for long-haul truckers to almost $70,000 in 2021, an 18% jump from 2019. KLLM Transport has boosted pay 33% for over-the-road truckers and for trainees coming out of its academy, while regional company drivers and independent contractors will receive a pay hike of 10% to 16%. Recently, Walmart announced that it is paying its drivers up to $110,000 in their first year at the company and more once they work longer for the company. At the same time, higher pay increases shippers'

rates and consumer prices. Autonomous trucks could address the fundamental problem that recruiting long-distance drivers is difficult because truck driving is such grueling work.

Finally, by increasing the speed and reliability of deliveries and by mitigating the shortage of truck drivers, autonomous trucks could help ameliorate the cost of periodic supply chain disruptions that delay shipments at ports arriving from abroad and that delay shipments throughout the supply chain.

<div align="center">Recent Developments in Autonomous Trucking</div>

Autonomous trucking operations are steadily evolving and expected to become an integral part of the US surface freight transportation system. Several companies have already attracted funds and are testing automated trucks, including TuSimple, PlusAI Corp, Embark Trucks Inc, Waymo Via, and Aurora Innovation Inc. In a step towards full automation, Locomation is developing truck convoying, which is of interest to Amazon, Walmart, and Target because those operations have a lead vehicle that keeps automated follower trucks with more total capacity moving all the time.[213]

Several companies are focusing on providing automated trucking service between cities. Autonomous trucking companies, such as Embark, and real estate investors, such as Alterra Property Group, are forming alliances to create autonomous trucking hubs so that after carrying its cargo on a highway, an autonomous truck can park at a hub and enable a human driver to deliver the cargo at destinations in the city. Embark plans to provide service on an autonomous trucking lane between Houston and San Antonio, and Aurora plans to open a new 635-mile route in Texas. In addition, logistics firms and autonomous truck companies are collaborating to provide service.

---

[213] Locomation's initial operating plan combines automated platooning with team driving, so the two-truck convoy has two drivers, one actively driving and one in the sleeper berth complying with Federal Motor Carrier Safety Administration Hours of Service rules.

CEVA Logistics is working with Kodiak Robotics Inc., a developer of self-driving trucks, to deliver freight autonomously on routes from Dallas-Fort Worth to Austin and from Dallas-Fort Worth to Oklahoma City. Waymo and CH Robinson (a logistics provider) will use Waymo's test fleet of autonomous trucks to make deliveries between Dallas-Fort Worth and Houston.

Another autonomous trucking company, Gatik, is developing short haul service. Gatik is operating fully driverless trucks to move customer orders between a Walmart dark store in Bentonville, Arkansas (which provides fulfillment for online orders only) and a neighborhood market.

Chapter Overview

The purpose of this chapter is to explore the potential benefits of autonomous trucking to the US economy. Although most of the potential benefits summarized above are difficult to determine quantitatively, it is possible for us to build on previous work to model firm-level, cost-minimizing logistics behavior. We then simulate empirically how firms' logistics behavior would be affected by improvements in the cost, speed, and reliability of freight transportation produced by autonomous truck operations and use those simulations to offer a plausible estimate of firms' annual logistics cost savings.

Using representative product and freight data, we find that the widespread adoption of autonomous trucking could reduce firms' annual logistics costs by roughly $140 billion. Those benefits would be in addition to other benefits generated by autonomous trucking, which we discuss qualitatively. In what follows, we describe our approach to modeling a firm's logistics behavior and formulate its cost minimization problem, indicate the parameters and assumed numerical values that we use for our numerical simulation, report our simulated benefits from

autonomous trucking, and discuss qualifications to our analysis and additional benefits that autonomous trucking could provide.

2. Modeling Inventory Logistics Behavior

We wish to develop a model of logistics costs, where transportation attributes affect those costs and can be adjusted to capture the potential service improvements from the introduction of autonomous trucking. It is difficult to develop such a model that can be used for a numerical simulation because of the absence of detailed data for key variables in the model and the limited information about the values of the model's key parameters. We therefore use an approach to modeling logistics behavior developed by Tyworth and Zeng (1998), as discussed by Shirley and Winston (2004), which has been used with data for numerical analysis and can be updated and modified for our purposes here.

The model is based on an economic order quantity (EOQ) transportation-inventory model that incorporates the effects of freight transportation cost, travel time, and the reliability of travel time on annual logistics costs and service levels. We use an EOQ model because transportation characteristics play an important role by affecting the costs of holding inventories, such as warehousing costs and the costs of the capital, and the costs of not having inventories when they are needed, such as backorder costs and the costs of lost sales. The EOQ model applies to inventories, such as raw materials, that are acquired through movement by a transportation mode. Note transportation characteristics do not play a role in a production smoothing model, which is characterized by increasing marginal costs of production.

The inventory policy used in an EOQ model to minimize logistics costs is called an (S,s) policy because the important values for the firm to decide are the target inventory level (S) and a

reorder point (s). When inventories fall below the reorder point, an order is placed. If inventories remain above the reorder point, they continue to be depleted without being replenished. A depiction of inventory levels over time produces a saw-tooth shape frequently seen in the literature on money demand (for example, Blinder and Maccini (1991)).

Generally, the target inventory level is influenced by a firm's expectations of future sales, the storage costs of inventory, and the interest costs of inventory investment. The reorder point is a buffer stock used to protect against stockouts, which occur when a product is in demand, but none is in inventory. Stockouts force a firm to incur extra costs to expedite delivery of the product (backorder costs), or to lose sales, such as for retail and wholesale inventories or when production is delayed if the product is used as an input. Thus, the reorder point is affected by expectations of future sales in conjunction with the stockout/shortage cost and by the expected delivery time.

The target inventory and reorder point values, S and s, are chosen by a firm to minimize the costs associated with placing orders, the costs from suffering a shortage of the good, and the unit costs of transportation, while trading those costs off against the costs of holding larger inventories. In the model, cost-minimizing firms continuously monitor inventory levels, while product demand is independent from one time period to the next. Transit times (the amount of time involved in transporting an order) are assumed to be independent, stationary random variables to realistically incorporate freight transit time and reliability into the EOQ model.

Logistics costs are comprised of ordering costs, transportation costs, holding costs, and shortage costs. In our analysis:

— *Ordering costs* consist of employees' time and information technology and communications costs.

— Annual *transportation costs* equal the number of units demanded each year and the freight rate per unit.

— *Holding costs* of inventory are complex and depend on a combination of in-transit, cycle, and safety stocks.

- In-transit stocks are those being transported. Their holding costs depend on the holding cost for each expected unit in transit, which include the cost of capital, depreciation, and obsolescence. The expected level of in-transit stocks depends on the mean number of units demanded per day and the mean number of days each shipment is in transit. Transit times, measured in days, are allowed some random variation.

- The remaining holding costs of inventory accrue to units in a warehouse. The units not in transit are either part of the cycle stock or are considered as safety stock. Cycle stock units are those on hand between shipments and are thus influenced by the transit times. Cycle stocks entail the same holding costs incurred by in-transit stocks, plus storage and insurance costs.

- Safety stock units are those still on hand when a new shipment is received. We consider the effects of transportation on safety stock holding costs by letting lead-time demand depend on the lead time and the daily demand. The lead time is directly affected by the time involved in transportation, plus a constant representing a fixed time spent on non-transportation activities in the order cycle, such as order preparation, order processing, manufacturing, or assembly, depending on the context. As with cycle stocks, the holding costs for safety stocks depend on the holding costs for each expected unit stored.

— *Shortage costs* include the costs of any expedited delivery to satisfy customer demand once inventory is depleted, plus the cost of any lost sales or lost customers that result from a shortage. In a raw materials or work-in-progress inventory, shortage costs may take the form of expedited delivery costs, overtime, or lost production. The shortage cost term depends primarily on the expected level of shortage during the lead time between each order and delivery over the year.

3. Product and Transportation Service Parameters

We use our model of a representative firm's logistics costs to estimate economy-wide logistics cost changes from the adoption of autonomous trucking, which affect transportation costs and the mean and standard deviation of transit time. To do so, we use a set of representative firm data for selected industries on product and shipping attributes, representative data on the characteristics of the transportation system, and we assume plausible changes in those characteristics that could be caused by a firm substituting non-autonomous truck transportation for autonomous trucking.

Detailed representative firm data are obtained from a survey conducted by LaLonde et al. (1988). It should be stressed that obtaining the necessary detail about a firm's logistics costs is very difficult. Indeed, discussions with logistics researchers led us to conclude that the LaLonde et al. survey was a one-of-a-kind endeavor; we were unable to uncover any more up-to-date, comparable data that were publicly available. We therefore used the LaLonde et al. original survey values and other standardized logistics cost assumptions and as summarized below, adjust them in different ways to characterize current economic conditions.

With help from the Council of Logistics Management, LaLonde et al. (1988) surveyed shippers in various industries about their logistics behavior. Three hundred and eleven responses were sorted according to nine different industry groupings: Food and Beverage, Chemical and Plastics, Pharmaceutical, Automotive, Paper, Electronics, Clothing and Textiles, Merchandising, and Other Manufacturing. Respondents included manufacturers, wholesalers, and retailers.

The survey asked for data on several inventory logistics and transportation variables, including average order sizes, order fill rates, unit product values, shipment distances, shipment travel times, and shipment travel time reliability. To increase the accuracy of the survey, respondents were directed to select a representative product and a typical customer for that product and to answer many of the questions with that specific product and customer in mind. The authors reported the mean response for each industry group. In a few instances where large outliers would unduly influence the mean, they reported the median response.

Our base-case adjustments to the survey data to produce plausible 2019 values are as follows; we also conduct sensitivity analysis to explore how our findings are affected by alternative adjustments.

— Changes in quantities demanded are increased to allow for subsequent growth in the US economy. The adjustment uses the increase in real GDP, but because inventories are specific to the sites where they are held, the change also takes into consideration growth in establishments. Establishment numbers for manufacturers' finished goods, materials and supplies, and work in process inventories, and retailers' and merchant wholesalers' inventories are taken from the US Census Bureau's County Business Patterns, which are available to 2018. The standard deviation of mean demand is scaled along with the mean.

— The average values of the items being sold are adjusting according to changes in the producer price index (PPI) for different industries; for years before a separate PPI was calculated for a specific industry, the overall composite PPI was used.

— Trucking costs (described further below) are inflated according to the change in the PPI for trucking, an increase of about 90% over the period. Transit times also are assumed to have increased by 10% because of greater offshoring of products and an increase in shipping distances.

— Order processing times are assumed to have decreased 20% because of the significant improvements in information technology during the past three decades. Ordering costs per order also are assumed to have dropped significantly, 40%, because of the same improvements in information technology that were assumed to reduce order processing times.

— Warehousing costs are adjusted by comparing the PPI for warehousing with the industry PPIs and applying the relative increase or decrease to the difference between the warehousing holding cost and the in-transit holding cost. Higher values were used when they resulted in an order size that more closely matched the average order size for the industry, with the justification that products that suffer from greater obsolescence and depreciation generally warrant higher holding costs.

— Unit shortage penalties are increased by a fifth to take account of greater competition in the US economy from globalization and lower information and search costs attributable to the advent of the Internet.

Table 1 presents the updated data that we use in the logistics cost model, including average demand, lead time, and product data for a typical product and consumer for each of the nine

industry categories. The values of the variables appear to be plausible by inspection, although we acknowledge that we are not aware of any comparable data sets including those variables that we could use for comparison.

Following Tyworth and Zeng (1998), both daily demand and the transit time component of lead time are assumed to be random variables that take a gamma distribution. Daily demand, for instance, ranges from less than 10 units on average (for merchandise) to nearly 200 units (for food), depending on industry. The important service time variables that we assume will be directly affected by autonomous truck operations are mean transit time, which ranges from 2.75 to 5.5 days, and the standard deviation of transit time, which ranges from 0.55 to 2.42 days. Product values per unit range from $12 to $356 and weights per unit range from less than half a pound to nearly 40 pounds.

To execute the numerical simulations of a firm's logistics cost-minimizing behavior, we also need values for the inventory parameters, a specification of the freight rate function, and counterfactual values of the transportation variables that capture improvements from the adoption of autonomous trucking.

<div align="center">Inventory Parameter Values</div>

As shown in table 1, inventory costs are expressed as a percentage of the unit value of a good. Holding costs include costs for storage, insurance, taxes, handling, depreciation, interest, and obsolescence, and range from 23% to 51% across industries. Holding costs for in-transit stock are generally lower than those for warehoused goods, as storage and insurance costs are typically avoided. For the simulations they are taken to be 20%. Unit shortage costs, 30%, represent the sum of the probabilistically adjusted costs of lost sales and lost customers that result when a product is unavailable.

Freight Rate Function

The freight rate function for an order size Q is based on two alternative power functions presented in Tyworth and O'Neill (1997) and Tyworth and Zeng (1998), which were derived from data originally supplied to them by Yellow Freight Systems. Each specification takes the form: $f(Q) = a \times (z/100) \times (Q \times w)^b$, where a and b are constants, z is the freight class, and w is the weight per unit of the good in pounds. Specification (1) takes $a = 306.94$ and $b = -0.3071$, while specification (2) takes $a = 463.8$ and $b = -0.3325$. Both specifications originated with 300-mile shipments at 1995 rates and then consider changes in trucking costs as described above. Preference was given to the freight rate specification that achieved the reorder size that more closely matched the LaLonde et al. (1988) average order size for each industry. We indicate the freight rate specification, (1) or (2), which we used for each industry category in table 1.

Counterfactual Values of Transportation Variables

As noted, we assume that the widespread adoption of autonomous trucks will affect the transportation time variables, which in turn will affect a firm's logistics costs. Recall the example of large time savings that would accrue to shippers for shipments from Los Angeles to Dallas if they were transported by autonomous instead of non-autonomous trucks. Given their ability to reduce labor, insurance, and accident costs, we also assume that autonomous trucks reduce transportation costs.

The Ryder System and the Socially Aware Mobility Lab at the Georgia Institute of Technology (2021) estimate that autonomous trucking could reduce the cost of transportation 29% to 40%. Yang (2021) estimated that self-driving trucks would reduce overall freight prices by 25.6% by lowering per-mile freight transportation costs, increasing the daily driving range, and eliminating other human factors related to truck driving. We therefore make conservative

assumptions compared with the preceding figures that the widespread adoption of autonomous trucking would yield a 20% improvement in transportation cost, time, and reliability. We perform a sensitivity analysis assuming an even lower value of 10% improvement of the transportation variables.[214]

The base case assumption for the improvement in the transportation variables is consistent with previous work on the effects of the development of the Interstate Highway System on costs and service times. Keeler and Ying (1988) estimated that trucking costs were reduced approximately 20% over a 23-year period that included the Interstate Highway System construction. Friedlaender (1965) assumed average vehicle speed increases of between 5 and 10 miles per hour after construction of the new highway system, or roughly 20% of the pre-interstate speed for different types and sizes of vehicles, including trucks. Our simulations assume comparable improvements in average transit time of roughly a day (given average transit times of as much as 5 days) and reductions in the standard deviation of transit time of up to half a day.

In the appendix, we summarize our complete model of a representative firm's total logistics costs, including the updated values of the parameters and variables. Given those values, we assume a firm chooses the reorder quantity (Q) that is expected to reach the desired target inventory level (S) and a reorder point (s) to minimize total logistics costs. We then assume, as discussed,

---

[214] It is not clear how the full cost of firms' trucking operations will be affected by the adoption of autonomous vehicles. The capital costs of autonomous trucks may be higher than the capital cost of non-autonomous trucks because of additional hardware, in the form of additional sensors and powerful computers. Those costs will decline when a sufficient volume of autonomous trucks is produced that exhausts scale economies. At the same time, autonomous trucks are likely to have lower insurance costs than non-autonomous trucks. In addition, the operating costs of autonomous trucks are likely to be lower than the operating costs of non-autonomous trucks because electric autonomous trucks will have lower energy costs than non-autonomous trucks. Gasoline powered autonomous trucks also will have lower operating costs than non-autonomous trucks because they will get better fuel economy by operating in a smoother traffic flow that is less subject to disruptions, such as stop and go driving. Finally, as noted, autonomous trucks also will have much lower operating costs than non-autonomous trucks because they will have lower labor costs and because they will be operating most of the time instead of stopping by law for several hours to enable a human driver to rest.

counterfactual values of transportation cost, transit time, and the reliability of transit time, which would exist in an environment where autonomous trucks were adopted and re-optimize a firm's total logistics costs. We then compare the logistics costs in the two environments to determine the total logistics cost savings for the nation.

4. Gains from Widespread Adoption of Autonomous Trucking

Table 2 presents the results of the counterfactual simulation of the effects of autonomous trucking on firms' logistics costs for each of the nine industries under the current conditions of non-autonomous trucking, our base case counterfactual scenario of a 20% reduction in transportation costs, travel time, and travel reliability, and an alternative scenario of a 10% reduction in those variables. In the base case, the reductions in logistics costs for the different industries range from 13% to 16%, with an overall average of 15%. Inventories and their attendant costs decline because of the improvements in transportation. Firms place more frequent orders, reflected in higher ordering costs, of smaller shipment sizes and at lower reorder points.

The results for the electronics industry illustrate the nature of the changes in logistics costs and firm behavior. The effect of the transportation improvements yields a $1,104 decrease in expected annual logistics costs for the included product, representing a reduction of 14.9%. Order costs rise from $166 to $191, but order sizes and reorder levels fall, and inventories, backorders, and their associated costs decrease. The transportation charge also drops from $3,551 to $2,968.

Extrapolating the results from the nine industries to the overall economy, we obtain an estimate of annual savings of $140 billion (2019 dollars) for an average reduction in logistics costs of 15%. Assuming the logistics cost savings for the firms in the sample are representative of the savings that could be achieved by other firms in the economy, our national estimate also can be

obtained by scaling the estimated cost savings in the sample based on the logistics costs in the rest of the economy, which were \$1.62 trillion in 2019 (Zimmerman et al. (2021)). Note that we make two adjustments to appropriately scale the results. First, we account for the fact that 75% of freight transportation expenditures are trucking-related (Zimmerman et al. (2021)). Second, because all inventories are not influenced by changes in transportation, specifically, final goods inventories and work-in-progress inventories are not necessarily determined by the kind of EOQ model presented here, the model is applicable to the manufacturing raw materials inventories, wholesale inventories, and retail inventories that account for 77% of total inventories.[215]

If we invoke the more conservative assumption of a 10% reduction in transportation costs, travel time, and the reliability of travel time, the average reduction in logistics costs would be 7.4% instead of 15%, resulting in \$69 billion (2019 dollars) of annual savings, which is still a sizable gain.

We also conducted another sensitivity analysis where we used the unadjusted (original) values as reported by LaLonde et al. (1988) for the demand, product, inventory, and transportation cost variables to explore the effect of our adjustments. Perhaps surprisingly, we obtained estimates of national savings from the adoption of autonomous trucking that were within \$1 billion of the findings for both 20% and 10% reductions in transportation costs and service times that were based on the adjusted values of the variables.

Essentially, our adjusted figures that increased unit demand, shortage costs, and trucking rates lead to larger percentage savings from autonomous trucking, all else equal. However, our

---

[215] The use of an EOQ model should be qualified that it is tailored for industries where shortages are a small fraction of sales (Tyworth and Zeng (1998)). The results are less accurate for industries where shortages are a large fraction of sales, but the direction of the bias is not clear. The treatment of shortage costs also may not be straightforward because the cost of shortages may reflect the cost of delayed or forgone use of a product by a consumer, which may be larger or smaller than the cost of shortage to a firm. Again, it is not clear how our findings are affected.

adjusted figures that increased unit product values, holding costs, and order costs lead to smaller percentage savings from autonomous trucking. The intuition behind those results is that changes that lead to a higher share of transportation costs within logistics costs, such as higher trucking rates and higher unit demand, provide a greater opportunity for autonomous trucking to reduce those costs. Among those factors, changes in trucking rates and holding costs tend to have the largest impact on national savings because they have the most direct impact on transportation costs and inventory costs, respectively. However, the effect of a change in any single parameter on the estimated national savings is small, and even a doubling of any of those parameters would most likely lead to less than a percentage point increase in the estimated 15% logistics costs savings from autonomous trucking under the 20% cost and service time reduction scenario. Overall, the changes in those parameters that were realized over the past few decades represented a mix of positive and negative influences on the estimated national savings from the adoption of autonomous trucking.

The gains also might be lower if we assume that autonomous trucking is used only for certain types of trucking operations, such as long-distance trucking. However, in the long run, it is not clear why autonomous trucking, in combination with autonomous vans or other local delivery vehicles, would not be used on all types of roads and for all types of trucking operations.[216]

---

[216] Recent work by the US Department of Transportation analyzes the macroeconomic effects of the introduction of automation in the long-haul trucking sectors in the United States by introducing productivity shocks to that form of trucking (Waschik et al. (2021)). Although that work takes a different approach than we take here, the estimated benefits are broadly consistent with what we obtained from our simulations. DOT's calculations of cost savings are based on a computable general equilibrium model of the US economy that includes details on transportation-related industries including for-hire and in-house trucking. Half of for-hire and 10% of in-house trucking are assumed to be long haul in the report, based on employment shares. In DOT's simulation, the estimated benefit is a 2019 GDP equivalent of $68 billion by the end of the 30-year period, with much of the benefit achieved within 15 years under a fast adoption scenario. Most of the economic effects come from the labor and capital savings aspects of autonomous trucking, with additional benefits in the form of fuel savings and capital investment.

5. Additional Benefits from Autonomous Trucking

We have found that the adoption of autonomous trucking could produce substantial benefits to the entire US economy by reducing firms' logistics costs. Those benefits do not include reducing the likelihood of a supply chain disruption or shortening its duration should one occur like the one the US experienced during 2021 and 2022. Such disruptions can lead to a large increase in trucking costs. Indeed, the US Bureau of Labor Statistics estimated that the annual inflation rate for overall trucking costs during December 2021 was 17% and the inflation rate for long-distance trucking was 25% (Tett (2021)).

The adoption of autonomous trucking also could generate other substantial benefits to the economy. Autonomous trucking will represent a loss of employment to drivers in the seat[217], but also make the industry less vulnerable to the growing shortage of drivers. Importantly, Winston and Karpilow (2020) argue that autonomous trucking is likely to increase employment because some drivers will be needed to provide security for a vehicle's cargo and to handle certain logistical matters, while additional jobs will develop at transportation centers that monitor the autonomous road network used by trucks. Jobs also will develop to maintain, repair, and clean autonomous trucks, and to keep maps for optimal routings updated. Finally, by increasing the nation's productivity and output, autonomous trucking could increase economy-wide employment.

Autonomous trucking also would produce gains for firms as they change production operations beyond their logistics operations and realize X-efficiencies from the use of improved inputs shipped longer distances. Firms also would benefit from greater economies of scale (Mohring and Williamson (1969)) and scope because their product could be sold at lower cost to

---

[217] Viscelli (2018) concludes that the current trucking jobs most at risk of displacement are long-distance driving jobs with few specialized tasks, representing fewer than 300,000 drivers. Gittleman and Monaco (2019) reach a similar conclusion.

more geographically dispersed markets. Firms also may realize savings from changing input mixes, including their use of land and labor, to take advantage of the change in relative factor prices that is attributable to the lower cost of transporting freight by truck. And firms may realize savings by using autonomous freight technology to improve their operations. For example, equipment manufacturers are investing in self-driving tractors that can plow fields by themselves, which can be used by farmers to boost productivity in agriculture. A share of the preceding sources of cost savings to firms will be passed on to consumers in lower product prices. Consumers also will benefit as firms increase the variety of products they offer or create innovative new products in response to changing input prices from lower transportation and logistics costs.

Autonomous trucking also will help reduce the externality costs of highway travel. Drivers and passengers of all vehicles on the road will benefit from reduced congestion as truckers increase off-peak and night-time driving. The improvements in the safety of trucking operations will reduce the cost of accidents and incident delays. And, as in the case of automobiles, autonomous trucks will eventually operate with emission-free vehicles, although large truck makers have only begun to mass-produce such vehicles.[218] As discussed in chapter 6, autonomous vehicles will facilitate the adoption of efficient highway pricing policies, including congestion pricing of all vehicles and, in the case of trucking, axle-weight and weight-based charges to reduce the damage that trucks do to road pavement and bridges (Winston (2021a)).

Finally, by significantly reducing the costs of shipping freight and providing faster and more reliable service, autonomous trucks will enable truckers to present a stronger competitive challenge to non-autonomous and, as discussed in the previous chapter, autonomous railroads. By

---

[218] Ewing (2022) describes the dilemma truck makers face of whether to build trucks with batteries (Traton, Volkswagen's truck unit has made that choice), or to build trucks with fuel cells that convert hydrogen into electricity (Daimler Truck favors that choice although it does not plan to begin mass-producing a hydrogen fuel cell truck until after 2025).

providing truck convoys or "road trains," which could carry much larger shipments without stopping until they reach their destination, it is possible that autonomous trucking operations could provide service to so-called "captive" shippers of bulk commodities, like coal and grain.

Current government policies are slowing nationwide adoption of autonomous trucking by limiting their operation in some states. Aurora, for example, is testing its autonomous Class 8 trucks in Texas but cannot operate those trucks in California without human drivers. Autonomous intermodal operations could expand in states with autonomous trucking and further increase the efficiency of the nation's surface freight transportation system, but such operations would be prevented in states that do not allow autonomous trucking.

An environment that could particularly benefit from automation—or be harmed by its absence—is the delivery of cargo carried by ocean transportation from abroad, which could facilitate smooth freight flows from ports. Traffic congestion and stranded freight at West Coast ports have been a prominent source of recent disruptions in the supply chain.

Accounting for all the sources of potential benefits from autonomous trucking, the magnitude is likely to be in the hundreds of billions dollars, or a notable fraction of the magnitude of the potential benefits from autonomous cars. Society is likely to realize the benefits from autonomous trucking sooner than it realizes the benefits of autonomous cars because trucks will operate in autonomous mode primarily on freeways instead of on more complicated city streets. In addition, autonomous trucking has the support of firms that wish to use autonomous operations and that can improve the overall performance of local and state economies by improving industry productivity and efficiency. Autonomous trucking may therefore pave the way for faster adoption of autonomous cars by demonstrating to the public that autonomous vehicles are safe and could significantly increase the efficiency of the transportation system and the economy.

Appendix

Profit-maximizing firms will minimize their expected annual logistics costs ($EALC$), which can be decomposed into the sum of ordering costs, transportation costs, holding costs, and shortage costs. We specify those costs mathematically in this appendix and we indicate the firm's logistics cost-minimization problem that we solve numerically.

Ordering Costs. The first component of logistics costs is the annual cost of placing orders (c_a). Annual ordering costs are the product of the number of orders per year, which is composed of the annual demand, R, divided by the order size, Q, multiplied by the cost per order, A: $c_a = (R/Q) \times A$.

Transportation Costs. The second component of logistics costs is annual transportation costs (c_t). Annual transportation costs depend on the number of units (R) demanded each year and the rate (F) at which the freight is being carried. This freight rate may be affected by the quantity (Q) of units being carried per shipment, because a volume discount may be offered when shipping a greater quantity. Thus, the freight rate can be specified as $f(Q)$, where $f(\cdot)$ can be an empirically estimated relationship between the freight rate and the unit lot size. The expression for annual transportation costs is $c_t = f(Q) \times R$.

Holding Costs. The third component of logistics costs is the annual cost of holding inventories (c_h). Holding costs depend on a combination of in-transit, cycle, and safety stocks. In-transit stocks are those in the process of being transported. Product demand (D) is a random variable with a mean value demanded of μ_D units per day with standard deviation σ_D. Transit time for a shipment is a random variable with a mean value of μ_T days with standard deviation σ_T. The expected level of in-transit stocks can be expressed as the product of the mean number of units demanded per day times the mean number of days each shipment is in transit ($\mu_D \times \mu_T$). The

holding costs for these expected in-transit stocks then depend on the holding cost for each expected unit in transit (Y). The rest of the holding costs accrue to units on hand, as in a warehouse. These units can be broken down into cycle stock units and safety stock units. Cycle stock units are those that are on hand between shipments. Because shipment sizes are Q units and the distribution of inventory levels is uniform, the expected level of cycle stock at any point in time is simply half of the shipment size ($Q/2$). Safety stock units are those still on hand when a new shipment is received. New orders are placed when stocks reach the reorder point (s). Expected demand during the lead time before the shipment is received is (μ_X). The stock level at the reorder point will be depleted by the expected lead time demand, so the expected level of safety stock is the difference between the two, $s - \mu_X$,

To consider the effects of transportation on safety stock holding costs, mean lead-time demand can be expressed as the product of the mean lead time and the mean daily demand, $\mu_X = \mu_L \times \mu_D$. Mean lead time can be further broken down into a sum that includes the time involved in transportation, $\mu_L = \mu_T + v$, with μ_T the expected transit time and v a constant representing a fixed time spent on non-transportation activities in the order cycle (such as order preparation, order processing, manufacturing, or assembly, depending on context). The holding costs for these expected cycle and safety stocks then depend on the holding costs for each expected unit stored in a warehouse (W). The expression for total annual holding costs is:

$$c_h = (\mu_T \times \mu_D) \times Y + (Q/2 + s - (\mu_T + v) \times \mu_D) \times W$$

Shortage Costs. The fourth and final component of logistics cost is annual shortage costs (c_s). Shortage costs include the costs of any expedited delivery to satisfy demand once inventory is depleted, plus the cost of any lost sales or lost customers that result from a shortage. In a raw materials or work-in-progress inventory, shortage costs may take the form of expedited delivery

costs, overtime, or lost production. The shortage cost term depends primarily on the expected level of shortage during the lead time between each order and delivery over the year (ES). These expected shortages can be expressed as a weighted average of conditional expected shortages over the lead time distribution

$$ES = \int (x - s) f(x) dx$$

where $x > s$. Total shortage costs are then derived by multiplying this expectation of the number of units short per order by the annual number of orders (R/Q) and the shortage cost per unit short (B).

$$c_s = ES \times (R/Q) \times B$$

or

$$c_s = \left[\int (x - s) f(x) dx \right] \times (R/Q) \times B$$

A firm is assumed to minimize expected total annual logistics costs, which are the sum of the ordering (c_o), transportation (c_t), holding (c_h), and shortage (c_s) cost elements; thus,

$$EALC = c_o + c_t + c_h + c_s$$

Expressed as a function of the decision variables, reorder point (s) and reorder quantity (Q), the expected total annual logistics costs become:

$$EALC(s, Q) = \{(R/Q) \times A\} + \{f(Q) \times R\} + \{\mu_T \times \mu_D \times Y + (Q/2 + s - (\mu_T + v) \times \mu_D \times W\}$$
$$+ \{[\int (x - s) f(x) dx] \times (R/Q) \times B\}$$

For the simulations, we minimized those logistics costs using the solver function in Excel.

Table 1. Data Used in the Simulations

	Food	Chemical	Pharma	Auto	Paper	Electronics	Clothing	Other Mfg	Merch
DEMAND									
Mean of daily demand (units), μ_D	193	41	14	26	21	46	26	33	6
Std. of daily demand (units), σ_D	115.8	24.6	8.4	15.6	12.4	27.6	15.6	19.8	3.6
LEAD TIME									
Constant order processing days, v	1.6	1.6	0.8	0.8	3.2	2.4	2.4	0.8	0.8
Mean transit time (days), μ_T	2.75	5.5	3.3	4.4	3.3	4.4	4.4	4.4	4.4
Std. deviation of transit time (days), σ_T	0.55	1.32	1.1	1.76	1.32	2.42	2.42	2.2	2.2
PRODUCT									
Unit value, V	$27.06	$356.63	$229.67	$108.80	$52.58	$11.80	$59.61	$52.07	$55.57
Unit weight (lbs.), w	4.4	37.4	0.4	6	1.5	0.4	4.3	1.6	3.4
INVENTORY									
Holding cost (warehouse), W	38%	34%	23%	27%	37%	51%	27%	27%	38%
Holding cost (in-transit), Y	20%	20%	20%	20%	20%	20%	20%	20%	20%
Ordering cost per order, A	$6	$6	$6	$6	$6	$6	$6	$6	$6
Unit shortage penalty fraction, B	30%	30%	30%	30%	30%	30%	30%	30%	30%
TRANSPORTATION									
Rate discount, d	6%	6%	6%	6%	6%	6%	6%	6%	6%
Freight class, z	70	65	100	100	65	100	100	100	100
Freight-rate specification	1	1	2	2	1	1	1	1	1

Table 2. Simulation Results

	Food			Chemical			Pharmaceuticals			Auto		
	Pre	10%	20%	Pre	10%	20%	Pre	10%	20%	Pre	10%	20%
Transportation Variables												
Real transportation cost (% of pre)	100%	90%	80%	100%	90%	80%	100%	90%	80%	100%	90%	80%
Mean transit time (days)	2.75	2.5	2.25	5.5	5	4.5	3.3	3	2.7	4.4	4	3.6
Std. deviation of transit time (days)	0.55	0.5	0.45	1.32	1.2	1.08	1.1	1	0.9	1.76	1.6	1.44
Decision Variables												
Order size (Q)	2,430	2,253	2,071	373	345	318	77	73	69	473	438	402
Reorder level (s)	1,084	1,026	971	421	394	366	116	109	102	238	223	207
Logistics Cost												
Expected annual logistics cost (EALC)	$55,606	$51,455	$47,225	$115,871	$107,201	$98,363	$11,174	$10,459	$9,739	$31,257	$28,953	$26,595
Percentage change from pre value		7.5%	15.1%		7.5%	15.1%		6.4%	12.8%		7.4%	14.9%
Inventory Analysis												
Cycle stock	1,215	1,127	1,036	186	173	159	38	36	35	236	219	201
In-transit stock	531	483	434	226	205	185	46	42	38	114	104	94
Safety stock	244	235	228	130	123	116	59	56	53	103	98	93
Expected backorders	6.46	5.73	5.01	1.3	1.13	0.97	0.15	0.14	0.12	1.71	1.46	1.24
Order-fill level	99.70%	99.70%	99.80%	99.70%	99.70%	99.70%	99.80%	99.80%	99.80%	99.60%	99.70%	99.70%
Cost Analysis												
Ordering	$172	$185	$201	$238	$256	$279	$395	$415	$438	$119	$128	$140
Cycle stock	$12,746	$11,824	$10,875	$22,848	$21,182	$19,485	$2,029	$1,929	$1,830	$7,070	$6,550	$6,018
In-transit stock	$2,931	$2,665	$2,400	$16,264	$14,789	$13,314	$2,128	$1,934	$1,741	$2,535	$2,305	$2,076
Safety stock	$2,509	$2,416	$2,346	$15,759	$14,902	$14,028	$3,112	$2,949	$2,784	$3,017	$2,871	$2,721
Backorders	$1,500	$1,435	$1,365	$5,510	$5,169	$4,832	$695	$654	$614	$1,105	$1,022	$940
Transportation cost	$35,750	$32,930	$30,037	$55,253	$50,902	$46,426	$2,816	$2,578	$2,332	$17,411	$16,076	$14,700

Table 2. Simulation Results (continued)

	Paper			Electronics			Clothing			Other Manufacturing		
	Pre	10%	20%	Pre	10%	20%	Pre	10%	20%	Pre	10%	20%
Transportation Variables												
Real transportation cost (% of pre)	100%	90%	80%	100%	90%	80%	100%	90%	80%	100%	90%	80%
Mean transit time (days)	3.3	3	2.7	4.4	4	3.6	4.4	4	3.6	4.4	4	3.6
Std. deviation of transit time (days)	1.32	1.2	1.08	2.42	2.2	2	2.42	2.2	2	2.2	2	1.8
Decision Variables												
Order size (Q)	189	178	167	600	560	520	526	487	447	454	422	389
Reorder level (s)	161	151	141	446	417	388	263	246	230	345	322	299
Logistics Cost												
Expected annual logistics cost (EALC)	$7,527	$6,947	$6,357	$7,417	$6,863	$6,313	$18,899	$17,434	$15,961	$15,446	$14,310	$13,151
Percentage change from pre value		7.7%	15.5%		7.5%	14.9%		7.8%	15.5%		7.4%	14.9%
Inventory Analysis												
Cycle stock	95	89	83	300	280	260	263	243	224	227	211	194
In-transit stock	69	63	57	202	184	166	114	104	94	145	132	119
Safety stock	25	21	17	133	122	112	86	80	74	174	164	154
Expected backorders	0.73	0.64	0.56	5.85	5	4.29	2.69	2.28	1.93	2.04	1.74	1.46
Order-fill level	99.60%	99.60%	99.70%	99.00%	99%	99.20%	99.50%	99.50%	99.60%	99.60%	99.60%	99.60%
Cost Analysis												
Ordering	$239	$255	$272	$166	$178	$191	$107	$115	$126	$157	$169	$183
Cycle stock	$1,861	$1,749	$1,637	$1,841	$1,717	$1,597	$4,323	$3,999	$3,677	$3,233	$3,001	$2,767
In-transit stock	$736	$669	$602	$487	$443	$399	$1,392	$1,266	$1,140	$1,531	$1,392	$1,253
Safety stock	$477	$401	$324	$802	$737	$676	$1,391	$1,284	$1,187	$2,443	$2,305	$2,163
Backorders	$462	$430	$399	$571	$524	$483	$854	$784	$722	$831	$765	$699
Transportation cost	$3,752	$3,441	$3,122	$3,551	$3,265	$2,968	$10,832	$9,986	$9,110	$7,251	$6,677	$6,086

Table 2. Simulation Results (continued)

	Merchandise		
	Pre	10%	20%
Transportation Variables			
Real transportation cost (% of pre)	100%	90%	80%
Mean transit time (days)	4.4	4	3.6
Std. deviation of transit time (days)	2.2	2	1.8
Decision Variables			
Order size (Q)	126	117	109
Reorder level (s)	54	50	47
Logistics Cost			
Expected annual logistics cost (EALC)	$5,786	$5,367	$4,930
Percentage change from pre value		7.2%	14.8%
Inventory Analysis			
Cycle stock	63	59	54
In-transit stock	26	24	22
Safety stock	23	22	20
Expected backorders	0.83	0.71	0.6
Order-fill level	99%	99.40%	99.40%
Cost Analysis			
Ordering	$103	$111	$119
Cycle stock	$1,367	$1,274	$1,176
In-transit stock	$302	$275	$246
Safety stock	$480	$457	$432
Backorders	$238	$218	$199
Transportation cost	$3,296	$3,033	$2,758

Chapter 13. Water Transportation and Drones

Clifford Winston   Jia Yan

1. Introduction

Surface freight transportation is complemented by water transportation provided by barges, container and cargo ships, and tankers for domestic shipments transported on major rivers and the Great Lakes and for international shipments transported across the oceans. Drones are an innovation that could improve the transportation system by delivering small freight shipments and mail in a timelier fashion than current surface modes.

In this chapter, we briefly discuss how competition and innovation in water transportation and drones could further improve freight transportation. We also discuss appropriate government policy to enhance the efficiency of those modes.

2. Water Transportation

Water transportation operations, like highway and airport operations, are not governed by efficient pricing and investment of the natural water environment and the port facilities that they use. Freight transported on container and cargo ships are therefore subject to significant congestion and delays as indicated by the increasing average number of days that containers are waiting on the docks for handling at major US ports during 2021-2022. These so-called dwell times have significantly delayed shipments that were unloaded at the container port complex at Los Angeles and Long Beach. As shippers seek less-congested destinations for their shipments, dwell times have increased at the Ports of Oakland, California and at other ports throughout the country including Savannah, Georgia, Charleston, South Carolina, and, most recently, the Port of New York and New Jersey.

Government policies also have exacerbated supply chain delays by limiting potential efficiency improvements that would be spurred by more competition and innovation in water transportation. In this section, we discuss competition and innovation primarily in ocean freight transportation, but the ideas also apply to freight transportation on rivers and lakes.

<p align="center">Ocean Transportation</p>

Levinson (2006) pointed out that stacking containers— the basis for containerization—was facilitated by the development of a lock connected to the corners of containers that crane operators could mechanically open and close from their viewing platforms. Thus, the shipping container became the workhorse of globalization.

Recently, ocean shipping has been marked by floating traffic jams that arise because dockworkers are overwhelmed by the volume of freight to unload, dock space is not available to unload all the freight, and truck drivers are sometimes stranded at the entry to a dock and cannot access freight to carry it from a dock to its destinations. Containers that cannot be unloaded also cannot be loaded, adding to delays. The breakdown of the supply chain has caused freight rates to skyrocket and contribute to the rise in inflation that began in 2021.

Additional competition and innovation have long had the potential to improve ocean shipping and could help prevent future breakdowns of the supply chain. However, government must reform its policies to facilitate those sources of improvement.

Historically, price competition in ocean transportation had been limited by an antitrust exemption that permitted ocean carriers to fix rates at shipping conferences to prevent destructive competition (Sagers (2006)). Entry into ocean shipping is regulated by the 1920 Jones Act, which decreed that ships allowed to call at two consecutive American ports must be built in the United States, owned by American companies, fly the American flag, and be operated by American crews.

Inefficiencies created by price and entry regulations have been compounded by inefficient operations of US ports.

The Ocean Shipping Reform Act of 1998 took a first step toward ending rate conferences by allowing shippers to negotiate contracts with carriers, causing rates to decline on certain routes (Reitzes and Sheran (2002)). Disbanding all ocean liner rate conferences would effectively end the remaining price-fixing arrangements and could reduce ocean-shipping rates at least 30% (Fink, Mattoo, and Neagu (2002)). Yet, ocean liner rate conferences still exist.

The 2010 Open America's Waters Act called for repealing the Jones Act, which according to a U.S. International Trade Commission study, would have generated a $1 billion efficiency gain by reducing the costs of ocean shipping (Kashian, Pagel, and Brannon (2017)). However, Congress did not pass the Act. The Trump administration explored eliminating the Jones Act but decided against doing so without providing a clear explanation for its decision. Thus, the Jones Act continues to prevent competition that could be provided by foreign ocean container ships.

<u>Port Operations</u>

Most US ports are owned by a public authority, where the authority varies in different locations by federal, state, or local government. In 2021, The World Bank and IHS Markit published The Container Port Performance Index 2020, the first effort to offer a systematic ranking of the performance of 351 ports around the world.[219] In general, US ports faired very poorly. As the report notes, the main West Coast US ports are ranked near the bottom of the 351 ports with Los Angeles (#328), Oakland (#332), Long Beach #333), and Tacoma (#335). The main US East Coast ports rank higher but even the best performers are not among the top 20% with New York

---

[219] https://ihsmarkit.com/info/0521/container-port-performance-index-2020.html

and New Jersey (#89), Savannah (#279), Virginia (#85), and Charleston (#95). The main US Gulf of Mexico port in Houston also is ranked poorly (#266).

Although much of the discussion of US ports focuses on international trade, their inefficiencies affect domestic trade as well. Substantial US flows of goods could be shipped up and down the East Coast, the West Coast, or in and out of the Gulf of Mexico. However, much of that ocean-based shipping does not occur because of costly and inefficient ports and entry barriers to less-costly foreign ships created by the Jones Act. Some goods are therefore shipped via surface transportation when it could have been less costly to ship them via water transportation if ports and carriers had the incentive and ability to reduce their costs.

The increase in international trade caused by globalization has caused ocean carriers to use bigger and bigger ships to realize economies of vehicle (container ship) size. Because those ships have difficulty fitting into most US ports, a large share of US imports are concentrated at the few ports, such as Long Beach and Los Angeles, that can accommodate large ships. By concentrating traffic in a few ports along the West Coast, congestion is increased and a breakdown of the supply chain is more likely, which motivates the importance for the US port system to distribute the volume of ocean freight more evenly by increasing the number of ports that can accommodate large ships.[220]

Deregulation of Ocean Shipping and Privatization of Ports

Ocean transportation has gained attention as an unexpected source of inflation. As reported by Swanson (2022), the spot price to transport a container from China to the West Coast of the United States costs 12 times as much as it did two years ago, while the time it takes a container to make that journey has nearly doubled. The inflated out-of-pocket and delivery time costs raise

---

[220] Another major source of freight congestion derives from bottlenecks at full warehouses that prevent shippers from unloading their containers and thus quickly picking up containers from ports and rail ramps.

consumers prices by increasing the prices that companies pay for source products and parts from overseas and shrink the scope of their spatial search for inputs.

Congress passed and President Biden signed the Ocean Shipping Reform Act, which includes regulations that seek to improve shipping opportunities provided by foreign ocean carriers and to prevent them from charging excessive late fees, known as detention and demurrage charges. Biden characterized the new ocean shipping legislation as an effective way to fight inflation; however, Biden passed on an opportunity to truly fight inflation by deregulating ocean shipping and privatizing US ports.

Deregulation of airlines and trucking occurred during a period of high inflation and could plausibly be characterized as an anti-inflation policy, especially because it succeeded in significantly reducing the price of air and surface transportation. Policymakers should learn from the past and pass legislation that would eliminate rate conferences and the Jones Act. By deregulating ocean shipping and privatizing ports, policymakers could increase competition in ocean transportation and spur innovations that could reduce shipping rates and improve service.

The Effects of Deregulation and Privatization

As noted above, eliminating the Jones Act and all rate conferences could generate billions of dollars in efficiency gains and reduce shipping rates. In addition, privatizing ports to stimulate competition could significantly reduce the accumulated cost inefficiencies and encourage ports to improve their operations by adopting the latest technologies to improve the flow of traffic and the loading and unloading of shippers' freight.

US shippers that send and receive international shipments may be located close to or far from a US port. Shippers that are located close to oceanic ports will have a competitive advantage over shippers located further away from such ports because they will incur lower shipping time

and out-of-pocket costs for the domestic part of their international shipment. However, many shippers are not located near an oceanic port, and, in theory, the domestic part of their international shipment could be handled by several alternative US ports. For example, shippers in the East and Midwest could have their shipments to Japan go through ports in San Diego, Los Angeles, Long Beach, Hueneme, Oakland, Portland, or Seattle.

Privatization of the nation's ports could create a more intense competitive environment for international freight and provide the incentive and ability for ports to make investments to increase their market share by reducing costs, using the latest technology, and increasing dredging and expanding their facilities to accommodate larger container ships. Competition among ports for freight from Asia, for example, is limited because the two ports at Los Angeles and Long Beach are among the few that can accommodate the largest ships. Thus, they typically handle more than 40% of imports, while their combined share has climbed to more than 50% during the global supply chain disruption. Private port competition that results in several US west coast ports increasing their market share of traffic from Asia could significantly reduce the monetary and time-related costs of ocean shipping.

Unions representing port workers have been effective at negotiating high pay and benefits for those workers. For example, union dockworkers represented by the International Longshore and Warehouse Union make an average of $171,000 per year plus free healthcare (Read (2020)). Privatization that increases port competition could put downward pressure on elevated labor costs, which would be passed on along the supply chain. Union agreements also limit ports' hours of operations, making it more difficult to reduce ships' dwell times during supply chain disruptions. For example, ports in Asia operate 24 hours a day, matching the 24-hour-a-day pace of factories, while the ports of Los Angeles and Long Beach generally operate only 16 hours a day as part of

union labor agreements. Although some docks eventually increased their operating hours as supply chain disruptions intensified, it is likely that the competition spurred by privatization would be necessary to increase all ports' operating hours permanently.

Privatization may overcome unionized dockworkers resistance to the greater use of autonomous cranes and straddle carriers that can load and unload containers more quickly than manually operated vehicles used by loaders to handle containers. Automation also can reduce maritime accidents and a vessel's berthing and unberthing time at a port. The dockworkers' union has reacted to automation as a threat to high-paying jobs and is seeking in contract negotiations with US West Coast port officials to draw a line against its use (Berger (2022)).

US port facilities are generally less automated and less efficient than those in other advanced economies. Their lower total factor and labor productivity results in it taking, for example, twice as long on average to move a container from a large ship at the Port of Los Angeles than it does to move a container from a large ship at top ports in China. Compared with US ports, European and Asian ports can clear backlogs more quickly because they have automated cargo-handling equipment that operates around the clock. A minority of those ports, including in China, are private. Tongzon and Heng (2005) provide evidence that privatization of ports throughout the world has led to efficiency gains and would likely generate such gains for privatized US ports.

Finally, privatization may give some smaller ports an incentive to attract more traffic and, if necessary, to expand their dock capacity. For example, smaller ports on the west coast like Oakland and Seattle can currently handle just a fraction of the shipping containers processed at the ports of Los Angeles and Long Beach. Private owners of those ports, and others in Canada, such as Vancouver and Prince Rupert, which have an advantage of shorter sailing times to and from

ports such as Shanghai, would be likely to consider new investments so they could accommodate large ships that seek alternative, less congested ports to unload their cargo.

<div align="center">Autonomous Ships</div>

The adoption of autonomous ships could help raise the performance of ocean transportation by improving routing efficiency and reducing congestion delays. Autonomous ships operate smoothly in crowded shipping lanes and close to shore and can operate 24/7 at all ports because they require far less labor on land to assist with their operations.

Avikus, which is a subsidiary of the world's largest shipbuilding company HD Hyundai, announced that the ultra-large vessel Prism Courage has completed the world's first transoceanic voyage of a large vessel using autonomous navigation technology on June 2, 2022. Prism Courage set off from the Gulf of Mexico and sailed through the Panama Canal before crossing the Pacific Ocean to the Boryeong LNG Terminal in South Korea. According to Avikus, the voyage took 33 days to complete, with route optimization increasing fuel efficiency by around 7 per cent and reducing greenhouse gas emissions by around 5 per cent.[221]

The Yara Birkeland, an autonomous container ship, is scheduled to be tested on a journey between two Norwegian ports at the end of 2022. Policymakers can help expedite the adoption of autonomous ships by setting congestion charges at ports to reduce traffic at peak times, and by using the revenues to fund infrastructure improvements that would enable ships to engage safely in autonomous operations by communicating in real time with each other and port coordination facilities to avoid collisions.

Autonomous trucking operations, discussed in the preceding chapter, would complement autonomous water transportation operations by eliminating delays caused by manual trucking

---

[221] In January 2022, the ocean container ship Mikage, sailing under the flag of Japan, completed a two-day voyage entirely autonomously from Tsuaruga to Sakai.

operations at ports and by expediting traffic flows in and out of the port that could reduce ships' dwell times and trucks' delivery times.

3. Drones and Delivery Robots

Drones, a new emerging technology, represent the future of a privatized mail system and the delivery of small packages and shipments. Although they were not initially designed to address the financial problems of the United States Postal Service, commercial drones have significant potential to provide more efficient service than any land-based delivery company because of their automated operations and the vast capacity in low-altitude air space. The technology also could reduce emissions and vehicle traffic, while making extremely fast deliveries in as little as five minutes.[222]

Currently, no drone delivery company in the US is fully certified to fly everywhere, without a human controlling or at least monitoring the aircraft. The Federal Aviation Administration claims to be developing regulations that would allow it to safely issue such authorization. For the time being, commercial drone delivery services are getting them certified through the same process by which passenger aircraft are certified.

Wing, an offshoot of Google, is the first drone operator to receive US governmental approval to operate as an airline and to drop products to actual customers. It provides service in southwest Virginia and in Finland and Australia. UPS has formed a new drone delivery subsidiary,

---

[222] Drones also can be used to improve transportation safety. For example, companies, such as BNSF Railway, use drones instead of employees to inspect parts of their plant that could pose safety hazards, such as underneath a highly elevated railroad bridge. States also are using drones to improve safety. For example, the Alaska and North Carolina Departments of Transportation are using a drone startup company, Skydio, to inspect their bridges and highways, New York and West Virginia transportation officials are using drone inspections to lower costs and improve safety, and Connecticut's senate passed a transportation bill that formalizes the use of drones for infrastructure inspections and traffic monitoring.

called UPS Flight Forward, and has received FAA approval to operate a drone delivery network.

UPS is striking agreements with healthcare groups to use drones to deliver medical supplies—for

example, to hospital campuses in Utah and elsewhere—and is making an agreement with CVS to

evaluate the use of drones for home delivery of prescriptions and other products. Finally, Amazon

launched a trial of its Prime Air drone delivery service for select customers in Cambridge, England,

in December 2016, and has recently received approval from the FAA to operate a fleet of delivery

drones and to begin testing customer deliveries. Other companies, including Microsoft, Uber, and

FedEx, have announced plans for drone delivery service trials, as well.[223]

Drones also can be used to move medical supplies quickly and to reduce human contact.

However, the FAA requires drones to be certified as a Part 135 air carrier, a process that can take

months or years to achieve. Zipline, a medical drone delivery business, has been active in Africa

for years.[224] It has been getting calls for help in the United States, but it was not yet certified by

the FAA to make deliveries. Finally, at the end of May 2021, Zipline's partner, the hospital system

Novant Health in North Carolina, received an FAA waiver to enable Zipline to make its first US

delivery of medical supplies from a field in Kannapolis, North Carolina, to a designated spot fifteen

miles away, near Novant Health's Huntersville Medical Center. The FAA should allow all drone

companies to expand their operations in a prompt and orderly fashion.[225]

Turing to the allocation of air space capacity, FAA recently announced that it would allow

fully automated commercial drone flights without hands-on piloting or direct observation by

---

[223] Dronamics is using the company's proprietary cargo drone, the Black Swan, to develop an international air delivery network in Europe. Drone Delivery Canada has begun operating drone flights carrying cargo from Edmonton International Airport, Canada.

[224] Zipline operates 1,000 drones and makes parachute deliveries globally.

[225] Draganfly uses its drones to deliver vaccines in rural parts of Texas; Volansi uses its drones to provide similar medical deliveries. Drone Delivery Canada will soon begin flying supplies from a distribution center in Milton, Ontario to the Oakville Trafalgar Memorial Hospital.

human controllers or observers. American Robotics is the first US company to gain this approval for its network.

For its part, the US commercial drone industry wants to create a privately funded and operated air traffic control network separate from the US air traffic control that would facilitate operations and prevent collisions. The FAA is unlikely to allow private air traffic control for drones, but it is preparing to test technologies at five US airports designed to ensure safe cohabitation of airplanes and drones. However, it is not clear how the FAA will allocate air space for airplanes and drones efficiently. Alternatives include efficient peak-load pricing and using bidding to lease airspace ((Skorup (2022)). Failure to allocate airspace efficiently will compromise the potential benefits from drones, especially for shipments needed in an emergency.

Various agencies and public officials are likely to raise security concerns about drones; hence, law enforcement agencies must work constructively with the drone industry to address those concerns and to enable the private sector to develop a technologically advanced mail and small package delivery industry.

Finally, sidewalk delivery robots are starting to make some so-called last mile deliveries. Scout, developed by Amazon, is making deliveries in parts of the US and Helsinki. Other companies using delivery robots include Domino's, FedEx, Walmart, and Grubhub. Delivery robots also might eventually be used to deliver mail.[226]

4. Conclusion

Greater competition and innovation in ocean freight transportation could enable the transportation system to become more efficient and the nation to fully realize the benefits of

---

[226] We are not aware of any issues that have arisen with picking up deliveries by drones and robots.

globalization and international trade. Deregulating ocean transportation and privatizing ports would provide incentives and opportunities to significantly improve the efficiency of ocean shipping. Autonomous water transportation could provide faster and more reliable service for cargo shipped from abroad and within the United States to ports. Ports could become more efficient by automating more of their operations. And autonomous trucks could deliver and pick up shipments at US ports more easily and with less delay and safely deliver them to receivers or could coordinate with rail to provide intermodal operations. Autonomous railroads also could ship cargo from ports to North American destinations. Finally, drones and delivery robots could increase the speed that smaller shipments and mail reach their final destinations.

Like other parts of the US transportation system, policymakers must reform their policies toward freight transportation to spur the vast improvements that would result from greater competition and innovation.

PART IV

SYNTHESIS AND POLICY RECOMMENDATIONS

Chapter 14. Synthesis and Policy Recommendations

Clifford Winston and Jia Yan

1. Introduction

The urban and intercity passenger and freight transportation systems are critical components of the transportation system that greatly affects the efficiency and equity of the US economy. Unfortunately, the US transportation system is often taken for granted by the public because it draws relatively little attention to itself—until it does.

During 2021, a heavier than expected snowfall caused motorists on East Coast highways to be stuck for days; US airline pilots who were infected by COVID-19 and could not work caused airlines to cancel thousands of flights during the Christmas holidays and created long delays to reschedule flights; and a surge in aggregate demand and logistical problems at ports led to congestion and delays that disrupted the supply chain, caused product shortages, and increased producer and consumer prices. When people and freight cannot move, travelers, shippers, and transportation modes are adversely affected as is the broader economy by lower productivity, higher inflation, and fewer personal interactions that could lead to new ideas and innovations.

This book has argued that competition and innovation have enabled the urban and intercity passenger and freight transportation systems to improve significantly over time and that new sources of competition and innovation could lead to additional improvements, which also could enable the system to rid itself of persistent inefficiencies and to be more resilient and less likely to experience future calamities.

Airline, intercity bus, trucking, and rail deregulation; open skies agreements on international airline routes; and the emergence of ridesharing have reduced transportation costs

and improved travel time and reliability. As a result, more people and more freight have moved more cheaply, quickly, and reliably and revitalized all sectors of the economy. Importantly, this significant improvement has been accomplished without compromising safety.

We discussed new sources of competition and innovation that could further improve the system if policymakers adopted policy reforms that expedite the adoption of autonomous modes, air taxis, and drones; negotiate open skies agreements with all countries and grant cabotage rights to foreign airlines to begin the process of fully deregulating the global air transportation system; eliminate the Jones Act and rate conferences in ocean shipping; and privatize airports, air traffic control, and ports.

Consider how some of those improvements could have reduced the effects of the calamities that occurred in 2021. Foreign airlines granted cabotage rights could have helped passengers whose flights were cancelled to obtain alternative flights more quickly to complete their journey. Privatized US ports and ships operating without rate conferences and the Jones Act's entry restrictions would have the incentive and ability to improve their operations and significantly reduce the time it takes to unload and load ships' cargo and get it moving to their final destinations. And the adoption of air taxis and drones could have provided travelers and truckers who were stuck on a highway with emergency transportation services if they were needed.

Historically, government has regulated prices, entry, and exit of private modes and owned and operated public modes and infrastructure because much of the transportation system appeared to be subject to market failures. However, government regulatory and operating policy inefficiencies gradually became transparent, technological change and new sources of competition indicated that some transportation markets had the potential to correct their alleged failures and the government's actual failures, and those markets proceeded to do so when given the

opportunity. This theme has consistently played out in parts of the urban, intercity, and freight components of the transportation system and it could play out more fully in all the components.

The theme also may eventually play out as transportation in outer space evolves. Poole (2022) compares the approaches taken by the National Aeronautics and Space Administration (NASA), a federal government agency, and the private sector. NASA has been using obsolete technology, left over from the space-shuttle program, to launch its Space Launch System moon rocket. Failure of its obsolete technology forced NASA to cancel its scheduled launch twice. US private sector companies, including SpaceX and Blue Origin, are launching rockets for commercial travel in outer space and, along with startup companies, are planning to go to the Moon and Mars. Private sector companies are competing on innovation and driving down costs, while NASA is stuck in the past and inflating costs to taxpayers. NASA may eventually improve the efficiency and performance of space travel by relying on services provided by a competitive market in space travel.

2. Summarizing New Sources of Competition and Innovation and Their Potential Benefits

The new sources of competition and innovation that we discussed, their potential to address the inefficiencies of the current transportation system, and their potential benefits to travelers, shippers, and the broader economy can be summarized as follows.

Urban Transportation

Public transit's subsidies continue to grow, and its market share continues to shrink. Automobile externalities of congestion, accidents, and emissions continue to generate large social costs. We provided evidence that ridesharing has benefited travelers by providing a superior service to taxis and transit. Electric and autonomous vehicles could build on ridesharing and

eliminate transit subsidies by enabling travelers to effectively have their own transit service and could dramatically reduce congestion, accident, and emissions externalities.

The improved mobility from autonomous electric vehicles could generate enormous benefits to the broader economy and help alleviate social issues, such as social distancing during a pandemic and violent altercations arising from police stops. It is possible that the full benefits from autonomous electric vehicles would be more likely to be realized if following successful experiments, highways were privatized. Thus, competition, if it worked effectively, would encourage highway operators to adopt efficient operations and modern technology, and to manage finances so the highway system does not experience financial crises that result in widespread bankruptcies.

Intercity Transportation

Airline and bus deregulation transformed intercity passenger transportation into an efficient, highly competitive system, which generated large benefits to travelers. We provided evidence that negotiating open skies agreements with all countries and expanding deregulation in air transportation by granting cabotage rights to foreign airlines could provide further benefits to air travelers. We also provided evidence that privatizing airports could potentially lead to more airline competition and innovations, such as heated runways, that benefit travelers, and we argued that privatizing air traffic control could make airline operations' safer and more efficient.

Deregulation also transformed trucking and railroads into more efficient and innovative services. We provided evidence that autonomous trucking operations could substantially reduce shippers' inventory logistics costs, and provide other benefits, and we argued autonomous railroad operations would strengthen rail's competitive position with autonomous trucks while further benefiting shippers.

Creating a more competitive environment by eliminating the Jones Act and rate conferences and privatizing ports would cause ocean freight transportation to become more efficient. Autonomous ships and port operations would add to the improvement. Finally, the adoption of new modes, air taxis and drones, would enrich the transportation system by giving travelers and shippers new faster transportation options that would overcome the cost of distance more effectively than current options.

As in the case of the urban system, the greatly improved intercity system would generate enormous benefits to the broader economy by, for example, lowering the cost and expanding the volume of domestic and international trade and increasing agglomeration economies by facilitating more productive engagements between Americans and between Americans and individuals from foreign countries. More seamless airline and ocean travel would reduce the large costs of transferring passengers and freight in international markets.

Table 1 summarizes the sources of new competition and innovation that could improve the US transportation system and their potential welfare gains. Of course, considerable uncertainty is associated with the estimates, but previous experience with deregulation indicated that predicted benefits were significantly underestimated in large part because analysts did not anticipate the full range of innovative responses that led to significant reductions in costs and improvements in service quality. The potential benefits from the new sources we discussed are so huge, especially from automation, that even if we have significantly overestimated the benefits, welfare gains for the United States would still amount to hundreds of billions of dollars.

The major obstacles to obtaining those benefits are whether the innovations in autonomy will work in practice and whether policymakers will adopt policies that enable competition and innovation to flourish.

3. Criticisms of Autonomous Transportation Innovations

Autonomous transportation modes represent a large part of the innovations that we have argued could greatly benefit the US transportation system. However, we identified in chapter 6 the existence of naysayers who have made harsh comments about autonomous vehicles, such as, they are "one of the most hyped technology experiments of this century," and are "incapable of reversing the growing death toll on American roads for many years to come—if ever." We are not aware of similar comments directed toward autonomous trucks, trains, ocean ships, drones, and air taxis, but it is likely that the naysayers have similar attitudes toward those modes.

As noted, we agree with Bishop (2022) that it is premature to make any strong predictions about what will happen when autonomous vehicles are deployed. At the same time, the history of technological innovations offers some useful perspective. To the best of our knowledge, with the possible exception of the nuclear fusion industry, no US industry, or in the case of autonomous transportation modes, no global industry has made a large-scale investment in a new technology that generated zero returns because it failed to produce a product or service that was commercially viable.

Instead, there are many examples where firms within an industry invested in a new technology that failed because it was replaced by a somewhat different technology that led to a similar but better product. In other words, there is considerable learning when firms seek to develop a new technology and there are plenty of failures until commercial success is achieved. However, while some firms fail in their efforts to develop a new technology to produce a new product or service, entire industries have yet to fail when they have made substantial investments in a new technological innovation.

As summarized by Eadicicco et al. (2017), examples of industries where firms failed but other firms subsequently succeed include but are not limited to: BlackBerry, which was replaced by iPhone; AOL, which was replaced by Gmail and Hotmail; Palm Pilot, which was replaced by iPod and iPhone; Netscape, which was replaced by Microsoft; MySpace, which was replaced by Facebook; Segways, which were replaced by electric scooters and hoverboards; MapQuest, which was replaced by Google Maps; and TiVo, which was replaced by DVR. The process of competition, innovation, and leapfrog technologies that eclipse previous technologies will continue for the foreseeable future.

When we consider the future of autonomous modes, it is possible that some or even several automotive and technology firms will fail, for example, to develop a commercially safe and viable autonomous vehicle. However, we think it is extremely unlikely that the global autonomous vehicle industry will lose its entire $100+ billion investment because no firm in the world can build a safe autonomous vehicle that was a significant advance over a non-autonomous vehicle. We also doubt that consumer demand for autonomous vehicles will be insufficient to enable a producer to make an adequate return on its investment.

An important feature of the autonomous vehicle industry is the extensive collaboration between producers and between producers and users. The former is important for improving the technology and for recognizing and overcoming technological challenges. The latter is important for encouraging society to realize the potential benefits of autonomous vehicles and to have a vested interest in their success.

Users of autonomous cars and shuttle services in experiments in various cities are unlikely to generate broad interest and excitement in the technology. However, shipping companies that collaborate with autonomous trucking services in experiments in various states can envision the

potentially large benefits in reduced inventory and logistics costs from autonomous trucking, which could generate interest and excitement throughout private industry and translate into political support that could expedite testing and adoption before autonomous cars are adopted. Thus, autonomous trucks are likely to be the first autonomous vehicle operations to refute the naysayers' predictions and to eventually lead the way for autonomous cars to do so also.

Other autonomous, and non-autonomous, transportation technologies that we identified as emerging in intercity passenger and freight transportation are less developed both technologically and financially than autonomous cars and trucks, so considerable uncertainty still exists about their long-run commercial viability. For example, the current valuation of the emerging eVTOL industry is only a fraction of the commercial helicopter business, which it would be expected to displace if eVTOLs becomes a commercially viable transportation technology. In addition, new eVTOL infrastructure, vertiports, must attract investors and overcome public resistance to noise created by eVTOL operations. As suggested by Robert Poole, eVTOLs may be best suited commercially for regional instead of urban markets to connect cities within a few hundred miles and to use existing airports for their operations.[227]

Like any new technology, eVTOLs have both time and the potential to attract additional investment and human capital to help solve their various challenges to becoming a viable transportation service. Generally, the greatest threat in our view to the widespread adoption and commercial success of new transportation technologies is poor government policy that impedes their development and adoption not the failure of the technology.

---

[227] https://reason.org/aviation-policy-news/aviation-policy-news-vertiports-for-regional-air-mobility-congress-oblivious-to-changing-airline-competition-and-more/

4. Government Policy Recommendations and Reality

Given that we have argued that competition and innovation have been the primary sources of improvements in the US transportation system, it naturally follows that government policy that is attuned to improving the performance of the transportation system should stimulate new ways that urban, intercity, and freight transportation can be more competitive and innovative.

<u>Policy Recommendations</u>

We have made specific immediate recommendations that policymakers increase competition by expanding open skies agreements, granting cabotage rights, privatizing airports and seaports and exploring highway privatization, and eliminating the Jones Act and rate conferences. We also have recommended that policymakers increase innovation by privatizing air traffic control to modernize the system.

Looking toward the future, we have called for policymakers to increase competition and innovation by establishing testing procedures, and modernizing highway, air, and water infrastructure to expedite the adoption and increase the performance of autonomous cars, trucks, air taxis, drones, and ships, and to implement efficient pricing and investment policies to enhance the efficiency of those autonomous operations.

Unfortunately, policymakers have generally eschewed our immediate policy recommendations because they prefer to address transportation problems by increasing public spending and intervening to correct alleged market failures. Both policy approaches are often wasteful, and they have not significantly improved the transportation system.

Specifically, in November 2021, President Biden signed the $1 trillion Infrastructure Investment and Jobs Act, which increased government spending on roads, bridges, electric charger stations, airports, seaports, public transit, and passenger rail. However, without efficient pricing

and investment guidelines, spending on roads, bridges, airports, and seaports will be excessive and unlikely to reduce congestion and improve travel. Spending public funds on electric charging stations is inappropriate because the private sector should be the sole investor in those stations as motorists' increase their adoption of electric vehicles. And greater public expenditures for bus and rail transit are likely to increase those modes' deficits and do little to increase their market shares.

The STB's concerns about rail competition are misplaced and if they implement a policy that forces a railroad to provide reciprocal switching, it is likely the policy would do more harm to shippers than good. Instead of taking steps to spur competition in ocean transportation by deregulating it, Congress is considering legislation that would hand more power to the Federal Maritime Commission by authorizing it to act against alleged anticompetitive behavior, require shipping companies to comply with certain service standards, and regulate how they impose certain fees on their customers. Again, it is unclear how more government oversight and enforcement will bring down shipping rates and costs, which are being driven in large part by soaring consumer demand and persistent bottlenecks in an inefficient transportation environment.

Policymakers have yet to take any of the actions that we have recommended that have an eye toward the future to prepare the nation for autonomous transportation. NHTSA will have an opportunity to show its interest in expediting autonomous vehicle adoption in its response to GM's Cruise's recent petition for permission to put its Cruise Origin (Level 4 full automation within a restricted operational design domain) into commercial service. If NHTSA grants GM's petition, it could inform how NHTSA will grant future petitions for passenger AVs, and how NHTSA will proceed in future rulemakings to update Federal Motor Vehicle Safety Standards to begin the

process of fully integrating AVs into the nation's vehicle fleet. If NHTSA rejects or delays the petition, it will signal the federal government's disinterest in expediting AV adoption.[228]

Reality

It is important to acknowledge that, in the immediate future, policymakers are highly unlikely to adopt comprehensive reforms of their transportation policies in ways that are broadly consistent with our recommendations. Because cities compete with other cities by seeking status that is associated with having an urban rail transit system, they go to great lengths to continue to fund their systems, although they operate with large deficits and low ridership.

In chapter 3, we found that cities are willing to raise bus transit costs to support rail transit operations. In addition, the New York and Washington, D.C., metropolitan areas, for example, plan to use an influx of toll revenues from new automobile pricing programs to reduce rail transit deficits and fund system improvements. Thus, it is unrealistic to envision that policymakers will take any actions to reduce public transit subsidies by setting efficient fares and curtailing service where little demand for it exists. Similarly, policymakers pay little attention to the inefficiencies associated with publicly owned and operated airports, highways, and ports. They appear to perceive that they may benefit politically from continuing to spend more money on them.

So, what realistic possibilities exist for competition and innovation to further improve the US transportation system? Government has been inclined to improve transportation by simply getting out of the way, which suggests that it is possible that policymakers would be willing to further deregulate transportation by expanding open skies agreements, granting cabotage rights, and eliminating the Jones Act and rate conferences. To be sure, the government is not eager to

---

[228] NHTSA's delays can amount to nearly a decade. For example, in 2013, Toyota first petitioned NHTSA to allow for adaptive beam lights. Finally, in 2022, NHTSA issued a final rule to allow adaptive beam lights, which had been used for two decades in Europe and Japan, to update the Federal Motor Vehicle Safety Standards that had allowed only high- and low-beam lights.

seek market solutions to problems it has not solved or has created, and it is prone to criticisms from many ideological opponents of free markets.[229]

The federal government appeared to seriously consider corporatizing air traffic control, in part because Nav Canada served as a credible alternative model. Reforming ATC is still a possibility in the future. Privatization of airports, highways, and ports also may become a possibility if government at lower levels conducts privatization experiments that show that privately run transportation infrastructure facilities perform significantly better than publicly run facilities. Alternatively, a state or city government that experiences a major budgetary shortfall may decide to address it by selling a major infrastructure facility to a private operator. If the private operator improves the facility's performance, then other states or cities might consider privatizing major infrastructure facilities.

Markets have shown that if government is unable or unwilling to reduce transportation inefficiencies, then the private sector is willing to attempt to do so. Ridesharing was an innovation from the private sector that improved urban transportation, which, except for private cars, had generally been provided or regulated solely by the government. Autonomous modes have the potential to greatly improve the entire system and private market participants are proceeding without any government assistance. Hopefully, successful demonstrations of the technology will encourage policymakers to do their part to facilitate widespread adoption.

A plausible path therefore exists where new competition produced by additional deregulation and privatization and new innovations derived from artificial intelligence could revitalize the US transportation system. It also is possible that policymakers could make that path

---

[229] Winston (2022) provides examples of the *New York Times* publishing essays that provide a biased discussion against deregulated transportation markets and points out that the *New York Times* was unwilling to publish a response to point out and correct the biases because they generally avoid pieces criticizing the media.

easier; however, they are more likely to do so if they are persuaded that the political benefits from greater international or domestic prestige would be large and that the political costs would be small.

Historically, a select group of visionary political and transportation industry leaders have been able to successfully make the political case for instituting transportation policies that improved the system's performance. It is time for a new generation of leaders to come forward to do the same.

Table 1. New Sources of Competition and Innovation and Potential Welfare Gains

Urban Transportation	*Potential Welfare Gains*
Rideshare Companies	Travelers benefit several billion dollars annually
Autonomous Electric Shared Vehicles	GNP could increase at least 1% annually and the cost of automobile emissions and accident externalities would be substantially reduced
Intercity Transportation	
Granting Cabotage Rights	Travelers benefit at least $1.5 billion annually, not including greater benefits from seamless travel
Privatizing Airports	Travelers could benefit by several billion dollars annually from private airport competition
Privatizing Air Traffic Control	Estimates not yet available
Freight Transportation	
Autonomous Trucking Operations	Shippers would gain at least $140 billion in lower inventory logistics costs
Autonomous Rail and Ocean Operations	Estimates not yet available
Eliminating the Jones Act and Rate Conferences	Benefits to shippers would be several billion dollars annually
Privatization of Ports	Estimates not yet available
Drones and Air Taxis	Estimates not yet available

References

Adadie, Alberto, and Guido Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economic Statistics* 29 (1): 1–11.

Adler, M., S. Peer and T. Sinozic. 2019. "Autonomous, Connected, Electric Shared Vehicles (ACES) and Public Finance: An Explorative Analysis", *Transportation Research Interdisciplinary Perspectives*, Vol. 2, 1-13.

Alderighi, Marco, Allessandro Cento, Peter Nijkamp, and Piet Pietveld. 2012. "Competition in the European Aviation Market: The Entry of Low-Cost Airlines." *Journal of Transport Geography* 24: 223–33.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129 (3): 1085–140.

Anderson, Michael L. 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review* 104 (9): 2763–96.

Anderson, Michael L., and Lucas W. Davis. 2021. "Uber and Alcohol-Related Traffic Fatalities." Working Paper 29071. National Bureau of Economic Research.

Angrist, Joshua D., Sydnee Caldwell, and Jonathan V. Hall. 2021. "Uber versus Taxi: A Driver's Eye View." *American Economic Journal*: *Applied Economics* 13 (3): 272–308.

Aryal, Gaurab, Federico Cilberto, and Benjamin T. Leyden. Forthcoming. "Coordinated Capacity Reductions and Public Communication in the Airline Industry." *Review of Economic Studies*.

Aryal, Gaurab, Charles Murry, and Jonathan W. Williams. 2021. "Price Discrimination in International Airline Markets." Working paper. https://ssrn.com/abstract=3288276

Austen, Ian. 2022. "Canada's Slow and Troubled Path to Rapid Transit," *New York Times*, July 9.

Azar, José, Martin C. Schmalz, and Isabel Tecu. 2018. "Anticompetitive Effects of Common Ownership." *Journal of Finance* 73 (4): 1513–65.

Bachwich, Alexander R., and Michael D. Wittman. 2017. "The Emergence and Effects of the Ultra-Low Cost Carrier (ULCC) Business Model in the U.S. Airline Industry." *Journal of Air Transport Management* 62: 155–64.

Baker, Jonathan B., and Timothy F. Bresnahan. 1985. "The Gains from Merger or Collusion in Product-Differentiated Industries." *Journal of Industrial Economics* 33 (4): 427–44.

Baranowski, Michael R., and Nathaniel S. Zebrowski. 2022. "Reciprocal Switching." Verified Statement and Written Testimony before the Surface Transportation Board, STB Ex Parte No. 711 (Sub-No. 1), February 14.

Barrett, Sean D. 2004. "The Sustainability of the Ryanair Model." *International Journal of Transport Management* 2 (2): 89–98.

Baum-Snow, Nathaniel, and Matthew E. Kahn. 2005. "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970–2000." *Brookings–Wharton Papers on Urban Affairs*: 147–206.

Baumgaertner, Emily, and Russ Mitchell. 2021. "Car Crash Deaths have Surged during Covid-19 Pandemic. Here's Why." *Los Angeles Times*, December 8.

Bereskin, C. Gregory. 2001. "Sequential Estimation of Railroad Costs for Specific Traffic." *Transportation Journal* 40 (3): 33–45.

Berger, Paul. 2022. "A Deep Divide on Automation Hangs Over West Coast Port Labor Talks," *Wall Street Journal*, June 9.

Bishop, Richard. 2022. "Self-Driving Cars: An Epidemic of Questionable Assertions." *Forbes*, February 21.

Bitzan, John D., and Theodore E. Keeler. 2003. "Productivity Growth and Some of Its Determinants in the Deregulated U.S. Railroad Industry." *Southern Economic Journal* 70 (2): 232–53.

Blinder, Alan S., and Louis J. Maccini. 1991. "Taking Stock: A Critical Assessment of Recent Research on Inventories." *Journal of Economic Perspectives* 5 (1): 73–96.

Blundell, Richard W., and James L. Powell. 2004. "Endogeneity in Semiparametric Binary Response Models." *Review of Economic Studies* 71 (3): 655–79.

Boguslaski, Charles, Harumi Ito, and Darin Lee. 2004. "Entry Patterns in the Southwest Airlines Route System." *Review of Industrial Organization* 25 (3): 317–50

Borrás, Jo. 2022. "BP Claims EV Charging Stations 'On the Cusp' of Being More Profitable than Gas Pumps." *Electrek*, January 16.

Bosman, Julie, Sophie Kasakove, Jill Cowan, and Richard Fausset. 2022. "Cities Want to Return to Prepandemic Life. One Obstacle: Transit Crime." *New York Times,* April 25.

Brinkman, Jeffrey, and Jeffrey Lin. 2022. "The Costs and Benefits of Fixing Downtown Freeways." *Economic Insights* (Federal Reserve Bank of Philadelphia), Quarter 1: 17–22.

Brueckner, Jan K., Darin Lee, and Ethan Singer. 2014. "City-Pairs versus Airport-Pairs: A Market-Definition Methodology for the Airline Industry." *Review of Industrial Organization* 44 (1): 1–25.

Brueckner, Jan K., Darin Lee, and Ethan S. Singer. 2013. "Airline Competition and Domestic US Airfares: A Comprehensive Reappraisal." *Economics of Transportation* 2 (1): 1–17.

Bunten, Devin Michelle, Ellen Fu, Lyndsey Rolheiser, and Christopher Severen. 2022. "The Problem Has Existed over Endless Years: Racialized Difference in Commuting, 1980–2019." Working paper.

Burtraw, Dallas. 1996. "The SO_2 Emissions Trading Program: Cost Savings without Allowance Trades." *Contemporary Economic Policy* 14 (2): 79–94.

Carlton, Dennis, Mark Israel, Ian MacSwain, and Eugene Orlov. 2019. "Are Legacy Airline Mergers Pro- or Anti-Competitive? Evidence from Recent U.S. Airline Mergers." *International Journal of Industrial Organization* 62: 58–95.

Chandar, Bharat, Uri Gneezy, John A. List, and Ian Muir. 2019. "The Drivers of Social Preferences: Evidence from a Nationwide Tipping Field Experiment." Working Paper 26380. National Bureau of Economic Research.

Cho, Woohyun, Robert J. Windle, and Martin E. Dresner. 2015. "The Impact of Low-Cost Carriers on Airport Choice in the US: A Case Study of the Washington–Baltimore Region." *Transportation Research Part E: Logistics and Transportation Review* 81: 141–57.

Cho, Woohyun, Robert J. Windle, and Christian Hofer. 2012. "Route Competition in Multi-Airport Cities: An Analysis of US Air Fares." *Transportation Journal* 51 (3): 265–88.

Choi, Ki-Hong and Choon-Geol Moon. 1997. "Generalized Extreme Value Model and Additively Separable Generator Function." *Journal of Econometrics* 76 (1): 129–40.

Cline, Andrew. 2022. "Biden's EV Push May Hit Mining Regs Pothole." *Inside Sources*, April 8.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. 2016. "Using Big Data to Estimate Consumer Surplus: The Case of Uber." Working Paper 22627. National Bureau of Economic Research.

Cook, Cody, Rebecca Diamond, Jonathan V. Hall, John A. List, and Paul Oyer. 2021. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers." *Review of Economic Studies* 88 (5): 2210–38.

Correal, Annie. 2018. "Inside the Dollar Van Wars." *New York Times*, June 8.

Cosslett, Stephen R. 1983. "Distribution-Free Maximum Likelihood Estimation of the Binary Response Model." *Econometrica* 51 (3): 765–82.

Cramer, Judd, and Alan B. Krueger. 2016. "Disruptive Change in the Taxi Business: The Case of Uber." *American Economic Review* 106 (5): 177–82.

Davidson, Russell, and James G. MacKinnon. 1981. "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica* 49 (3): 781–93.

Davis, River. 2021. "A City Tailor-Made for Self-Driving Cars? Toyota is Building One," *Bloomberg,* April 26.

de Freytas-Tamura, Kimiko. 2019. "Can 'Scooby Doo' and the Rest of the Dollar Vans go High-Tech?" *New York Times*, December 11.

Dorsey, Jackson, Ashley Langer, and Shaun McRae. 2022. "Fueling Alternatives: Gas Station Choice and the Implications for Electric Charging." Working Paper 29831. National Bureau of Economic Research.

Dresner, Martin, Jiun-Sheng Chris Lin, and Robert Windle. 1996. "The Impact of Low-Cost Carriers on Airport and Route Competition." *Journal of Transport Economics and Policy* 30 (3): 309–28.

Eadicicco, Lisa, Matt Peckham, John Patrick Pullen, and Alex Fitzpatrick. 2017. "The 20 Most Successful Technology Failures of All Time." *Time*, April 3.

Engel, Eduardo, Ronald D. Fischer, and Alexander Galetovic. 2021. "When and How to Use Public-Private Partnerships in Infrastructure: Lessons from the International Experience," In *Economic Analysis and Infrastructure Investment*, edited by Edward L. Glaeser and James M. Poterba. University of Chicago Press.

Ewing, Jack. 2022. "Truck Makers Face a Tech Dilemma: Batteries or Hydrogen?" *New York Times*, April 11.

Ewing, Jack, and Neal E. Boudette. 2022. "Why This Could Be a Critical Year for Electric Cars," *New York Times*, February 8.

Fink, Carsten, Aaditya Mattoo, and Ileana Cristina Neagu. 2002. "Trade in International Maritime Services: How Much Does Policy Matter?" *World Bank Economic Review* 16 (1): 81–108.

Friedlaender, Ann F. 1965. *The Interstate Highway System: A Study in Public Investment*. North-Holland.

Friedlander, Ann F., and Richard Spady. 1981. *Freight Transport Regulation: Equity, Efficiency, and Competition in the Rail and Trucking Industries*. MIT Press.

Friedman, Milton, and Daniel J. Boorstin. 1996. "How to Plan and Pay for the Safe and Adequate Highways We Need." In *Roads in a Market Economy*, edited by Gabriel Roth. Avebury Technical.

Gao, Yang, and David Levinson. 2021. "COVID-19, Travel Time Reliability, and the Emergence of a Double-Humped Peak Period." *Transport Findings*. https://doi.org/10.32866/001c.27013

Gelinas, Nicole. 2022a. "Whatever Happened to Congestion Pricing?," *City Journal*, June 23.

Gelinas, Nicole. 2022b. "Train Wrecks," *Washington Examiner*, June 2.

George, Justin. 2022. "1 in 3 Metrobus Rides Goes Unpaid amid Rise in Fare Evasions during Pandemic." *Washington Post*, February 8.

Gittleman, Maury, and Kristen Monaco. 2019. "Automation Isn't About to Make Truckers Obsolete." *Harvard Business Review*, September 18.

Gómez-Ibáñez, José A., and John R. Meyer. 1984. *Autos, Transit, and Cities*. Harvard University Press.

Goolsbee, Austan, and Chad Syverson. 2008. "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." *Quarterly Journal of Economics* 123 (4): 1611–33.

Gorback, Caitlin. 2021. "Ridesharing and the Redistribution of Economic Activity." Working paper.

Gore, Christina. 2021. "What Drives Battery Electric Vehicle Adoption? Willingness to Pay to Reduce Emissions Through Vehicle Choice." Job Market Paper. Ohio State University.

Green, Richard K. 2007. "Airports and Economic Development." *Real Estate Economics* 35 (1): 91–112.

Greene, William, H. 2003. *Econometric Analysis* (5th ed.). Prentice Hall.

Grimm, Curtis, and Clifford Winston. 2000. "Competition in the Deregulated Railroad Industry: Sources, Effects, and Policy Issues." In *Deregulation of Network Industries: What's Next*, edited by Sam Peltzman and Clifford Winston. Brookings Institution.

Grimm, Curtis, Clifford Winston, and Carol Evans. 1992. "Foreclosure of Railroad Markets: A Test of Chicago Leverage Theory." *Journal of Law and Economics* 35 (2): 295–310.

Groom, Nichola, and Tina Bellon. 2021. "EV Rollout Will Require Huge Investments in Strained US Power Grids." *Reuters*, March 5.

Hall, Jonathan D., Craig Palsson, and Joseph Price. 2018. "Is Uber a Substitute or Complement for Public Transit?," *Journal of Urban Economics* 108 (6): 36-50.

Hall, Jonathan V., and Alan B. Krueger. 2018. "An Analysis of the Labor Market for Uber's Driver-Partners in the United States." *Industrial and Labor Relations Review* 71 (3):705–32.

Holland, Stephen P., Erin T. Mansur, and Andrew J. Yates. 2021. "The Electric Vehicle Transition and the Economics of Banning Gasoline Vehicles." *American Economic Journal: Economic Policy* 13 (3): 316–44.

Holland, Stephen P., Erin T. Mansur, Nicholas Z. Muller, and Andrew J. Yates. 2020. "Decompositions and Policy Consequences of an Extraordinary Decline in Air Pollution from Electricity Generation." *American Economic Journal: Economic Policy* 12 (4): 244–74.

Holland, Stephen P., Erin T. Mansur, Nicholas Z. Muller, and Andrew J. Yates. 2016. "Are There Environmental Benefits from Driving Electric Vehicles? The Importance of Local Factors." *American Economic Review* 106 (12): 3700–29.

Hoppe Edward J. 2001. "Evaluation of Virginia's First Heated Bridge." *Transportation Research Record* 1741 (1): 199–206.

Horowitz, Joel L. 2011. "Applied Nonparametric Instrumental Variables Estimation." *Econometrica* 79 (2): 347–94.

Hughes, Jonathan E., Daniel Kaffine, and Leah Kaffine. 2022. "Decline in Traffic Congestion Increased Accident Severity in the Wake of COVID-19." *Transportation Research Record*

Ivaldi, Marc, and Gerard McCullough. 2001. "Density and Integration Effects on Class I U.S.Freight Railroads." *Journal of Regulatory Economics* 19: 161–82.

Jia, Panle. 2008. "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retail Industry." *Econometrica* 76 (6): 1263–316.

Jerch, Rhiannon, Matthew E. Kahn, and Shanjun Li. 2017. "The Efficiency of Local Government: The Role of Privatization and Public Sector Unions." *Journal of Public Economics* 154: 95–121.

Kahneman, Daniel and Alan B. Krueger. 2006. "Developments in the Measurement of Subjective Well-Being," *Journal of Economic Perspecitves*, volume 20, Winter, pp. 3-24.

Kashian, Russ, Jeff Pagel, and Ike Brannon. 2017. "The Jones Act in Perspective: A Survey of Costs and Effects of the 1920 Merchant Marine Act." Grassroot Institute of Hawaii. www.grassrootinstitute.org/2017/04/the-jones-act-in-perspective/

Keeler, Theodore E., and John S. Ying. 1988. "Measuring the Benefits of a Large Public Investment: The Case of the U.S. Federal-Aid Highway System." *Journal of Public Economics* 36 (1): 69–85.

Koustas, Dmitri, James Parrott, and Michael Reich. 2020. "New York City's Gig Driver Pay Standard: Effects on Drivers, Passengers, and the Companies." The New School, Center for New York City Affairs.

LaLonde, Bernard J., Martha C. Cooper, and Thomas G. Noordewier. 1988. *Customer Service: A Management Perspective*. Council of Logistics Management.

Langer, Ashley, Vikram Maheshri, and Clifford Winston. 2017. "From Gallons to Miles: A Disaggregate Analysis of Automobile Travel and Externality Taxes." *Journal of Public Economics* 152: 34–46.

Lave, Charles, and Lester Lave. 1999. "Fuel Economy and Auto Safety Regulation: Is the Cure Worse than the Disease?" In *Essays in Transportation Economics and Policy*: *A Handbook in Honor of John R. Meyer*, edited by José A. Gómez-Ibáñez, William B. Tye, and Clifford Winston. Brookings Institution Press.

Leard, Benjamin, and Jianwei Xing. 2020. "What Does Ridesharing Replace?" Working Paper 20-03. Resources for the Future.

Leslie, Alex and Dan Murray. 2022. "An Analysis of the Operational Costs of Trucking: 2022 Update," American Transportation Research Institute, Arlington, Virginia.

Levelton Consultants Limited. 2007. "Guidelines for the Selection of Snow and Ice Control Materials to Mitigate Environmental Impacts." Report 577. National Cooperative Highway Research Program. https://doi.org/10.17226/23178.

Levinson, Marc. 2006. *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton University Press.

Ley, Ana. 2022. "When the Only Way to Get to Work is This Slow Bus," *New York Times*, May 31.

Li, Shanjun, Lang Tong, Jianwei Xing, and Yiyi Zhou. 2017. "The Market for Electric Vehicles: Indirect Network Effects and Policy Design," *Journal of the Association of Environmental and Resource Economists,* volume 4, March, pp. 89-133.

Li, Ziru, Chen Liang, Yili Hong, and Zhongju Zhang. 2022. "How Do On-Demand Ridesharing Services Affect Traffic Congestion? The Moderating Role of Urban Compactness." *Production and Operations Management* 31 (1): 239–58.

Lieberman, Ben. 2021. "Why So Many Subsidies for Electric Vehicles?" *InsideSources*, February 23.

Lindsey, Robin. 2012. "Road Pricing and Investment." *Economics of Transportation* 1 (1): 49–63.

Liu, Meng, Erik Brynjolfsson, and Jason Dowlatabadi. 2021. "Do Digital Platforms Reduce Moral Hazard? The Case of Uber and Taxis." *Management Science* 67 (8): 4643–5300.

Lokshin, Michael M., and David Newsom. 2020. "Automated Transport Could Propel Development Forward. Can We Turn the Vision into Reality?" *World Bank Blogs*, August 11.

Lund, John W. 1999. "Reconstruction of a Pavement Geothermal Deicing System." *Geo-Heat Center Quarterly Bulletin* 20 (1): 14–17.

Luo, Dan. 2014. "The Price Effects of the Delta/Northwest Airline Merger." *Review of Industrial Organization* 44 (1): 27–48.

Maheshri, Vikram, and Clifford Winston. 2016. "Did the Great Recession Keep Bad Drivers Off the Road?" *Journal of Risk and Uncertainty* 52 (3): 255–80.

Malyshkina, Nataliya V., and Fred Mannering. 2008. "Effect of Increases in Speed Limits on Severities of Injuries in Accidents." *Transportation Research Record* 2083 (1): 122–27.

Mannering, Fred, and Clifford Winston. 1991. "Brand Loyalty and the Decline of American Automakers." *Brookings Papers on Economic Activity: Microeconomics*: 67–114.

Marsh, Joanna. 2022. "Railroads, Union Clash Over Use of Track Inspection Technology," *FreightWaves*, May 27.

Matulka, Rebecca. 2014. "The History of the Electric Car." US Department of Energy. https://www.energy.gov/articles/history-electric-car

Matzkin, Rosa L. 1992. "Nonparametric Identification and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models." *Econometrica* 60 (2): 239–70.

Minsk, L. David. 1999. "Heated Bridge Technology." Report FHWA-RD-99-158. Federal Highway Administration.

McWeeny, Dennis B. 2019. "Spatial Competition in the Airline Industry." Job Market Paper. University of Wisconsin. https://ssrn.com/abstract=3465104

Meredith-Karam, Patrick, Hui Kong, Shenhao Wang, and Jinhua Zhao. 2021. "The Relationship between Ridehailing and Public Transit in Chicago: A Comparison before and after COVID-19." *Journal of Transport Geography* 97: article 103219.

Meyer, John R., John F. Kain, and Martin Wohl. 1965. *The Urban Transportation Problem*. Harvard University Press.

Meyer, John R., Merton J. Peck, John Stenason, and Charles Zwick. 1959. *Economics of Competition in the Transportation Industries*. Harvard University Press.

Morrison, Steven A. 2001. "Actual, Adjacent, and Potential Competition: Estimating the Full Effect of Southwest Airlines." *Journal of Transport Economics and Policy* 35 (2): 239–56.

Morrison, Steven A., and Clifford Winston. 1985. "An Econometric Analysis of the Demand for Intercity Passenger Transportation." *Research in Transportation Economics* 2: 213–35.

Morrison, Steven A., and Clifford Winston. 1986. *The Economic Effects of Airline Deregulation*. Brookings Institution Press.

Morrison, Steven A., and Clifford Winston. 1987. "Empirical Implications and Tests of the Contestability Hypothesis." *Journal of Law and Economics* 30 (1): 53–66.

Morrison, Steven A., and Clifford Winston. 1995. *The Evolution of the Airline Industry*. Brookings Institution Press.

Morrison, Steven A., and Clifford Winston. 1999. "Regulatory Reform of U.S. Intercity Transportation." In *Essays in Transportation Economics and Policy*: *A Handbook in Honor of John R. Meyer*, edited by José A. Gómez-Ibáñez, William B. Tye, and Clifford Winston. Brookings Institution Press.

Morrison, Steven A., and Clifford Winston. 2000. "The Remaining Role for Government Policy in the Deregulated Airline Industry." In *Deregulation of Network Industries: What's Next?* edited by Sam Peltzman and Clifford Winston. Brookings Institution Press.

Morrison, Steven A., and Clifford Winston. 2007. "Another Look at Airport Congestion Pricing." *American Economic Review* 97 (5): 1970–77.

Morrison, Steven A., and Clifford Winston. 2008. "The Effect of FAA Expenditures on Air Travel Delays," *Journal of Urban Economics*, volume 63, March, pp. 669-678.

Morrison, Steven A., Clifford Winston, and Tara Watson. 1999. "Fundamental Flaws of Social Regulation: The Case of Airplane Noise," *Journal of Law and Economics*, volume 42 (2): pp. 723-744.

Nunes, Ashley, Lucas Woodley, and Philip Rossetti. 2022. "Re-thinking Procurement Incentives for Electric Vehicles to Achieve Net-Zero Emissions," *Nature Sustainability*, volume 5, pp. 527-532

Olatubi, Williams O., and David E. Dismukes. 2000. "A Data Envelopment Analysis of the Levels and Determinants of Coal-Fired Electric Power Generation Performance." *Utilities Policy* 9 (2): 47–59.

Oum, Tae H., Jia Yan, and Chunyan Yu. 2008. "Ownership Forms Matter for Airport Efficiency: A Stochastic Frontier Investigation of Worldwide Airports." *Journal of Urban Economics* 64 (2): 422–35.

Perrine, Kenneth, Kara Kockelman, and Yantao Huang. 2020. "Anticipating Long-Distance Travel Shifts Due to Self-Driving Vehicles." *Journal of Transport Geography* 82: article 102547.

Picard, Francois. 2022. "Driverless Trucks Set to Take Over Roads in Texas." *RTE News*, January 21. https://www.rte.ie/news/newslens/2022/0121/1275129-driverless-trucks/

Poole, Robert. 2022. "The Last Gasp of 20th-Century NASA," *Wall Street Journal,* September 1.

Porter, Robert H. 1983. "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886." *Bell Journal of Economics* 14 (2): 301–14.

Powell, James L. 1994. "Estimation of Semiparametric Models." In *Handbook of Econometrics, Volume 4*, edited by Robert F. Engle and Daniel L. McFadden. North-Holland.

Quiroz-Gutierrez, Marco. 2021. "It Could Take Nearly a Decade for Public Transit to Return to Pre-Pandemic Levels." *Fortune*, November 7.

Rapson, David S., and Erich Muehlegger. 2021. "The Economics of Electric Vehicles." Working Paper 29093. National Bureau of Economic Research.

Read, Richard. 2020. "A West Coast Union Faces Bankruptcy. Here's Why Unions Nationwide are Unnerved." *Los Angeles Times*, January 28.

Reitzes, James D., and Kelli L. Sheran. 2002. "Rolling Seas in Liner Shipping." *Review of Industrial Organization* 20 (1): 51–59.

Reiss, Aaron. 2014. "New York's Shadow Transit." *The New Yorker*, June 27.

Retallack, Angus Eugene, and Bertram Ostendorf. 2019. "Current Understanding of the Effects of Congestion on Traffic Accidents." *International Journal of Environmental Research and Public Health* 16 (19): 1–13.

Roback, Paul, and Julie Legler. 2021. *Beyond Multiple Linear Regression*: *Applied Generalized Linear Models and Multilevel Models in R*. CRC Press.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–S102.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39 (1): 33–38.

Rosenberg, Rebecca. 2022. "NYC Fare Beaters on Bus, Subway Lines Costing Taxpayers Millions." *Fox News*, February 23.

Rosenthal, Brian. 2017. "The Most Expensive Mile of Subway Track on Earth," *New York Times*, December 28.

Rothe, Christophe. 2009. "Semiparametric Estimation of Binary Response Models with Endogenous Regressors." *Journal of Econometrics* 153 (1): 51–64.

Rubin, Jonathan, Per E. Gårder, Charles E. Morris, Kenneth L Nichols, John M. Peckenham, Peggy McKee, Adam Stern, and T. Olaf Johnson. 2010. "Maine Winter Roads: Salt, Safety, Environment and Cost." Margaret Chase Smith Policy Center, University of Maine.

Ryder System and the Socially Aware Mobility Lab. 2021. "The Impact of Autonomous Trucking: A Case-Study of Ryder's Dedicated Transportation Network." White paper.

Sagers, Chris. 2006. "The Demise of Regulation in Ocean Shipping: A Study in the Evolution of Competition Policy and the Predictive Power of Microeconomics." *Vanderbilt Journal of Transnational Law* 39 (3): 779–818.

Said, Carolyn. 2019. "On-Transit Shuttles to Serve Apple Hometown's Transit Desert." *San Francisco Chronicle*, October 7.

Sarika, Kemal, and Ilhan Or. 2007. "Efficiency Assessment of Turkish Power Plants Using Data Envelopment Analysis." *Energy* 32 (8): 1484–99.

Scherer, Frederic M., and David Ross. 1990. *Industrial Market Structure and Economic Performance* (3rd ed.). Houghton Mifflin.

Schmidt, Stephen. 2001. "Market Structure and Market Outcomes in Deregulated Rail Freight Markets." *International Journal of Industrial Organization* 19 (1–2): 99–131.

Schulz, Dorothy Moses. 2022. "The Gentrification of Fare Evasion." *City Journal*, May 3.

Scribner, Marc. 2021a. "Increasing Access to Cars Advances more Equitable Outcomes." Reason Foundation, June 21.

Scribner, Marc. 2021b. "Pathways and Policy for 21st Century Freight Rail." Reason Foundation, September 8.

Shen, Weibin, Halil Ceylan, Kasthurirangan Gopalakrishnan, Sunghwan Kim, Peter C. Taylor, and Chris R. Rehmann. 2016. "Life Cycle Assessment of Heated Apron Pavement System Operations." *Transportation Research Part D: Transport and Environment* 48: 316–31.

Shen, Ying. 2017. "Market Competition and Market Price: Evidence from United/Continental Airline Merger." *Economics of Transportation* 10: 1–7.

Shepardson, David, Hyunyoo Jin, and Joseph White. 2022. "Self-Driving Car Companies Zoom Ahead, Leaving U.S. Regulators Behind." *Reuters*, February 2.

Shi, Xianming, Laura Fay, Chase Gallaway, Kevin Volkening, Marijean M. Peterson, Tongyan Pan, Andrew Creighton, Collins Lawlor, Stephanie Mumma, Yajun Liu, and Tuan Anh Nguyen. 2009. "Evaluation of Alternative Anti-Icing and Deicing Compounds Using Sodium Chloride and Magnesium Chloride as Baseline Deicers – Phase I." Report CDOT-2009-1. Colorado Department of Transportation.

Shirley, Chad, and Clifford Winton. 2004. "Firm Inventory Behavior and the Returns from Highway Infrastructure Investments." *Journal of Urban Economics* 55 (2): 398–415.

Siddiqui, Faiz. 2022. "How Auto Regulators Played Mind Games with Elon Musk." *Washington Post,* March 28.

Singh, Prashant, Maxim A. Dulebenets, Junayed Pasha, Ernesto D. R. Santibanez Gonzalez, Yui-Yip Lau, and Raphael Kampmann. 2021. "Deployment of Autonomous Trains in Rail Transportation: Current Trends and Existing Challenges." *IEEE Access* 9: 91427–61.

Skorup, Brent. 2022. "Drone Airspace: A New Global Asset Class," Center for the Study of Partisanship and Ideology, https://www.cspicenter.com/p/drone-airspace-a-new-global-asset

Small, Kenneth A., and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. Routledge.

Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73 (4): 1367–82.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell.

Smith, Kevin. 2019. "Rise of the Machines: Rio Tinto Breaks New Ground with AutoHaul." *International Railway Journal*, August 9.

Smyth, Mark, and Brian Pearce. 2008. "Air Travel Demand." Economics Briefing 9. International Air Transport Association.

Solow, Robert M. 1957. "Technological Change and the Aggregate Production Function." *Review of Economics and Statistics* 39 (3): 312–20.

Starkie, David. 2009. "The Airport Industry in a Competitive Environment: A United Kingdom Perspective." In *Competitive Interaction between Airports, Airlines, and High-Speed Rail*. Organization for Economic Cooperation and Development.

Swanson, Ana. 2022. "With Inflation Surging, Biden Targets Ocean Shipping." *New York Times*, March 21.

Sweeting, Andrew, James W. Roberts, and C. Gedge. 2020. "A Model of Dynamic Limit Pricing with an Application to the Airline Industry." *Journal of Political Economy* 128 (3): 1148–93.

Swietlik, William. 2010. *The Environmental Impacts of Airport Deicing—Water Quality,* US Environmental Protection Agency, https://apps.dtic.mil/sti/pdfs/ADA554090.pdf#:~:text=Aircraft%20deicing%20fluids%20also%20contain%20additives%2C%20and%20some,processes%20and%20the%20quality%20of%20finished%20drinking%20water.

Tajalli, Mehrdad, Ramin Niroumand, and Ali Hajbabaie. 2022. "Distributed Cooperative Trajectory and Lane Changing Optimization of Connected Automated Vehicles: Freeway Segments with Lane Drop," *Transportation Research Part C: Emerging Technologies*, volume 143, October.

Tang, Cheng Keat, and Jos van Ommeren. Forthcoming. "Accident Externality of Driving: Evidence from the London Congestion Charge." *Journal of Economic Geography*.

Tefft, Brian C., Leon Villavicencio, Aaron Benson, Lindsay Arnold, Woon Kim, Victoria Añorve, and William J. Horrey. 2022. "Self-Reported Risky Driving in Relation to Changes in Amount of Driving During the COVID-19 Pandemic." AAA Foundation for Traffic Safety.

Tett, Gillian. 2021. "Why America Has to Keep On Trucking." *Financial Times*, January 21.

Tongzon, Jose, and Wu Heng. 2005. "Port Privatization, Efficiency, and Competitiveness: Some Empirical Evidence from Container Ports (Terminals)." *Transportation Research Part A: Policy and Practice* 39 (5): 405–24.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation* (2nd edition). Cambridge University Press.

Tyworth, John E., and Liam O'Neill. 1997. "Robustness of the Normal Approximation of Lead-Time Demand in a Distribution Setting." *Naval Research Logistics* 44 (2): 165–86.

Tyworth, John E., and Amy Zhaohui Zeng. 1998. "Estimating the Effects of Carrier Transit-Time Performance on Logistics Cost and Service." *Transportation Research Part A: Policy and Practice* 32 (2): 89–97.

US Department of Transportation. 2016. "The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations—Revision 2 (2016 Update)."

US Department of Transportation. 1995. "Bus–Rail Integration," Departmental Guidance, January 17.

Verma, Pranshu. 2020. "We're Desperate: Transit Cuts Felt Deepest in Low-Income Areas." *New York Times*, August 15.

Vickrey, William. 1968. "Automobile Accidents, Tort Law, Externalities, and Insurance: An Economist's Critique." *Law and Contemporary Problems* 33: 464–87.

Viscelli, Steve. 2018. "Driverless? Autonomous Trucks and the Future of the American Trucker." Center for Labor Research and Education, University of California, Berkeley.

Viton, Philip A. 1981. "On Competition and Product Differentiation in Urban Transportation: The San Francisco Bay Area." *Bell Journal of Economics* 12 (2): 362–79.

Wang, Guihua, Ronghuo Zheng, and Tinglong Dai. Forthcoming. "Does Transportation Mean Transplantation? Impact of New Airline Routes on Sharing of Cadaveric Kidneys." *Management Science*.

Ward, Jacob W., Jeremy J. Michalek, and Constantine Samaras. 2021. "Air Pollution, Greenhouse Gas, and Traffic Externality Benefits and Costs of Shifting Private Vehicle Travel to Ridesourcing Services." *Environmental Science & Technology* 55 (19): 13174–85.

Ward, Jacob W., Jeremy J. Michalek, Constantine Samaras, Inês L. Azevedo, Alejandro Hanao, Clement Rames, and Tom Wenzel. 2021. "The Impact of Uber and Lyft on Vehicle Ownership, Fuel Economy, and Transit Across U.S. Cities." *iScience* 24 (1): article 101933.

Waschik, Robert, Catherine L. Taylor, Daniel Friedman, and Jasmine Boatner. 2021. "Macroeconomic Impacts of Automated Driving Systems in Long-Haul Trucking." Report FWHA-JPO-21-847. US Department of Transportation.

Windle, Robert J., and Martin E. Dresner. 1995. "The Short and Long Run Effects of Entry on U.S. Domestic Air Routes." *Transportation Journal* 35 (2): 14–25.

Winston, Clifford. 2022. "Debunking the Mainstream Liberal Media's Bias Against Free Markets." RealClear Markets, March 28.

Winston, Clifford. 2021a. "Comment on 'Can America Reduce Highway Construction Costs? Evidence from the States.'" In Economic Analysis and Infrastructure Investment, edited by Edward L. Glaeser and James M. Poterba. University of Chicago Press.

Winston, Clifford. 2021b. *Gaining Ground: Markets Helping Government*. Brookings Institution Press.

Winston, Clifford. 2021c. "Autonomous Vehicles." *Milken Institute Review* 23 (1): 56–65.

Winston, Clifford. 2013. "On the Performance of the U.S. Transportation System: Caution Ahead." *Journal of Economic Literature* 51 (3): 773–824.

Winston, Clifford. 2010. Last Exit: Privatization and Deregulation of the U.S. Transportation System. Brookings Institution Press.

Winston, Clifford. 2006. *Government Failure Versus Market Failure: Microeconomics Policy Research and Government Performance*. Brookings Institution Press.

Winston, Clifford, Thomas M. Corsi, Curtis M. Grimm, and Carol A. Evans. 1990. *The Economic Effects of Surface Freight Deregulation*. Brookings Institution Press.

Winston, Clifford, and Quentin Karpilow. 2020. *Autonomous Vehicles*: *The Road to Economic Growth?* Brookings Institution Press.

Winston, Clifford, and Ashley Langer. 2006. "The Effect of Government Highway Spending on Road Users' Congestion Costs," *Journal of Urban Economics*, volume 60, November, pp. 463-483.

Winston, Clifford, and Vikram Maheshri. 2007. "On the Social Desirability of Urban Rail Transit Systems." *Journal of Urban Economics* 62 (2): 362–82.

Winston, Clifford, and Fred Mannering. 2014. "Implementing Technology to Improve Public Highway Performance: A Leapfrog Technology from the Private Sector is Going to be Necessary." *Economics of Transportation* 3 (2): 158–65.

Winston, Clifford, and Chad Shirley. 1998. *Alternate Route: Toward Efficient Urban Transportation*. Brookings Institution Press.

Winston, Clifford, and Joan Winston. 2022. "The Same Old Technophobia Is Holding Back Autonomous Vehicles." *Barron's*, January 7.

Winston, Clifford, and Jia Yan. 2021. "Vehicle Size Choice and Automobile Externalities: A Dynamic Analysis." *Journal of Econometrics* 222 (1): 196–218.

Winston, Clifford, and Jia Yan. 2015. "Open Skies: Estimating Travelers' Benefits from Free Trade in Airline Services." *American Economic Journal: Economic Policy* 7 (2): 370–414.

Winston, Clifford, and Jia Yan. 2011. "Can Privatization of U.S. Highways Improve Motorists' Welfare?" *Journal of Public Economics* 95 (7–8): 993–1005.

Wooldridge, Jeffrey M. 1999. "Distribution-Free Estimation of Some Nonlinear Panel Data Models." *Journal of Econometrics* 90 (1): 77–97.

Yan, Chiwei, Vikrant Vaze, Allison Vanderboll, and Cynthia Barnhart. 2016. "Tarmac Delay Policies: A Passenger-Centric Analysis." *Transportation Research Part A: Policy and Practice* 83: 42–62.

Yan, Jia, and Clifford Winston. 2014. "Can Private Airport Competition Improve Runway Pricing? The Case of San Francisco Bay Area Airports." *Journal of Public Economics* 115: 146–57.

Yang, Ron. 2021. "(Don't) Take Me Home: Home Bias and the Effect of Self-Driving Trucks on Interstate Trade." Job Market Paper. Harvard Business School.

Zakharenko, Roman and Alexander Luttmann. 2022. "Downsizing the Jet: A Forecast of Economic Effects of Increased Automation in Aviation," January. https://www.researchgate.net/publication/358263085_Downsizing_the_jet_A_forecast_of_economic_effects_of_increased_automation_in_aviation)

Zhang, Jing, Debendra K. Das, and Rorik Peterson. 2009. "Selection of Effective and Efficient Snow Removal and Ice Control Technologies for Cold-Region Bridges." *Journal of Civil, Environmental, and Architectural Engineering* 3 (1): 1–14.

Zhang, Yunzhou, and Linda Nozick. 2018. "Investigating the Pricing Impacts of the American Airlines and US Airways Merger." *Transportation Research Record* 2672 (23): 15–19.

Zimmerman, Michael, Balika Sonthalia, Alberto Oca, Arsenio Martinez-Simon, Korhan Acar, and Yan Sun. 2021. "Change of Plans." Annual State of Logistics Report. Council of Supply Chain Management Professionals. https://www.kearney.com/transportation-travel/state-of-logistics-report

Zou, Li, and Chanyan Yu. 2020. "The Evolving Market Entry Strategy: A Comparative Study of Southwest and JetBlue." *Transportation Research Part A: Policy and Practice* 132: 682–95.