(Don't) Take Me Home: Home Preference and the Effect of Self-Driving Trucks on Interstate Trade

Ron Yang*

Version: Sep 7, 2022

Abstract

How will self-driving trucks affect U.S. interstate trade flows? I argue that human drivers' preferences to return home generate geographic specialization in the trucking freight market, and that self-driving trucks would eliminate this "home preference." I build a model of trucking carriers who make dynamic decisions about where to work, given that they prefer to be at home. A large home preference increases the value of driving places that are likely to bring one home and increases the value of taking time off at home. Using trucking freight transactions and highway inspections data, I estimate the model parameters and find that carriers value being at home at $90 per day or about one-half of their daily wage. In a counterfactual where self-driving trucks lower per-mile costs, increase daily driving range, and eliminate home preference, overall freight prices fall by 28.6 percent. Without home preference, carriers shift from driver-rich states to driver-poor states and total driving increases as carriers spend less time off. This paper therefore shows how automation can reshape transportation costs by removing the role of worker preferences.

*Contact: rnyang@stanford.edu. Acknowledgements: I am deeply grateful to my advisors Ariel Pakes, Edward Glaeser, Mark Egan, Myrto Kalouptsidi, and Robin Lee. I am also thankful to Oliver Hart, Gabriel Kreindler, Isaiah Andrews, Elie Tamer, Alex McKay, Pol Antras, Marc Melitz, Kyle Mangum, Erica Moszkowski, Frank Pinter, Gregor Schubert, Nicola Rosaia, Pedro Degiovanni, Hanbin Yang, and the attendees at the UEA North American Meeting, HBS Rising Scholars Conference, Asia-Pacific IO Conference, UEA European Meeting, and AREUEA National Conference for helpful conversations and comments. I thank LEAP for funding. This project would have been impossible without the generosity of many industry participants who provided information, ideas, and advice.

1 Introduction

Trucking freight is an integral part of the U.S. economy, responsible for over 80 percent of domestic shipments and the plurality of international trade. In normal times, this market is an input into almost every consumed good found on a supermarket or store shelf. The post-2020 supply chain disruptions have particularly highlighted its relevance, when a tight supply of truckers collided with increased demand from e-commerce to create long delays and high transport costs. I argue that a first-order feature of the trucking freight market is truckers' preferences to return home, which I call home preference. Unlike other transportation markets, where truckers live affects where they drive, because different routes present different opportunities to return home. This horizontal heterogeneity among truckers means that the distribution of where truckers live will affect shipping costs and trucking flows.

Studying home preference is particularly timely because trucking freight is an exciting frontier for self-driving technology. Technologically, freight trucks are more straightforward to automate than passenger vehicles because highways are more predictable than city streets. Financially, high labor costs and driver turnover present an economic case that has invited significant financial investment: In 2021, four self-driving truck firms — TuSimple, Plus, Embark, and Aurora — raised $5.7 billion in funding and were collectively valued at about $26 billion. TuSimple and Plus began fully driverless tests in 2021 and 2022 and expect to start commercial production of autonomous trucks capable of operating on highways without a human driver in 2024. Further, by eliminating the driver, self-driving trucks could break the link between drivers' homes and trucking prices generated by home preference. Given the rapid pace of innovation, understanding the impact of self-driving trucks on freight is essential today.

In this paper, I build and estimate a model of the trucking freight market which captures home preference. I find that trucking carriers value starting a day at home at $93 per day, or about one-half of the average daily wage. Compared to a reference setting without home preference, home preference causes carriers to shift operations from carrier-poor states to carrier-rich states. This relocation of carrier supply means, for example, that driver-rich California sees a 9 percent decrease in average export prices while driver-poor Massachusetts sees a 9.6 percent increase in average export prices. Carriers also take less time off because it is less attractive to stay at home. The increase in capital utilization results in an 8 percent increase in overall freight prices. In the context of a counterfactual which also includes the effect of self-driving trucks on daily driving ranges and per-mile costs, I find that eliminating home preference is responsible for about 25 percent of the total fall in overall prices.

I use several overlapping data sources to analyze this historically fragmented industry. For trucking prices and quantities, I use transactions from DAT RateView, which aggregates transaction data from shippers, carriers, and brokers. I combine this data with pre-transaction job postings and searches on DAT's online electronic marketplace (the largest spot marketplace for trucking freight). I merge this DAT data with highway inspection data from the Federal Motor Carrier Safety Administration, which offers insights into carriers' home locations and travel patterns. These datasets reveal how carriers' geographies — both where they are based and where they work — affect aggregate prices and quantities. I document three facts about the industry: carriers are concentrated in particular states, carriers are more likely to work in surrounding states, and the mix of home locations among carriers working in a given location predicts prices.

I build a dynamic model of trucking carriers and shippers across the United States, where carriers are differentiated by their home locations. Each day, carriers can either choose to work one of a set of available routes, or elect not to work. For every day they start at home, carriers receive a flow payoff which I call "home preference." Carriers that work a given route receive a market price and pay route-specific travel costs. After travelling, they begin again from the destination of their previous route. On the other side of the market, shippers are differentiated by the origins and destinations of the goods they need shipped, and they decide whether to ship based on market prices. I study a steady-state equilibrium where shippers and carriers optimize, prices equalize supply and demand along each route, and the geographic distribution of truckers is constant over time.

In this model, the home preference directly increases the value of being home for a given carrier, which indirectly increases the value of being in any location which is likely to bring the carrier home. In equilibrium, home preference spills across the economy and shapes the desirability of all locations through carrier value functions.

The price of shipping along a given origin-destination route therefore depends on both the desirability of the destination and the mix of carriers at the origin. For example, if a large share of carriers working in Washington state are from California, then the preferences of Californian carriers will have a larger weight on prices than those of carriers from other states.

Next, I estimate shipper and carrier parameters separately. On the shipper side, I estimate the price elasticity of trucking demand with linear IV. I use route snowiness, which makes driving harder and increases carrier costs, as a price instrument. I find that demand is relatively price-inelastic: A one percent increase in trucking freight prices yields a 0.57 percent decrease in quantity demanded.[1]

On the carrier side, the key parameters of interest are the home preference, the price sensitivity of supply, and the observable components of route costs: distance and diesel. I estimate these parameters using two moments from the data: the geographic distribution of carriers and carrier choice probabilities. Intuitively, the geographic distribution of carriers identifies the level of home preference. At higher levels of home preference, carriers specialize more geographically. This specialization is visible in the inspection patterns of carriers from different home locations. Highway inspections are therefore informative about the level of home preference. Meanwhile, carrier choice probabilities are informative about how carriers react to prices and route characteristics. Finally, I use the availability of rivers as a price instrument. Rivers allow barge shipping, thereby increasing shippers' outside options and generating a demand shock to trucking prices independent of trucking costs.

I find a level of home preference where carriers value a day at home at $93 per day — and considering that the average wage of a U.S. truck driver is $200 per day, this is sizeable. I also recover reasonable estimates of carrier costs. I estimate the marginal cost to carriers of distance at $1.65 per mile, which is in line with industry cost estimates of $1.55 per mile. Routes with higher local diesel prices also have higher costs.[2]

[1] My demand estimates build on previous studies of trucking freight demand (e.g., Oum (1979); Graham and Glaister (2004); Litman (2013)).

[2] My cost estimation builds on previous work on estimating cost functions for trucking (Spady and Friedlaender (1978); Chiang and Friedlaender (1984); Daughety and Nelson (1988); Gagne (1990)), as well as more general work studying the effect of U.S. highways on trade (Duranton and Turner (2012); Duranton et al. (2014); Fajgelbaum (2016); Allen and Arkolakis (2019)).

Using my parameter estimates, I study the effect of home preference on the U.S. trucking market by comparing a simulation of the economy without home preference to a baseline simulation. Two effects emerge when home preference is added. First, I document a reallocation effect wherein carriers shift operations from carrier-poor states to driver-rich states. For example, when home preference is added, I find that the share of all carriers working in California rises by 4.5 percent, and California's average export prices fall by 9 percent. The opposite scenario occurs in carrier-poor regions such as New England: Fewer carriers working in New England, and fewer carriers willing to travel away from home to New England, leads to higher shipping costs in the area. Quantitatively, Massachusetts sees an 11 percent decrease in carrier supply and a 9.6 percent increase in export prices when home preference is added. Second, there is a capital utilization effect as more carriers choose to take the outside option at home. Less driving decreases the supply of carriers across the economy. Overall, home preference raises average prices by 8 percent and decreases total shipping volume by 4 percent.

I apply these estimates to a first counterfactual where I study the effect of a vehicle miles tax on trucking prices. I find that a 19 cent per mile vehicle miles tax leads to a 7.1% increase in the cost of shipping, and this increase is concentrated in regions with long average lengths of haul. Home preference causes carriers to relocate away from places with higher tax incidence. The decline in local carrier supply therefore amplifies the effect of the tax.

Next, I apply my estimates to a counterfactual where I consider the effect of a transition to self-driving trucks. I focus on three effects of self-driving trucks: the elimination of home preference, changes in per-mile costs, and daily driving ranges. Based on previous work in the transportation literature, I reduce estimated per-mile costs by 25 percent (to account for the increase in efficiency) and increase the maximum daily driving time from 11 hours to 22 hours (to account for the decrease in rest time). Together, these three changes to the economy lower average shipping prices by 28.6 percent and contribute $16.65B to annual shipper profits. The fall in per-mile costs makes up roughly 50 percent of the total effect of automation on prices and quantities, while eliminating home preference contributes 25 percent and is comparable to the effect of doubling daily driving range. The effects are concentrated in states which have high average length of haul, which benefit the most from lower per-mile costs and driving range, and carrier-poor states, which benefit the most from eliminating home preferences. Relative to existing traffic, I predict that trucking traffic will rise the most in the Northwest and Northeast.

In summary, my paper suggests that ,beyond homogeneous cost reductions from labor savings, automation can reshape the geography of transportation costs by removing workers' preferences over where production occurs.

This paper contributes to several literatures. At the most applied level, I contribute to the literature studying the effect of autonomous vehicles[3]. Recent academic work has considered the effects of trucking automation on per-mile cost savings (Huang and Kockelman (2020)), labor turnover (Burks et al. (2018)), labor demand (Gittleman and Monaco (2020)), traffic and road utilization (Carrone et al. (2021)), emissions (Liu et al. (2018)), the rail industry (Bao and Mundy (2018)), the macroeconomy (Waschik et al. (2021)), and interactions with electrification (Ghandriz et al. (2020)). Other papers which study the effects of self-driving vehicles more broadly include Ostrovsky and Schwarz (2018) and Clements and Kockelman (2017). Winston and Karpilow (2020) highlights the potential for automation to decrease congestion externalities from transportation. On the labor side, Viscelli (2018) and Viscelli (2020) survey scenarios for the trucking labor market under automation. I push in a new direction by highlighting the effect of autonomation on geographic specialization by transportation firms.

More generally, my model contributes to the literature on competition in search and matching markets, building on Lagos (2003). Recent papers in this literature focusing on transportation markets have studied competition in oceanic shipping (Brancaccio et al. (2020)), taxicabs (Buchholz (2022); Fréchette et al. (2019)), rideshare markets (Rosaia (2020)), and railroads (Chen (2021)). In these markets, heterogeneity in the transportation firms has been relatively unimportant: Brancaccio et al. (2020) notes that home ports are unimportant for ocean shipping because crews fly home a few times per year, while Buchholz (2022) finds that although taxi cabs differ by home garage, these initial conditions "wash out" and become unimportant by

[3]A broader literature has studied the general interaction between information technology and firm organization in trucking and transportation markets (Hubbard (1998, 2000, 2001); Baker and Hubbard (2003); Hubbard (2003); Baker and Hubbard (2004); Barla et al. (2010)), with a focus on monitoring and on-board computer adoption.

7:00 a.m. every day. An exception is Vreugdenhil (2018), which studies the offshore drilling industry and adds vertical heterogeneity in searching agents' productivities. In contrast, I study a setting where horizontal differentiation - home location - is first order and demonstrate how to incorporate it into this class of models.

My paper also builds on the literature on endogenous trade costs, especially research on the backhaul effect (Behrens and Picard (2011); Ishikawa and Tarui (2018); Hayakawa et al. (2020); Wong (2020)). In this literature, transportation firms are constrained to return to their home after completing any trip - this causes prices on the fronthaul and backhaul legs to be correlated. The most closely related paper to mine is Allen et al. (2020), which studies trucking markets in Colombia with a focus on the effect of remoteness on competition and prices. In their model, truckers choose where to live, choose a single route to service (potentially traveling to the origin and back from the destination), and set markups along each route. The home preference in my paper generalizes the backhaul effect in this literature by allowing the carriers to follow any arbitrary pattern of trips or trip chains. In equilibrium, demand, travel costs, and the magnitude of the home preference jointly determine when and how often carriers return home.

2 Industry and Data

2.1 Industry Background

The trucking freight industry is a significant component of the U.S. economy. In 2019, the truck transportation industry employed over 2 million people, generated $797.7 billion in revenue, and moved 11.84 billion tons of freight - representing 80.4 percent of total freight by value and 72.5 percent of total freight by weight (Adler (2020); Day and Hait (2019)). Compared to alternative modes, trucking freight is less cost-efficient but more flexible than rail or water and slower but cheaper than air. Trucking freight dominates most commodity classes except for very low-value density commodities such as bulk grains, very high-value density commodities such as electronics, and specific categories of hazardous materials. In 2020 and 2021, the Covid pandemic has only reiterated the importance of the trucking sector, as truckers played a major role in sustaining supply chains for vital medical supplies as well as increased e-commerce purchasing. Continuing capacity constraints in trucking, on both labor and capital, have driven freight costs up to 40 percent higher than pre-Covid levels.

In this paper, I focus on a specific segment of the freight industry: the over-the-road, general freight, truckload market. This market has several important implications for my paper. The over-the-road market refers to long distance trips between cities, and is the market segment most susceptible to automation in the near term. The general freight market refers to delivery of standardized van trailers, which reduces the potential degree of driver specialization. In the truckload market, firms buy and sell an entire truck's worth of capacity at a time. This means that unlike other markets such as the less-than-truckload or parcel markets, returns to scale from efficient sorting and routing are diminished. Consistent with this, the truckload market is more characterized by smaller, competitive firms - the Bureau of Labor Statistics reports that 88 percent of long-haul general freight business are owner operators. These features - non-local travel, lack of specialization, and small competitive firms - inform my modeling choices. For more details on market structure, see Appendix D.1.

As a mode of transportation, trucking has been historically resistant to substituting inputs from labor to capital. While freight trains have increased in length, and barges and ships in tonnage, trucks have maintained a high ratio of human operators to freight shipped. The necessity of a human creates an opportunity for drivers' preferences to play an important role, and for automation to disrupt the market. Ocean shipping in the 1800s may provide a historic parallel; Glaeser (2005) describes how larger ships de-emphasized sailing skills and lead to a decline in sailor-rich Boston.

Automation has generated significant industry and policy interest in recent years. Many firms have entered the self-driving truck space, from entrants like TuSimple, Waymo, Embark, and Otto, to incumbents like UPS. Plus, Aurora, Torc, and TuSimple began on-road testing of Level 4 autonomous trucks in 2021.[4] Plus and TuSimple expect to start mass production of Level 4 trucks in 2024. While passenger vehicles must achieve Level 5 automation to cover most use cases, freight trucks could run

[4]SAE defines Level 4 autonomy as "features that can drive the vehicle under limited conditions," where "limited conditions" could include the National Highway System but not local city streets. The U.S. Department of Transportation describes Level 4 systems as having "an [operational design domain] of limited-access highways" and "capable of operating within their ODD with no human operator in the vehicle."

long-haul and regional routes under Level 4 autonomy alone. On the other hand, tasks like loading/unloading at the warehouse and last-mile delivery are likely to remain human-operated.

In this paper, I will focus on the "Exit-to-Exit" model of Level 4 autonomous operation, which is described as the "odds-on favorite for near-time adoption" (Viscelli (2018), Viscelli (2020))[5]. Under this model, a trip begins at a customer warehouse with a human driver. The human driver takes the trailer along local roads to an autonomous truck port near a highway, where the trailer is transferred to an autonomous tractor. The autonomous tractor travels on the highway network until it reaches another truck port near the final destination. A human driver at the destination port then makes the last mile delivery to the customer's destination. This model purely employs local drivers - never taking anyone further than the nearest truck port - and is feasible with Level 4 autonomy. In practice, for example, in 2021 TuSimple contracted with Ryder to use Ryder's existing network of maintenance facilities to build a future truck port network.

In early 2021, the U.S. Department of Transportation released its Automated Vehicles Comprehensive Plan, detailing how it plans to support the integration of automated vehicles into the surface transportation system. The three key goals of this plan include promoting collaboration and transparency, modernizing the regulatory environment, and preparing the transportation system. The U.S. DOT and FMCSA have begun preparing the regulatory system for automation, e.g., eliminating legal assumptions that humans are always present in commercial motor vehicles. Policy discussions have centered on the safety of self-driving trucks and the potential disemployment effects; the International Brotherhood of Teamsters union has lobbied lawmakers to create stricter regulations over driverless vehicles. Meanwhile, enthusiasts point to the potential for heightened efficiency and safer roads without driver fatigue. In this context, my paper points in a different direction: the potential for self-driving trucks to integrate previously geographically segmented markets and shift carriers away from their home locations.

[5]Firms have explored other models, including human-drone platooning (Peloton 2011-2021, Locomation) and drone operation (Starsky 2015-2020). In the past few years, firms pursuing alternate models have largely exited the market.

2.2 Data

I combine several datasets to capture the spot trucking market.

First, I use a dataset of spot market transactions from DAT RateView. DAT Freight & Analytics operates the dominant trucking freight marketplace platform in the U.S. DAT also combines in-house data generated by this platform with external data to produce data analytics products. DAT collects transactions from shippers, carriers, and third-party brokers to construct the RateView dataset. In my version of the dataset, for each origin-destination-week from 2016 to 2020, I observe the number of trips and moments of transaction prices (mean, standard deviation, etc.) for spot and contract transactions. In total, I see 669 million trips over these four years and 107 million trips for 2019 alone. Origins and destinations are observed at the Key Market Area (KMA) level. A KMA is a collection of 3-digit zip code areas roughly comparable in size to an MSA.

Next, I use DAT Trucks in Market data to measure the distribution of active carriers in the spot market. On DAT's trucking freight marketplace platform, shippers (loads) and carriers (equipment) can search for counterparties, make posts or offers, and match with each other. In the Trucks in Market dataset, for each origin-day from 2016 to 2020, I observe the number of load and equipment searches and posts on the platform. In total, I see 458 million carrier searches over the four years and 50 million carrier searches for 2019 alone.

Thus far, the DAT datasets aggregate across carriers from different home locations. To study carrier heterogeneity, I turn to motor carrier highway inspections. Under the Federal Motor Carrier Safety Administration's (FMCSA) Motor Carrier Safety Assistance Program, inspectors conduct roadside inspections of motor carriers. These inspections ensure that drivers and vehicles are safely operating in compliance with FMCSA regulations and vary in intensity from driver-only inspections to the most thorough inspections which evaluate brake systems, suspensions, etc. For each inspection, I observe the date, the state of inspection, the DOT number of the inspected vehicle, and registration state of the tractor and trailer (if present). I use the tractor's registration state as the home location as, unlike trailers, it is difficult to register a tractor outside one's home state.[6] The inspections dataset spans 3.2 million

[6]Alternatively, I could link each inspection report with the inspected carrier's registration data in

inspections and 1.8 million unique tractors over 2019.[7]

To measure the total size of potential trucking freight demand, I use 2017 Census Bureau's Commodity Flow Survey (CFS). The Commodity Flow Survey samples shippers to estimate state-state shipping for each major freight modes: truck, rail, water, air, pipeline, etc. I also use the FMCSA's Census of Motor Carriers to measure the total population of potential carriers.

I use several data sources to construct route characteristics within the United States. I begin by using Open Street Maps to compute the fastest driving route between any two locations. Next, I use the Highway Performance Monitoring System, an annual census of U.S. highways, to measure physical road characteristics at the road segment level, and I use state average diesel prices from AAA. I take a weighted average of these characteristics over the fastest Open Street Maps route to construct route characteristics, where the weights are the mileage share elapsed in a given state. This construction represents the average road conditions experienced by drivers between origin and destination. For example, a carrier driving from Texas to Florida will spend a small share of the journey in Alabama. My method assigns a correspondingly small weight to Alabama diesel prices when constructing route-average diesel prices.

To construct a cost shifter, I construct the percentage of days in a year that a state experiences snow. I use a weather dataset constructed by Moosavi et al. (2019) which collects reports from airport weather stations. This paper shows that airport reports of snow positively predicts deterioration in traffic, which would raise the cost for a tracker to travel along a route. For each state, I compute the average share of days in the year that snow is reported, and for each route, I weight the average state snowiness by the mileage share to construct the route-level cost shock.

To construct a demand shifter, I construct the availability of river or water shipping from origin to destination using the Army Corps of Engineers' Fuel-Taxed Inland

the FMCSA's Census of Motor Carriers. However, the FMCSA Census has known data challenges, and multiple drivers/tractors may work for a given firm. For more details, see Burks et al. (2017).

[7] For more details on this dataset, see Liang (2021). She finds that states vary systematically in how intensely and frequently they inspect - I will incorporate this explictly in my estimation section. She also argues that, conditional on truck characteristics, there are no pre-trends between trucks which are inspected and trucks which are not.

Waterway system.[8] For an origin-destination pair, I define an indicator variable which is 1 if a river flows from the origin to destination, or if an intracoastal waterway connects the origin and destination, and 0 otherwise.

2.3 Descriptive Statistics

I begin by documenting three empirical patterns about the trucking freight industry which will motivate my economic analysis. First, I show that carrier homes are unevenly distributed across states. Some states, such as California, have a much larger population of registered carriers than other states, such as Massachusetts. Second, I show that the number of carriers living in the origin or destination states of a given rotue predicts prices and quantities. A larger number of carriers living in the origin predicts higher prices and higher quantity, while the number of carriers living in the destination predicts lower prices and higher quantity. Third, I focus on the role of where carriers drive by defining a home share variable, which captures the probability that a given route brings carriers in the origin location back to their home location. I find that this home share variable predicts lower prices and higher quantities. Finally, I run descriptive quantity and pricing regressions which summarize these patterns. The combination of these three patterns motivate my model of trucking freight with geographic specialization. For additional descriptive figures, see Appendix B.

In Figure 1, I document the unequal distribution of carriers across the United States. To select for firms actively engaged in inter-state freight, I combine carrier registrations with highway inspections to measure the population of active carriers. I define an active carrier as any U.S.-licensed tractor inspected in the U.S. at least once in 2019. I focus on tractors because tractors are more difficult to register outside one's state of residency than trailers. Compared to the population, the distribution of active carriers is much more concentrated in a few states: California, Texas, Illinois, and Florida.

[8]This system represents the main internal water shipping in the United States, and covers the navigable components of major river systems such as the Mississippi, the Missouri, the Ohio, the Arkansas, the Columbia, and others, major artificial waterways such as the Tennessee-Tombigbee Waterway, and the Gulf and Atlantic Intracoastal Waterways.

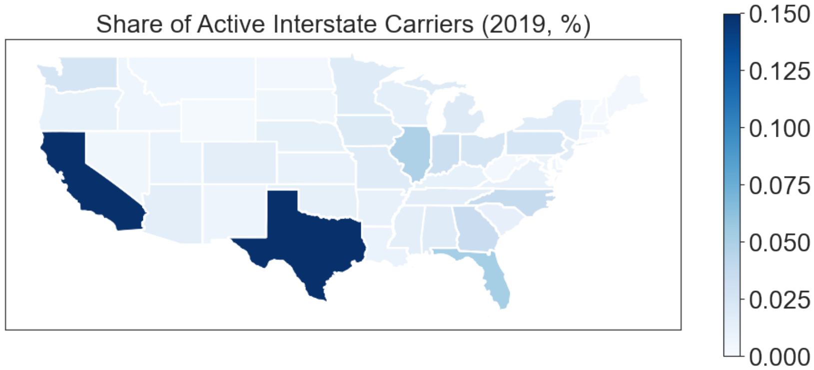

Figure 1: Geographic Distribution of Trucking Carriers

Notes: This figure plots the share of all active interstate carriers by U.S. state. Active interstate carriers are defined as carriers which (1) were inspected at least once, and (2) were registered for interstate, non-hazmat, non-passenger carriage.
Data sources: 2019 FMCSA Motor Carrier Inspections, 2019 FCMSA Registrations.

To explore the implications of this carrier distribution on the trucking market, I plot bin scatters of route-level prices and quantities on the number of carriers from the origin location in Figure 2. Routes where the origin state has a larger number of carriers tend to have higher prices and higher quantities. This is consistent if, for example, carriers are more likely to live in places which demand more trucking exports.

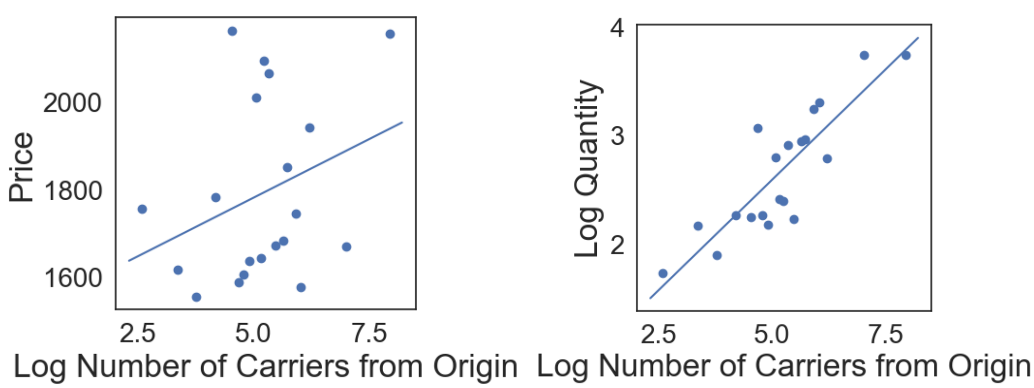

Figure 2: Effect of Number of Carriers from Origin

Notes: This figure plots bin scatters of price (Panel A) and quantity (Panel B) for a given route $i \to j$ against the number of carriers from the origin location i.
Data sources: DAT, 2019 FCMSA Registrations.

On the other side, in Figure 3 I plot bin scatters of route-level prices and quantities on the number of carriers from the destination location. Routes where the destination state has a larger number of carriers tend to have lower prices and higher quantities.

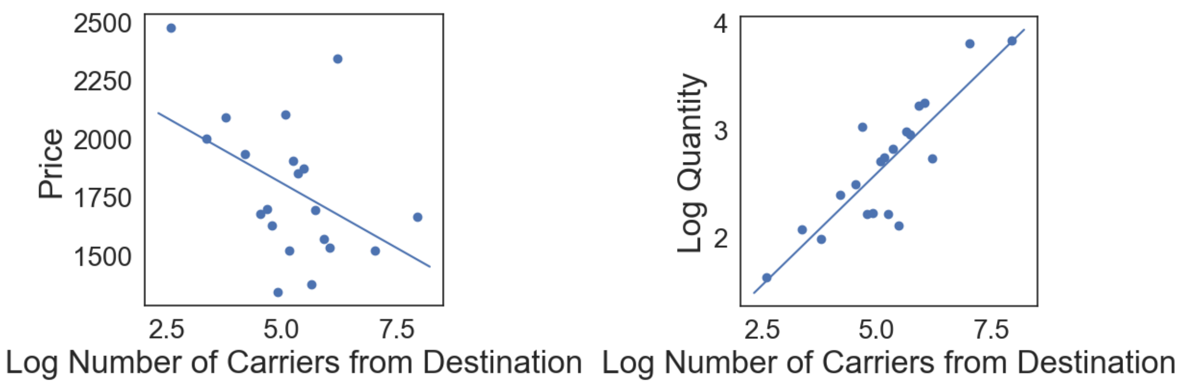

Figure 3: Effect of Number of Carriers from Destination

Notes: This figure plots bin scatters of price (Panel A) and quantity (Panel B) for a given route $i \to j$ against the number of carriers from the destination location j.
Data sources: DAT, 2019 FCMSA Registrations.

The previous figures did not use any information on where carriers are actually located when they are driving. To isolate the effect of carriers' location decisions, I use my highway inspections data to construct a home share variable h_i^j: the share of the carriers in the origin location who make their homes in the destination location. For example, for California to Massachusetts shipments, about 0.037 percent of carriers in California are from Massachusetts. Intuitively, this variable measures the share of available carriers for whom this trip would be a homecoming trip. If preferences to return home are important for prices and quantities, this home share variable should have predictive value.

In Figure 4, I plot the relationship between prices, quantity shipped, and the distribution of carriers. Panel A presents a bin scatter of price against the share variable in blue. Along lanes with larger home shares, prices are lower. Since carriers are more likely to work in states which are close to their home state, the share variable is negatively correlated with distance. To control the effect of distance, I plot a second bin scatter in orange which absorbs route distance. The slope is flatter but the overall pattern remains negative. Panel B presents a bin scatter of log quantity against the share variable in blue. Lanes with a higher share of carriers in the origin being from the destination location see higher quantity shipped. This relationship is robust to controls for distance, as shown by a second bin scatter in orange which controls for route distance.

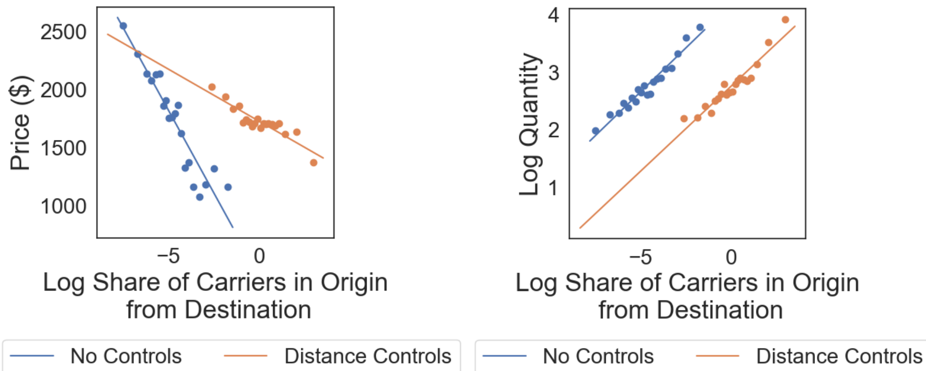

Figure 4: More Carriers in origin from destination are correlated with lower prices and higher quantities

Notes: This figure plots bin scatters of price (Panel A) and quantity (Panel B) for a given route $i \to j$ against the share of carriers in origin i that are from destination location j. I refer to this share as the home share.
Data sources: DAT, 2019 FCMSA Registrations.

2.3.1 Descriptive Regressions

I synthesize the previous patterns using a pair of descriptive regressions predicting trucking flows and prices.

In Table 1, I present the results of a regression of (log) trucking quantity on distance, the home share variable described in the previous section, and the number of carriers from the origin and destination states. Model 1 estimates the following equation,

$$\log Q_{ij} = \beta^d d_{ij} + \beta^h \log h_i^j + \beta^{orig} \log C_i + \beta^{dest} \log C_j$$

where h_i^j is the share of carriers in i who are from j, and C_i is the total number of carriers from i. Routes which are shorter, have a higher home share, and connect locations with more carriers, tend to have greater quantities.

In Model 2, I estimate the same equation with origin and destination fixed effects.

Model	1	2
Distance	-0.353 (0.0244)	-0.226 (0.0128)
Log Home Share	1.514 (0.505)	1.345 (0.255)
Log Carriers from Orig	0.415 (0.0138)	
Log Carriers from Dest	0.393 (0.0171)	
Constant	-0.132 (0.111)	3.764 (0.0609)
Fixed Effects	No	Yes
Observations	1936	1936
R-squared	0.522	0.929

Table 1: Descriptive Trucking Flow Regression

Notes: This table presents results from a regression of log trucking quantities on distance, home share, and carrier population. Robust standard errors in parentheses.
Data sources: DAT, 2019 FCMSA Registrations.

The fixed effects absorb the effect of origin and destination carriers.

$$\log Q_{ij} = FE_i + FE_j + \beta^d d_{ij} + \beta^h h_i^j$$

I continue find a negative relationship between distance and trucking flows, indicating a gravity-like relationship in this setting. Further, I confirm the patterns previously presented: quantity is rising in the share of potential homecoming carriers, and in the total number of carriers from the origin and destination locations.

I estimate a parallel pricing equation and present the results in Table 2. In Model 1, I run a regresion of prices on the same set of variables: distance, home share, and number of carriers from origin and destination.

$$p_{ij} = \beta^d d_{ij} + \beta^h \log h_i^j + \beta^{orig} \log C_i + \beta^{dest} \log C_j$$

In Model 2, I add origin and destination fixed effects, which absorb the effect of origin and destination carriers.

$$p_{ij} = FE_i + FE_j + \beta^d d_{ij} + \beta^h h_i^j$$

Model	1	2
Distance	1.190 (0.0106)	1.162 (0.00507)
Log Home Share	0.271 (0.219)	-0.391 (0.101)
Log Carriers from Orig	0.0495 (0.00598)	
Log Carriers from Dest	-0.123 (0.00744)	
Constant	0.730 (0.0481)	0.302 (0.0241)
Fixed Effects	No	Yes
Observations	1936	1936
R-squared	0.880	0.985

Table 2: Descriptive Price Regression

Notes: This table presents results from a regression of prices on distance, home share, and carrier population. Robust standard errors in parentheses.
Data sources: DAT, 2019 FCMSA Registrations.

I find that longer trips are associated with higher prices. This descriptive regression suggests a price-distance gradient of around $1.20 per mile. In Model 1, the share of potential homecoming carriers is positive and noisy, while in Model 2, the coefficient becomes negative. Finally, as in the previous figures, more carriers in the origin are associated with higher prices, and more carriers in the destination are associated with lower prices. This is consistent with a story where carriers must be compensated to leave home, and are willing to accept lower prices to return home.

In practice, both prices and the distribution of working carriers are determined in equilibrium. For example, the home share variable may be driven by carriers responding to underlying demand patterns. Similarly, prices both reflect the supply of available carriers, and the demand for shipping along each route. To disentangle these forces, I turn to my model in the following section.

3 Model

In this section, I develop a dynamic spatial equilibrium model to explore the effects of geographic specialization. For intuition, I begin by presenting a two-location version of the model. This simplified setting allows me to isolate the main economic forces in the model. After, I extend this to a many-location setting which I will take to

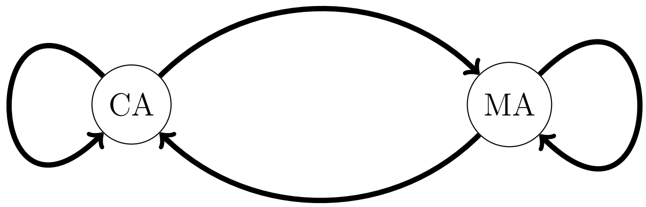

Figure 5: Two Location Model: Markets

estimation.

3.1 Two-Location Model

I begin by presenting a two-location model which captures the main economic forces in this market. There are two locations, CA and MA. Every day, there is symmetric demand for shipping between the two locations, and within each location,

$$D_{ij}(p_{ij}) = \sigma p_{ij} + \omega \tag{1}$$

for all $i, j \in \{CA, MA\}$, where p_{ij} is the price of shipping from i to j. Figure 5 presents the geography in this model, along with the four shipping routes.

There are a set of carriers in the economy. Each of them has a home location: CA or MA. At the beginning of every day, a carrier is either in a location (CA or MA) - available to begin a new trip - or en-route between a pair of locations ($CA \to MA$, $MA \to CA$, $CA \to CA$, or $MA \to MA$). There are therefore six total states that a carrier can be in.

Consider a CA carrier who begins the day in CA. Since she began the period at home, she receives a home preference flow utility b. Next, she has three choices, as summarized in the first panel of Figure 6: Taking an interstate job to MA, taking an intrastate job, or taking the day off and exercising an outside option. If she takes an interstate job to destination MA, she receives payoff

$$p_{CA,MA} - c + V^{CA}_{CA \to MA} + \epsilon_{CA,MA}$$

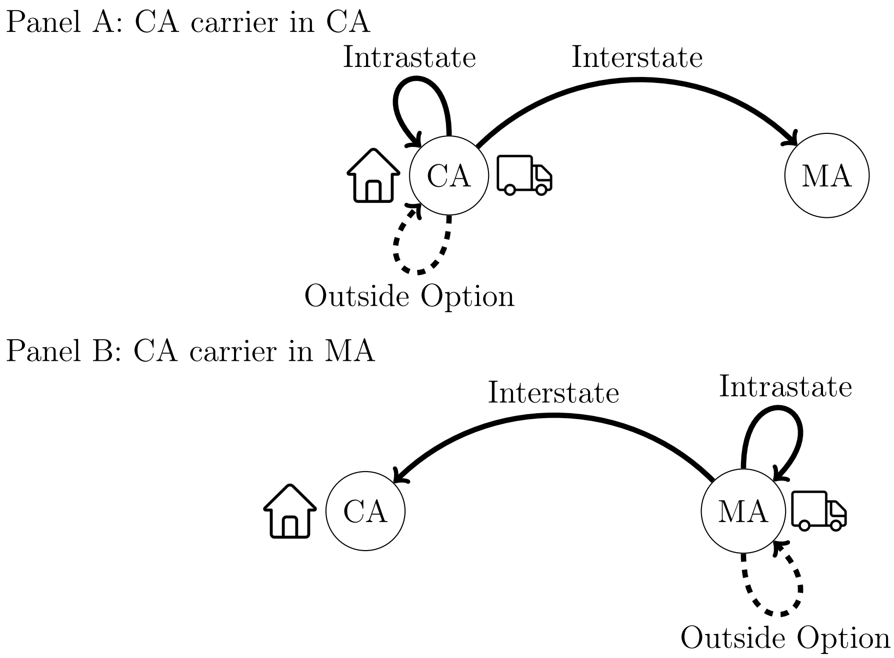

Figure 6: Two Location Model: Choices

where $p_{CA,MA}$ is the price of a trip from CA to MA, c is a travel cost, and $\epsilon_{CA,MA}$ is an i.i.d. Logit shock. The carrier takes prices as given. Next period, she will begin travelling to MA, and receive continuation value $V^{CA}_{CA\to MA}$ for being a CA carrier travelling from CA to MA. If she takes an intrastate job, she instead receives payoff

$$p_{CA,CA} - c + V^{CA}_{CA\to CA} + \epsilon_{CA,CA}$$

where $p_{CA,CA}$ is the price of a CA intrastate trip. Similarly, next period she will begin travelling to her destination, and her continuation value is $V^{CA}_{CA\to CA}$. Finally, she can take an outside option, choose not to work, and receive payoff

$$\beta V^{CA}_{CA} + \epsilon_{OO}$$

where ϵ_{OO} is i.i.d. Logit. Since she is available to work the following day, her continuation utility if she takes the outside option is V^{CA}_{CA}, discounted by her discount factor β.

Putting these together, the payoff of a carrier who starts in CA is

$$V_{CA}^{CA}(\epsilon) = \underbrace{b}_{\text{Home Preference}} + \max \begin{cases} p_{CA,MA} - c + V_{CA\to MA}^{CA} + \epsilon_{CA,MA} & \text{Interstate to MA} \\ p_{CA,CA} + V_{CA\to CA}^{CA} + \epsilon_{CA,CA} & \text{Intrastate within CA} \\ \beta V_{CA}^{CA} + \epsilon_{OO} & \text{Outside Option} \end{cases}$$

If she instead started the day in MA, she faces the same three choices, as depicted in the second panel of Figure 6. Her payoffs are symmetric, with the exception of the home preference utility,

$$V_{MA}^{CA}(\epsilon) = \max \begin{cases} p_{MA,MA} - c + V_{MA\to CA}^{CA} + \epsilon_{CA,MA} & \text{Interstate to CA} \\ p_{MA,MA} + V_{MA\to MA}^{CA} + \epsilon_{CA,CA} & \text{Intrastate within CA} \\ \beta V_{MA}^{CA} + \epsilon_{OO} & \text{Outside Option} \end{cases}$$

Note that she is free to choose to stay in MA for as long as she desires. There is no restriction on the pattern or order of trips, or a maximum number of days after which she must return to CA.

When travelling, It takes two days to travel between CA and MA, and one day to perform a within-state trip. This implies that the continuation value of beginning a trip is

$$V_{CA\to MA}^{CA} = \beta^2 V_{MA}^{CA}$$
$$V_{CA\to MA}^{CA} = \beta V_{CA}^{CA}$$

The shape of the logit shocks implies choice probabilities for this CA carrier in CA,

$$s_{CA\to MA}^{CA}(p) = \frac{\exp(p_{CA,MA} - c + V_{CA\to MA}^{CA})}{\exp(p_{CA,MA} - c + \beta^2 V_{MA}^{CA}) + \exp(p_{CA,CA} + \beta V_{CA}^{CA}) + \exp(\beta V_{CA}^{CA})} \quad (2)$$

$$s_{CA\to CA}^{CA}(p) = \frac{\exp(p_{CA,CA} + \beta V_{CA}^{CA})}{\exp(p_{CA,MA} - c + \beta^2 V_{MA}^{CA}) + \exp(p_{CA,CA} + \beta V_{CA}^{CA}) + \exp(\beta V_{CA}^{CA})} \quad (3)$$

as well as the expected value of beginning the period in CA,

$$V_{CA}^{CA}(p) = \log(\exp(p_{CA,MA} - c + \beta^2 V_{MA}^{CA}) + \exp(p_{CA,CA} + \beta V_{CA}^{CA}) + \exp(\beta V_{CA}^{CA})) \quad (4)$$

Given a set of prices p, the value function of the CA carrier can therefore be solved for when she is in CA and when she is in MA. Finally, the above equations hold symmetrically for carriers from MA.

There is a fixed population of carriers, C^{CA} and C^{MA}. Among the California carriers, there are C_{CA}^{CA} carriers in CA, and C_{MA}^{CA} carriers in MA. Since some carriers will be travelling, $C_{CA}^{CA} + C_{MA}^{CA} \neq C^{CA}$. Given a set of carrier locations and prices, total quantity supplied from CA to MA is a weighted average of the probability that a CA carrier chooses to go to MA, and the probability that an MA carrier chooses to go to MA,

$$Q_{CA \to MA} = C_{CA}^{MA} sh_{CA \to MA}^{MA} + C_{CA}^{CA} sh_{CA \to MA}^{CA}$$

where the relative share of CA or MA carriers in CA provides the weights.

I will focus on a steady-state equilibrium where the number of carriers of each type in each location are constant. For example, every day $C_{MA}^{CA} s_{MA \to CA}^{CA}$ carriers from CA arrive in MA, and $C_{CA}^{CA} s_{CA \to MA}^{CA}$ carriers leave. In steady state,

$$C_{MA}^{CA} sh_{MA \to CA}^{CA} = C_{CA}^{CA} sh_{CA \to MA}^{CA} \quad (5)$$

Finally, prices are set to clear markets,

$$D_{ij}(p_{ij}) = Q_{i \to j} \quad (6)$$

A steady state equilibrium of this two-state model is a vector of four prices $(p_{CA,CA}, p_{CA,MA}, p_{MA,CA}, p_{MA,MA})$, four quantities $(q_{CA,CA}, q_{CA,MA}, q_{MA,CA}, q_{MA,MA})$, eight choice probabilities $(s_{CA,CA}^{CA}, s_{CA,MA}^{CA}, s_{MA,CA}^{CA}, s_{MA,MA}^{CA}, s_{CA,CA}^{MA}, s_{CA,MA}^{MA}, s_{MA,CA}^{MA}, s_{MA,MA}^{MA})$, four value functions $(V_{CA}^{CA}, V_{MA}^{CA}, V_{CA}^{MA}, V_{MA}^{MA})$, and four equilibrium carrier locations $(C_{CA}^{CA}, C_{MA}^{CA}, C_{CA}^{MA}, C_{MA}^{MA})$, which satisfies four demand equations (1), four market clearing conditions (6), eight choice probability equations (2 and 3), four value function equations (4), and four steady-state equations (5).

Given this model, I run a set of simulations to illustrate how the model parameters, especially the magnitude of the home preference b, affect equilibrium outcomes. I begin with a base setting where the number of carriers in CA and MA are symmetric[9]. I compute a series of equilibria for different values of home preference b, and plot the results in Figure 7. Panel A plots the share of carriers in each location which are either at home (i.e., from that location) or away (i.e., from the other location). As home preference increases, the carriers in each location are more likely to be at home at home. While this does not change the total supply of carriers in each location, since the two locations have symmetric populations of carriers, carriers become more hesistant to travel to the other location. This raises inter-state prices and lowers intra-state prices, as seen in Panel B. Given constant demand, inter-state quantity shipped falls and intra-strate quantity rises as in Panel C. An increase in the degree of home preference generates a steeper negative relationship between quantity and distance.

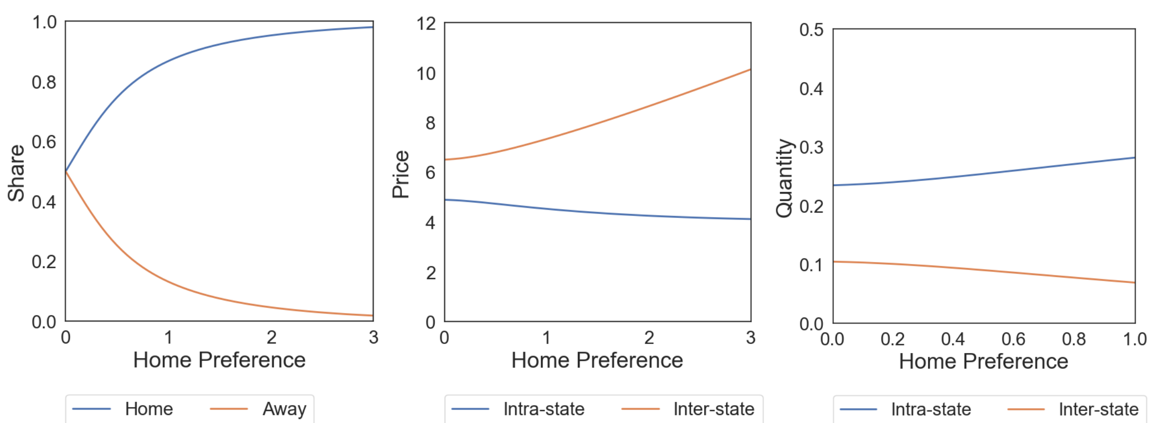

Figure 7: Two Location Setting, Symmetric Locations

Note: Figure 7 plots the results of simulations of the two-location model, as the home preference parameter varies. Panel A plots the share of carriers in each location which are home (from that location) or away (from the other location). Panel B plots the prices for intra-state and inter-state trips. Panel C plots the trade flows of intra-state and inter-state pairs.

Next, suppose that the two locations are asymmetric in their endowments of carriers. Specifically, suppose that one location (CA) has 75 percent of the carrier

[9]Specifically, I let $C^{CA} = C^{MA} = \frac{1}{2}$. For the other parameters, I set $\omega_{ij} = 1$ for all (i, j), $\sigma = -\frac{1}{2}$, $c = 1$, $\alpha = 1$.

population, and the other location (MA) has the remaining 25 percent, while holding all other parameters fixed. I compute a new set of equilibria for varying levels of home preference.

Panel A plots the share of carriers in each location which are from each location. At zero home preference, the share of carriers in each location matches the overall distribution of home locations - 75 percent of carriers in both CA and MA are from CA. As home preference increases, CA truckers make up more of the population in CA, and likewise for MA truckers. On net, this leads to fewer carriers in MA and more carriers in CA. The implications can be seen in Panel B, where intra-state prices now diverge between MA and CA. On average, inter-state prices rise relative to intra-state prices. For MA, this works in the opposite direction as the overall number of carriers, so MA sees a modest rise in intra-state prices. A higher home preference leads to a steeper price-distance gradient. This carries into quantities, as plotted in Panel C, where a higher home preference leads to more intra-state quantity for the carrier-rich location, modest changes in quantity for the carrier-poor location, and a decrease in inter-state quantity.

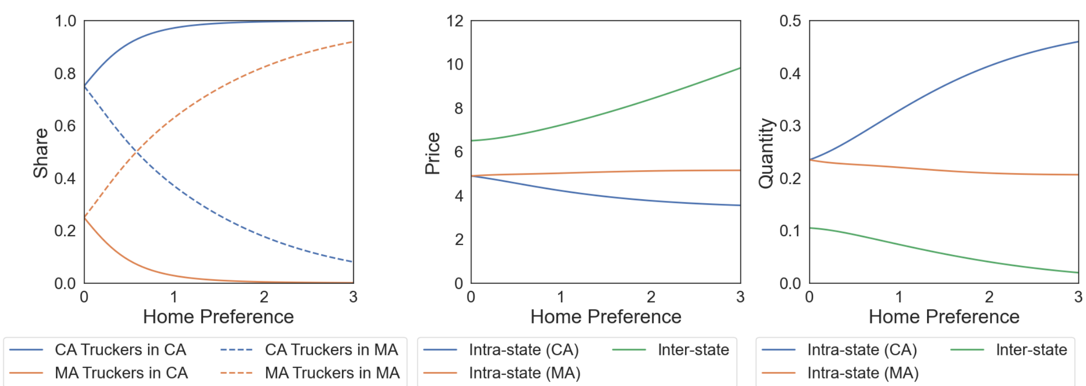

Figure 8: Two Location Setting, Asymmetric Locations

Note: Figure 8 plots the results of simulations of the two-location model, as the home preference parameter varies. Panel A plots the share of carriers in each location which are home (from that location) or away (from the other location). Panel B plots the prices for intra-state and inter-state trips. Panel C plots the trade flows of intra-state and inter-state pairs.

3.2 General Model

3.2.1 Setup, Agents, and Timing

There is a set of locations in the economy, L. There are two types of infinitesimal agents: shippers who demand shipping and carriers who supply shipping. Shippers are short-lived, myopic, and differentiated by their origin-destination. Carriers are risk-neutral, long-lived, and differentiated by their home location h.

Each carrier may either begin the period in a location $i \in L$, or begin the period en-route from i to j for $i, j \in L$. There is a fixed total quantity of carriers of each type h, C^h, and the number of carriers in each location i, C_i^h, or en-route along each route ij, C_{ij}^h, must sum up, so that

$$C^h = \sum_{i \in L} C_i^h + \sum_{i,j \in L} C_{ij}^h \quad \forall h \in L \tag{7}$$

In contrast, I refer to the number of carriers of each type present in location i

$$C_i = \sum_h C_i^h$$

as the supply of carriers in location i. Importantly, for example, the number of California carriers, C^{CA}, is not equal to the supply of carriers in California, C_{CA}, as carriers of all types move around the economy.

Each period is one day. First, shippers and carriers draw idiosyncratic cost and demand shocks. Second, prices are set to clear shipper demand & carrier supply. Third, matched carriers begin travelling. Finally, travelling carriers arrive at their destination with some probability.

3.2.2 Carrier Choices / Freight Supply

Each period, a carrier c of type h in location i can choose to either accept a job to any destination $j \in L$, or choose to take an outside option and remain in i.

A carrier c who accepts a job from i to j receives flow utility

$$u_{ijc} = \alpha p_{ij} + \kappa_{ijc}$$

where p_{ij} is the route price and κ_{ij} is the route cost[10]. Carriers are price-takers and observe $p = (p_{ij})_{i,j \in L}$, a vector of prices along all routes.

Route costs are decomposed as

$$\kappa_{ijc} = \bar{\kappa}_{ij} + \xi_{ij} + \epsilon_{cj} \tag{8}$$

where $\bar{\kappa}_{ij}$ is an observable component, ξ_{ij} is a systematic unobservable cost component, and ϵ_{cj} is an idiosyncratic cost shock which is i.i.d. Logit across all carriers. For example, distance may be an observable component of $\bar{\kappa}_{ij}$, while unobserved congestion or scarce public bathrooms may be unobserved components of ξ_{ij}.

A carrier who accepts a job to j arrives at the destination with probability λ_{ij}. If they arrive, they may immediately take another job the following period. With probability $1 - \lambda_{ij}$, the carrier does not arrive and instead begins the next period en-route from i to j. Each day, an en-route carrier arrives with probability λ_{ij}. The expected days of travel for a job from i to j is therefore $1/\lambda_{ij}$.

A carrier c who takes the outside option in location i receives flow utility

$$u_{i,OO} = \delta_{i,OO} + \epsilon_{c,OO} \tag{9}$$

where $\delta_{i,OO}$ is a common flow payoff and $\epsilon_{c,OO}$ is an i.i.d. Logit cost shock. Let $\epsilon_c = \{\epsilon_{cj}\}_j \cup \epsilon_{c,OO}$ be the entire vector of cost shocks received by carrier c.

Finally, a carrier which begins the day in its home location h_c receives a home preference flow payoff, b. Economically, b can represent both positive preferences for being at home as well as costs associated with being away at home.

Putting these together, the carrier's daily flow payoffs are

$$b \times 1_{h_c=i} + \begin{cases} u_{ijc} & \text{Accept job to } j \\ u_{ikc} & \text{Accept job to } k \\ \cdots & \\ u_{i,OO} & \text{Outside Option} \end{cases} = b \times 1_{h_c=i} + \begin{cases} \alpha p_{ij} + \bar{\kappa}_{ij} + \xi_{ij} + \epsilon_{cj} \\ \alpha p_{ik} + \bar{\kappa}_{ik} + \xi_{ik} + \epsilon_{ck} \\ \cdots \\ \delta_{i,OO} + \epsilon_{c,OO} \end{cases} \tag{10}$$

[10]The carrier pays the entire cost of a trip at the beginning of the trip. Equivalently, κ_{ij} can be interpreted as the expected discounted value of flow costs that the carrier pays during the trip.

For convenience, let δ_{ij} denote the component of flow payoffs which is common across carriers,

$$\delta_{ij} = \alpha p_{ij} + \bar{\kappa}_{ij} + \xi_{ij} \tag{11}$$

To complete the carrier optimization problem, let V_{ic}^h be the expected value function of a carrier with home h which starts a day available in i. Let W_{ijc}^h be the expected value function of a carrier with home h which starts a day en-route from i to j. Carriers have a discount factor of β. The carrier's choice problem is, given idiosyncratic shocks ϵ_{cj},

$$\max \begin{cases} \delta_{ij} + \epsilon_{cj} + \beta E[W_{ij}^h] & \text{Accept job to } j \\ \delta_{ik} + \epsilon_{ck} + \beta E[W_{ik}^h] & \text{Accept job to } k \\ \dots & \\ \delta_{i,OO} + \epsilon_{c,OO} + \beta E_{\epsilon_c}[V_{ic}^h(\epsilon_c)] & \text{Outside Option} \end{cases} \tag{12}$$

and the carrier's value function is defined by

$$V_{ic}^h(\epsilon_c) = b \times 1_{h_c = i} + \max \begin{cases} \delta_{ij} + \epsilon_{cj} + \beta E[W_{ij}^h] & \text{Accept job to } j \\ \delta_{ik} + \epsilon_{ck} + \beta E[W_{ik}^h] & \text{Accept job to } k \\ \dots & \\ \delta_{i,OO} + \epsilon_{c,OO} + \beta E_{\epsilon_c}[V_{ic}^h(\epsilon_c)] & \text{Outside Option} \end{cases} \tag{13}$$

Given the stochastic arrival probabilities, the value function in the en-route state is

$$E[W_{ij}^h] = \lambda_{ij} E_\epsilon[V_j^h] + (1 - \lambda_{ij})\beta E[W_{ij}^h] \tag{14}$$

where $E[W_{ij}^h]$ is a weighted average of the value function in the destination and the value function of being in the en-route state. Since there are no idiosyncratic shocks in the en-route state, I do not index W by c.

Given the logit cost shocks, the share of carriers of type h in origin i who choose destination j is given by

$$s_{ij}^h = \frac{\exp(\delta_{ij} + E[W_{ij}^h])}{\sum_{k \in L} \exp(\delta_{ik} + E[W_{ik}^h]) + \exp(\delta_{i,OO} + \beta E_\epsilon[V_{ic}^h])} \tag{15}$$

I call this the type-specific choice probability, as it is specific to carriers of type h.

The aggregate share of carriers who choose to take jobs from i to j

$$s_{ij} = \sum_h \left(\frac{C_i^h}{\sum_{h'} C_i^{h'}} \right) s_{ij}^h \tag{16}$$

is a weighted average of type-specific choice probabilities, where the weights are the share of carriers in i which are from h. Intuitively, a larger share of shipping flows from origin i to destination j if origin i has a large share of carriers who like destination j (high $\frac{C_i^h}{\sum_{h'} C_i^{h'}}$ for h with high s_{ij}^h). For example, suppose a large share of carriers working in Washington have homes in California. These carriers' preferences (e.g., to return to California) will have an outsized weight on aggregate exports from Washington.

Given the type-specific choice probabilities, aggregate supply

$$S_{ij} = \sum_h C_i^h s_{ij}^h \tag{17}$$

is the sum of choice probabilities weighted by the distribution of available carriers.

3.2.3 Shipper Optimization / Freight Demand

Every day, a mass of N_{ij} potential shippers enter with fixed origin $i \in L$ and destination $j \in L$. Each shipper draws a willingness-to-pay for freight $\omega_{ij} + \nu$ where ω_{ij} is a common demand shock, and v is an idiosyncratic shock distributed exponential with mean σ^{-1}.

Each shipper makes a binary decision over whether to pay price p_{ij} and ship via truck, or to take an outside option with payoff normalized to 0. The outside option can include alternative freight modes such as rail or barge. If the shipper does not ship, it exits at the end of the day. The shipper's choice problem is

$$\max\{\omega_{ij} + v - p_{ij}, 0\} \tag{18}$$

Aggregating over the mass of potential shippers, demand for shipping from i to j given prices p_{ij} is

$$D_{ij} = N_{ij} \times P(\omega_{ij} + v > p_{ij}) = N_{ij} \exp(-\sigma p_{ij} + \sigma \omega_{ij}) \tag{19}$$

and the share of shippers who choose trucking is

$$s_{ij}^{shipper} = \frac{D_{ij}}{N_{ij}} = \exp(-\sigma p_{ij} + \sigma \omega_{ij})$$

Since shippers have a fixed origin and destination, there is no substitution across destinations and the price of shipping along other routes does not enter demand for shipping from i to j. After taking logs, the log share of shippers who choose trucking

$$\log s_{ij}^{shipper} = \log D_{ij} - \log N_{ij} = -\sigma p_{ij} + \sigma \omega_{ij} \tag{20}$$

is linear in prices.

3.2.4 Equilibrium

This model, the number of carriers of type h in location i is constant period to period if

$$\sum_k \lambda_{ki} (\underbrace{C_{ki}^h}_{\text{inflows from en-route carriers}} + \underbrace{C_k^h s_{ki}^h}_{\text{inflows from new trips}}) = C_i^h \underbrace{\sum_j s_{ij}^h}_{\text{outflows}} \tag{21}$$

holds, where the left-hand-side captures all carriers of type h which arrive from all potential origin locations k, and the right-hand-side captures all carriers of type h who choose to leave i.

The number of carriers of type h in an en-route state from i to j is constant period to period if

$$\underbrace{C_i^h s_{ij}^h (1 - \lambda_{ij})}_{\text{new trips beginning to travel}} = \underbrace{\lambda_{ij} C_{ij}^h}_{\text{en-route carriers arriving}} \tag{22}$$

holds, where the left-hand-side captures all carriers of type h which began trips from i to j this period, and the right-hand-side captures the share of travelling carriers from i to j which arrive (and thereby leave the travelling state).

A steady-state equilibrium of this model is a set of prices p_{ij}, quantities Q_{ij}, and carrier locations C_i^h, C_{ij}^h , such that (1) markets clear, (2) the geographic distribution of carriers is constant over time, (3) carriers make optimal choices, and (4) shippers make optimal choices. Specifically, the following must hold:

- For each $i, j \in L$, total supply equals total demand

$$Q_{ij} = S_{ij}(p) = D_{ij}(p) \tag{23}$$

- For each available state i and en-route state $i \to j$, for each carrier type h, inflows equal outflows (Equations 21 and 22)
- For each $h, i \in L$, carrier market shares are optimal (Equation 15) and value functions are consistent (Equations 13 and 14)
- For each $i, j \in L$, shipper decisions are optimal (Equation 20)

3.3 Discussion

I make several simplifying assumptions to tailor my model for the long-haul truckload market.

First, I assume that all shipments are of equal size. This is true of the truckload market where a shipper purchases an entire truck's worth of capacity. Unlike less-than-truckload (LTL) and parcel markets, there are no complementarities across shipments along a route. For example, trucks have no reason to take a detour to drop off a portion of their load before continuing to their final destination.

Second, I assume that all carriers are optimizing independently. My model therefore rules out large carriers who may be able to exert market power. The existing literature characterizes the truckload market as more competitive than the less-than-truckload market (Corsi et al. (1992); Keeler (1989); Savage (1995); Laing and Nolan (2009)). In the less-than-truckload market, there are strong scale economies due to large firms' ability to operate hubs and repack freight efficiently. At the extreme end, the parcel market is dominated by just a few major firms.

Third, I abstract away from considerations of the spot, contract, and private fleet markets. This model focuses on the cross-sectional variation in freight prices, rather than the time-series interaction of spot and contract prices, or the contracting problem of whether to develop a private fleet. In practice, the spot and contract prices are highly correlated.

Fourth, I assume that there is a single market-clearing price for each origin-destination route. This prevents carriers of different types from charging different prices from shippers for a given route. Since I focus on the general freight market, there is less differentiation that would be true if I were considering a broader market including refrigerated or hazardous materials carriers.

4 Estimation

I begin by making parametric assumptions. I estimate the demand side of my model using a linear IV. I construct an estimator for my supply-side parameters. Finally, I discuss my estimated parameters and interpret their magnitudes.

4.1 Parametric Assumptions

For the remainder of this section, I assume that the observable route cost κ_{ij} is linear in observable route characteristics X_{ij}.

$$\kappa_{ij} = \gamma X_{ij} \tag{24}$$

The common component of the carrier flow payoff δ_{ij} is linear in prices, route characteristics, origin and destination fixed effects, and an unobservable route cost shock.

$$\delta_{ij} = \alpha p_{ij} + \gamma X_{ij} + FE_i + FE_j + \xi_{ij} \tag{25}$$

In this specification, destination fixed effects capture geoegraphic variation in the attractiveness of a destination. When FE_j is high, the flow payoff to travelling to j rises for carriers in all origin locations. Origin fixed effects capture geographic variation in the attractiveness of the outside option. When FE_i is high, the payoff to taking the outside option falls relative to the payoff of accept a trip, so the outside share falls. Since FE_i enters symmetrically for trips from i to all potential destinations, it does not affect the relative choice probabilities when leaving i.

Parameter	OLS	IV
Price ($)	-0.197 (0.0248)	-0.323 (0.101)
Constant	-0.184 (0.0411)	-0.0160 (0.129)
Fixed Effects	Yes	Yes

Table 3: Demand Estimation Results

Notes: Regression of log shipper trucking share $\log s_{ij}^{shipper}$ from i to j on the price of trucking p_{ij}. IV uses snowiness of the route from i to j as a cost shifter. Standard errors in parentheses, clustered by origin and destination.

For details on how I construct route characteristics and other elements of the data, see Appendix D.2.

4.2 Demand Estimation

My model implies a linear equation for the share of shippers who choose trucking,

$$\log s_{ij}^{shipper} = \sigma p_{ij} + \omega_{ij} \tag{26}$$

This equation faces the classic simultaneity problem: equilibrium prices p are a function of demand shocks ω. I use a linear IV strategy to estimate σ. Specifically, I use supply shifters which affect prices but are independent of demand shocks ω: the snowiness of the route from i to j as a cost shifter. Snowy conditions make driving more difficult and intensify traffic, and multiple routes from the same origin or to the same destination may have different degreees of snow risk. In Table 3, I report demand estimation results.. The price coefficient is significant and negative. For interpretation, the estimated price coefficient on -0.323 corresponds to an average demand elasticity of -0.574, which is comparable in magnitude to previous work in the trucking freight literature as plotted in 9. Beuthe et al. (2001) estimate short distance trucking demand elasticities at -0.58 and long distance elasticities at -0.63. In one survey of the literature, Litman (2013) reports truck demand elasticities between -0.25 and -0.47. In another survey, Graham and Glaister (2004) report that 66 percent of estimates fall between -0.5 and -1.3. Trucking demand appears comparable and slightly less elastic compared to rail or water. Chen (2021) estimates a rail freight demand elasticity of -0.739, which is in line with earlier literature surveyed in that

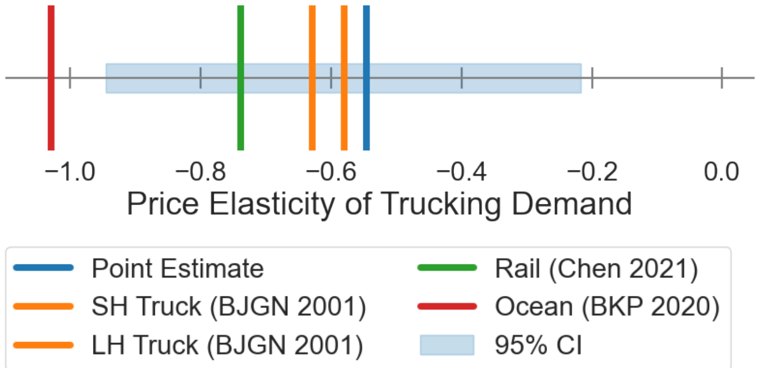

Figure 9: Demand Elasticity Comparison

Notes: This figure presents demand elasticity estimates in context with other point estimates from the literature.

paper. Brancaccio et al. (2020) finds a trade elasticity with respect to ocean freight shipping prices of -1.03.

4.3 Supply Estimation

The remaining structural parameters of my model are the discount factor β, the expected travel time λ_{ij}^{-1}, the home preference b, the supply price coefficient α, and the observable cost coefficients γ. I will calibrate β and λ and estimate the home preference and price and cost coefficients.

4.3.1 Overview

I present a high-level overview of my supply side estimation here; additional details can be found in Appendix A. My estimation problem is a standard dynamic discrete choice problem with two differences.

My first difference is that I do not observe choice data at the most granular level. Home preference is reflected in differences in choice probabilities across carriers of different types (i.e., from different home locations). However, my trucking share data is too aggregated - I do not observe the home location of the trucker engaged in a given trucking transaction. In the language of my model, I observe aggregate shares s_{ij} but not type-specific shares s_{ij}^h. To handle this, I will augment my choice data with auxiliary data which does contain the identify of the carrier: highway inspections. Each highway inspection contains a record of both the identity (registration) of the inspected carrier, along with the location of the inspection. This allows me to construct the differential geographic distributions of carriers of different types.

Differences in the inspection patterns of different types of carriers are informative about the magnitude of the home preference. Intuitively, suppose that the home preference is zero. All carriers are homogenous, so they will have identical inspection patterns across states. California carriers will make up the same share of inspected carriers in California as in Massachusetts. Alternatively, let the home preference go to infinity. Carriers will choose to only stay in their home state and do within-state jobs, so carriers will have very different inspection patterns. If the observed world is somewhere in the middle, then the degree of carrier geographic specialization will identify the level of home preference b. As a caveat, if I only observed data on the distribution of California carriers, any such pattern could be explained by the underlying pattern of trucking demand. For example, California carriers being mostly inspected near California could be consistent with any level of home preference if most trucking demand out of California is to nearby destinations. Crucially, I will observe different patterns between, for example, California and Massachusetts carriers, which allows me to control for the role of common demand shocks.

More formally, I assume that inspections are generated by the following process. For origin i, destination j, and intermediate location k, let m_{ijk} be the mileage share of the total ij route spent in k. A carrier is inspected in location k with probability ρ_k if the carrier is available in k, or with probability $\rho_i m_{ijk}$ if the carrier is en-route in ij. Carriers which take the outside option are not inspected. This process allows different states to have different inspection intensities, but restricts states from discriminating against carriers registered in any particular state. The probability of inspecting a type h carrier, conditional on making an inspection in k, is

$$\iota_k^h = \frac{\rho_k C_k^h + \sum_{i,j} \rho_k m_{ijk} C_{ij}^h}{\sum_{h'}(\rho_k C_k^{h'} + \sum_{i,j} \rho_k m_{ijk} C_{ij}^{h'})} = \frac{C_k^h + \sum_{i,j} m_{ijk} C_{ij}^h}{\sum_{h'}(C_k^{h'} + \sum_{i,j} m_{ijk} C_{ij}^{h'})} \tag{27}$$

In the first stage of my supply side estimation, I will estimate the magnitude of home preference which maximizes the likelihood of the observed pattern of inspections.

Second, my cost function contains unobservables shocks ξ_{ij}. To handle this, I will use an approach similar to BLP for demand estimation where I use demand shifters as instruments independent of the ξ_{ij} cost shocks. Recall the mean payoff of a route from i to j,

$$\delta_{ij} = \alpha p_{ij} + \gamma X_{ij} + \xi_{ij} \quad \forall (i,j) \tag{28}$$

For every level of home preference b, I can find a vector of mean payoffs which matches observed shares s^O_{ij}.

$$\delta^*(b, s^O) = \delta \quad s.t. \quad s_{ij}(b, \delta) = s^O_{ij} \in (0,1) \quad \forall (i,j) \tag{29}$$

For this vector to be unique, I must make L normalizations. I choose to normalize the flow payoff of taking the outside option $\delta_{OO,i}$ to zero for all locations i. I find this ex-ante more plausible than normalizing the payoff or the value of the unobservable cost for any other choice; the value of not working is plausibly more comparable across states than, for example, the value of a within-state trip. This normalization creates a convenient interpretation for b: the magnitude of the home preference is the difference in utils between not working at home versus not working somewhere else.

I solve for δ^* using an iterative algorithm which is analogous to the BLP contraction mapping, with the addition of accounting for how value functions and the equilibrium location of carriers depend on δ. For details on this algorithm, see Appendix A.1.

$$\delta'_{ij}(b) = \delta_{ij}(b) + \log s^O_{ij} - \log s_{ij}(\delta) \tag{30}$$

Given a value of δ^*, and a price instrument which is orthogonal to ξ_{ij}, I can recover price and cost coefficients with IV. Specifically, I use the availability of river and internal waterway shipping between i and j. Rivers increase the availability of substitutes for trucking (barge shipping), thereby shifting demand for trucking and varying prices independent of ξ_{ij}.

4.4 Estimation Results

I present my estimates in Table 4. Standard errors are taken from 500 bootstrap sample draws[11]. Given my estimated price coefficient, the daily flow payoff of being at home is about $93 per day. For context, this is one-half of the average daily wage of about $200 (Williams and Murray (2020)). Turning to the coefficients in Table 4, I find a positive coefficient on price and negative coefficients on the components of the cost function: distance and local diesel prices.

Parameter	Estimate (Utils)	Estimate ($)
Price ($)	0.315 (0.067)	-
Home Preference (daily)	29.69 (0.450)	93.46 (1.416)
Distance (miles)	-0.520 (0.080)	-1.648 (0.253)
Diesel ($/gal x miles)	-1.041 (0.033)	-3.305 (0.104)

The first column presents estimates from the supply-side estimation in util units. The second column presents a 95% bootstrap confidence interval from 500 bootstrap draws. The third column converts estimates into dollar units using the estimated price sensitivity.

Table 4: Supply Structural Estimates

For interpretation of the magnitudes, I divide the coefficients by the estimated price coefficient α to convert into dollar units. I estimate the marginal cost of distance at $1.648 per mile. This marginal cost is comparable with contemporaneous 2019 industry average cost estimates of $1.55 per mile.[12] I find that carriers value a route with $1 per gallon lower average diesel prices at $3.305 per mile.

To examine the effect of different model assumptions, I compare price and cost estimates across four specifications in Table 5. The first column contains estimates without origin and destination fixed effects, without using a price instrument, and holding home preference b to zero. The price coefficient is negative, which is implausible for the supply-side, and the diesel coefficient is positive. The second column

[11] I interpret my trip and inspection data, which is observed weekly, as coming from repeated samples from an equilibrium. I therefore perform a block bootstrap which resamples data with replacement at the weekly level. For robustness to correlation across time and states, I also perform a block bootstrap which resamples data at the state-week level. The resulting confidence intervals are similar to the original bootstrap confidence intervals.

[12] See Williams and Murray (2020) for this analysis using an accounting approach. They sum over fuel costs, capital lease and maintenance, license and toll expenses, and driver wages to arrive at their number.

adds origin and destination fixed effects. The price coefficient rises and is slightly positive, while the diesel coefficient becomes negative. This specification implies a marginal cost of distance of over \$17 per mile. The third column adds the price instrument. This specification implies a more plausible marginal cost of distance of \$2.34. The difference between the IV and OLS coefficients is consistent with prices being correlated with unobserved cost shocks, which would cause an upward bias in the price coefficient. For example, if highly demanded transcontinental routes also involve more difficult terrain, then carriers would appear less sensitive to prices than expected. The fourth column presents my main specification as in Table 4. The price coefficient falls and the distance coefficient falls. This is consistent with short-range trips returning home being more affected by home preference. If these trips tend to be lower distance and higher price, then failing to consider home preference would over-estimate carriers' sensitivity to price and distance.

Parameter	Spec 1	Spec 2	Spec 3	Main Specification
Price (\$)	-0.249	0.0272	0.386	0.315
Home Preference (daily)	0	0	0	29.69
Distance (miles)	-0.170	-0.482	-0.905	-0.520
Diesel (\$/gal x miles)	0.0357	-1.152	-1.269	-1.041
Fixed Effects	No	Yes	Yes	Yes
Price Instrument	No	No	Yes	Yes
Estimated b	No	No	No	Yes

This table presents estimates from alternative regression specifications. The first column omits origin and destination fixed effects, forces home preference to be zero, and omits a price instrument. The second column adds origin and destination fixed effects. The third column adds a price instrument. The fourth column estimates home preference, and is my main specification.

Table 5: Comparing Estimates Across Specifications

In Table 6, I study the effect of additional route covariates on the cost function. In specification five, I add the international roughness index (IRI), as a measure of road quality, and the density of truck parking spaces, as a measure of local amenities, as components of X_{ij}. For both, I construct the average characteristics over the entire length of the trip. I find a positive coefficient on IRI, indicating a preference for rougher roads, and a negative coefficient on truck parking spaces. Since road quality responds to local public investment, which may take trucking demand into account, I may be concerned about endogeneity. In specification six, I introduce an additional instrument, rail density, and treat IRI as potentially endogenous. I define rail density as the ratio of railroad tracks to highway miles, averaged over the length of the truck trip. Intuitively, this operates similarly to the river instrument as a shifter of the outside option of shippers. Along routes where rail access is better, shippers may substitute to rail, so trucking receives a demand shock. Specification six finds that carriers prefer less rough roads, and they prefer routes with more parking spaces.

Parameter	Main Specification	Spec 5	Spec 6
Price ($)	0.315	0.317	0.340
Home Preference (daily)	29.69	29.69	29.69
Distance (miles)	-0.520	-0.523	-0.566
Diesel ($/gal x miles)	-1.041	-1.017	-1.338
IRI	-	0.0329	-0.231
Spaces (per 1000 miles)	-	-0.0031	0.0314
IRI Instrument	-	No	Yes

This table presents estimates from alternative regression specifications which add additional covariates.

Table 6: Additional Covariates

4.5 Quantifying Effect of Home Preference

To evaluate the effect of home preference on the trucking market, I compare the status quo against a no-home-preference simulation. Specifically, I set the counterfactual home preference b to zero for all carriers and compute a new equilibrium while holding all other parameters constant. I compute counterfactual equilibria using a tatonnement procedure. Beginning at a candidate price vector p^0, I iterate the price vector with excess supply until convergence. For a given level of prices, I update the value of δ, re-compute value functions, and re-compute the equilibrium distribution of carriers.

$$p_{ij}^{t+1} = p_{ij}^t + (S_{ij}(p^t) - D_{ij}(p^t)) \tag{31}$$

where demand and supply are computed as

$$D_{ij}(p^t) = N_{ij} \exp(\sigma p_{ij}^t + \hat{\omega}_{ij}) \tag{32}$$

$$S_{ij} = \sum_h C_i^h(\delta) \frac{\exp(\alpha p_{ij}^t - \bar{\kappa}_{ij} - \xi_{ij} + \tilde{\beta}_{ij} V_j^h(b,\delta))}{\sum_{k \in L} \exp(\alpha p_{ik}^t - \bar{\kappa}_{ik} - \xi_{ik} + \tilde{\beta}_{ik} V_k^h(b,\delta)) + \exp(\beta V_i^h(b,\delta))} \tag{33}$$

I find two main effects of home preference. First, it causes carriers to work closer to home, so in aggregate, carriers reallocate from working in driver-poor states to driver-rich states. This benefits certain locations rather than others. Second, the overall attractiveness of taking the outside option increases when home preference is present. This effectively shifts the supply curve in, thereby raising prices and lowering quantities.

In Figure 10, I plot the percentage change in the number of available carriers C_i in each state when going from the no-home-preference setting to the status quo. States with a high number of carriers — California, Texas, Illinois, and Florida — see increases in the number of available carriers as carriers increase their probabilities of return home. By comparison, low carrier states see a relative decline in the number of available carriers. Within the high-carrier states, California sees a much higher increase in carrier supply compared to Texas, despite having similar numbers of carrier residents. This reflects the difference in demand for trucking trips between California and Texas.

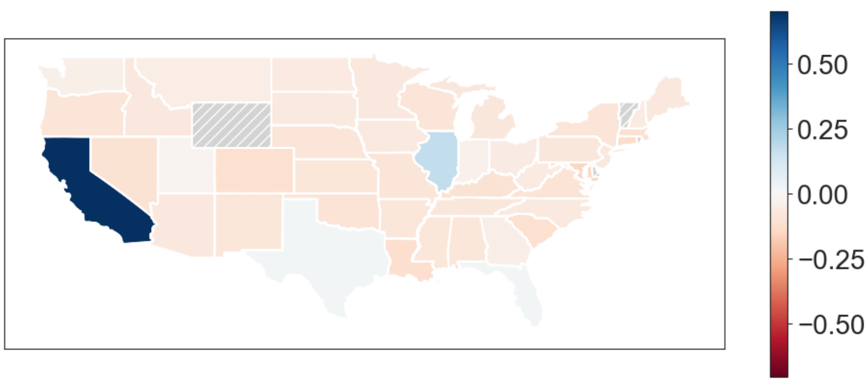

Figure 10: Home Preference: Carrier supply shifts toward carrier-rich states

Note: Figure 10 plots the percentage change in the supply of carriers in each state, where

$$\text{Percentage Change} = \frac{\sum_h C_i^h}{\sum_h C_{i,nohomepreference}^h} - 1$$

States which see rising supply are blue, while states which see falling supply are red. Larger magnitudes are indicated by darker colours.

This redistribution can also be seen in changing driving patterns. Table 7 presents statistics on driving patterns in the no-home-preference simulation and the baseline. When going from the no-home-preference simulation to the baseline, carriers spend four times as many days at home per year, and drive 17 percent closer to home on average.

	No Home Preference	Baseline
Time at Home (Days per Year)	8.30	32.26
Average Distance from Home (Miles)	1262.59	1050.75

Table 7: Effect of Home Preference on Driving Behavior

Carrier home preference also affects prices. I construct the change in a Laspeyres export price index for each state,

$$\Delta p_i^{export} = \frac{\sum_j p'_{ij} q_{ij}}{\sum_j p_{ij} q_{ij}} - 1 \tag{34}$$

which holds the composition of destinations fixed, where p' are the new set of prices being compared. Figure 11 plots the change in this export price index. Compared to a no-home-preference equilibrium, adding home preference causes prices to rise across the board due to decreased capital utilization. However, states which saw an increase in the supply of available carriers - especially California - see modest increases or decreases in prices. Home preference in the status quo subsidizes exports from California and other carrier-rich states relative to the rest of the U.S.

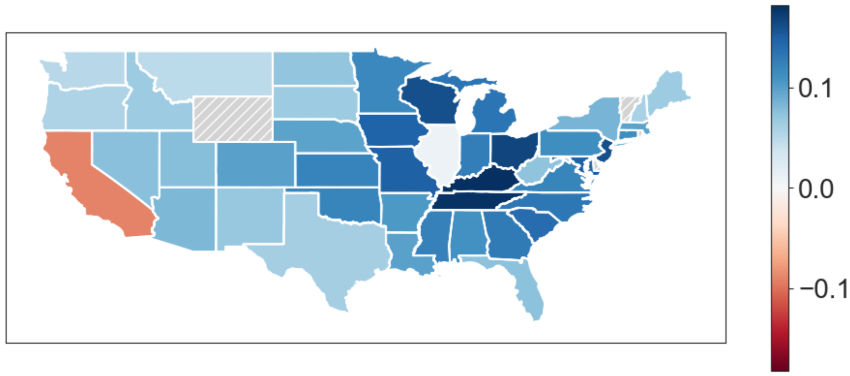

Figure 11: Home Preference: Higher average prices, lower prices for carrier-rich states

Note: Figure 11 plots the percentage change in the average trucking export price from each U.S. state, as defined by a Laspeyres index. States which see rising export prices are blue, while states which see falling export prices are red. Larger magnitudes are indicated by darker colours.

Diving deeper, I run a descriptive regression of prices on distance and empirical home shares. I present the regression coefficients in Table 8. In the no-home-preference world, routes with a high empirical home share have comparable prices with routes with low empirical home shares. When home preferences are present, the relationship becomes negative, in line with the pattern in the data. The coefficient on distance also increases modestly.

Price	No Home Preference	Baseline
Distance	1.155 (0.00683)	1.168 (0.00653)
Log Home Share	0.00199 (0.206)	-0.252 (0.129)
Constant	0.198 (0.0166)	0.352 (0.0161)
Fixed Effects	Yes	Yes
Observations	1936	1936
R-squared	0.987	0.987

Table 8: Pricing Regression (No Home Preference)

Note: This table presents results from a regression of prices in the no-home-preference simulation on distance and home shares. For consistency, I use empirical home shares rather than re-computing home shares in the simulation.

4.6 Extension: Choice over Home Locations

In the previous section, I have taken the home locations of carriers as exogeneous. However, carriers might choose their home locations given the wages and amenities of working as a long-haul truck driver in different states. In this section, I consider an extension where carriers first choose their home locations, and then participate in the steady-state equilibrium of my main model.

I use a nested logit model to capture this initial decision. There is a fixed population of potential carriers. Each of these carriers makes a one-time choice: they can choose to live in each location $i \in L$, and conditional on living in a location, they can choose to work as a long-haul driver, a non-long-haul driver, or an outside option. After they make their choice, they pay a local rent cost r_i, and then participate in the steady state equilibrium of the main model, taking all the individual carrier location choices as fixed.

If a carrier chooses to work as a long-haul driver in location i, they receive payoff

$$\pi(i, LH) = -\beta^r r_i + \beta^V E[V^i] + \beta^h h_i + \epsilon_{iLH}$$

where r_i is the rent in location i, $E[V^i]$ is the expected value function of a carrier from i, and h_i is the expected share of time spent at home for carriers from i.

If a carrier chooses to work as a non-long-haul driver in location i, they receive payoff

$$\pi(i, SH) = -\beta^r r_i + \beta^w w_i + \beta^h + \epsilon_{iSH}$$

where w_i is the average wages of a non-long-haul driver, and I assume that home time $h_i = 1$. Finally, the payoff of taking the outside option is

$$\pi(i, OO) = -\beta^r r_i + \epsilon_{iOO}$$

$(\epsilon_{iLH}, \epsilon_{iSH}, \epsilon_{iOO})$ are distributed a la a nested logit model, where all occupation choices within the same home location i are contained in a single nest. The carrier's problem, conditional on drawing a set of shocks ϵ, is

$$\max_{i \in L, j \in \{LH, SH, OO\}} \pi(i, j \mid \epsilon)$$

Model	Nested Logit
Value Function	1.167 (0.0324)
Home Time	1.146 (0.246)
Rent	-1.697 (0.401)
Outside Wages	0.294 (0.0816)
Observations	1936

Table 9: Extension: Choice over Home Locations

Notes: Robust standard errors in parentheses.

I bring a mix of predictions from my main estimation and additional data to estimate this extension. Expected value functions $E[V^i]$ and home time h_i are predictions from my estimated model. For each state, I construct rents r_i, wages w_i, and the estimated population of commercial drivers using the American Community Survey. I am interested in estimating three main parameters: β^r, the sensitivity of location choice to rent, β^V and β^w the sensitivity of location choice to long haul and non-long-haul profitability, and β^h, the sensitivity of location choice to the amount of time spent at home. I estimate these parameters using PyBLP and present the results in Table 9. As expected, carriers value both long-haul profits and non-long-haul trucking wages when making location choices. The latter is particularly relevant for states such as Massachusetts, which have very high wages for non-long-haul truckers. The presence of these attractive alternatives lowers the number of long-haul truckers. Carriers also dislike states with high rents, and prefer to work as long-haul truck drivers in states where the model-predicted time spent at home is high.

5 Counterfactual Simulations

I apply my model and estimates to two sets of counterfactuals. First, I consider the effect of a vehicle miles tax on trucking freight. Second, I consider the effect of a shift to self-driving trucks.

5.1 Counterfactual 1: Vehicle-Miles Tax

Freight trucking traffic is a major user and source of damages on U.S. road infrastructure. While fuel taxes have historically captured this externality in part, vehicle-miles taxes have been proposed as an alternative which directly targets driving behaviour. Advocates suggest that differential tax rates based on vehicle weight could account for the greater damage done by heavy trucks compared to lighter passenger vehicles and better capture the externalities of highway traffic (Boesen (2020)). A 2017 Congressional Budget Office found that, for example, the estimated total cost imposed by a fully-loaded combination truck weighing 80,000 pounds is about 19 cents per mile, significantly higher than the cost of a passenger vehicle (Office (2019)). Finally, implementing a vehicle miles tax has become more relevant recently as technological improvements have made monitoring more feasible, and rising gas mileage decreases the effectiveness of fuel taxes; the U.S.'s 2021 Infrastructure Investment and Jobs Act included an allocation to fund pilot programs for potential national, state, and local vehicle miles taxes.

Beyond the fiscal implications, a vehicle miles tax will also affect trucking prices and flows by changing the cost of travelling different routes. In my first counterfactual, I apply a vehicle miles tax by changing the marginal cost of distance $\gamma_{distance}$ and computing the resulting equilibrium. Specifically, I take the 19 cents per mile CBO estimate, and I use my estimated price sensitivity α to convert into utility units. Compared to my estimated marginal cost of distance of \$1.648 per mile, this represents a significant 11.5 percent increase in marginal costs. I plot the resulting change in the distribution of carriers in Figure 12. First, because longer routes are now more expensive relative to shorter routes, carriers prefer to be in states with shorter average length of haul. This lowers the number of carriers in the West (where lengths of haul is long) toward the East (where lengths of haul are shorter). Second, because travelling away from home is now more expensive, carriers are more likely to stay closer to their home state. States which are carrier-rich see increases in the number of available carriers.

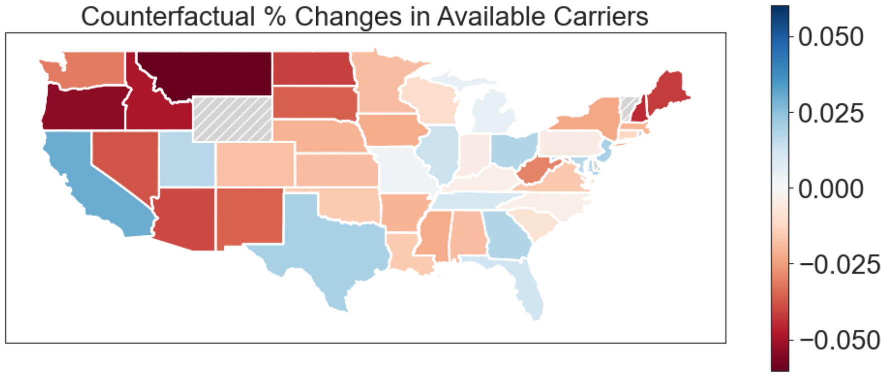

Figure 12: Counterfactual: Effect of Vehicle Miles Tax on Distribution of Carriers

Note: Figure 12 plots the percentage change in the supply of carriers in each state, where

$$\text{Percentage Change} = \frac{\sum_h C_i^h}{\sum_h C_{i,nohomepreference}^h} - 1$$

States which see rising supply are blue, while states which see falling supply are red. Larger magnitudes are indicated by darker colours.

Turning to prices, I compute changes in the export price index and plot the results in Figure 13. Prices rise across the country, and especially for long average length of haul states. Maryland, on the low end, sees a 4.5 percent increase in prices, while Iowa, on the high end, sees a 10.4 percent increase in prices. The effect of the vehicle miles tax will therefore not be felt uniformly by consumers across all states. On average for the entire country, prices rise by 7.1 percent, reflecting a passthrough of about 60 percent of the tax to prices.

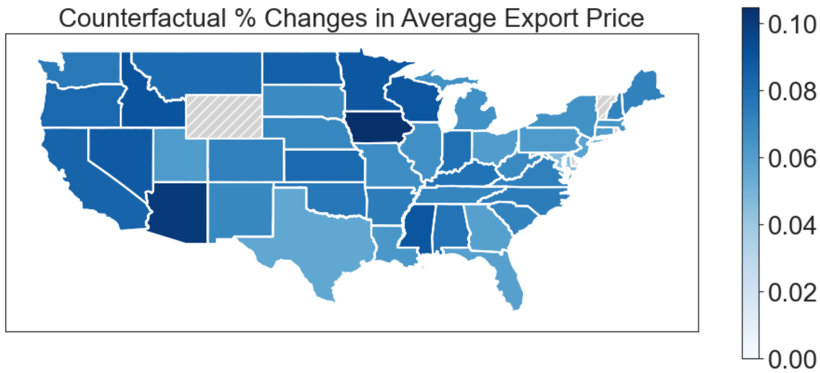

Figure 13: Counterfactual: Effect of Vehicle Miles Tax on Prices

Note: Figure 13 plots the percentage change in the average trucking export price from each U.S. state, as defined by a Laspeyres index. Larger magnitudes are indicated by darker colours.

Finally, if I run a parallel experiment where I apply a vehicle miles tax to the no-home-preference economy of Section 4.5, I find a slightly more modest price increase of 6.9 percent. Home preference has an amplifying effect on the vehicle miles tax, as carriers reallocate away from states where the tax is likely to have a large effect. This shrinks the local supply of carriers and raises prices further.

5.2 Counterfactual 2: Full Automation

In this second counterfactual, I study the effect of self-driving trucks by considering three channels: the elimination of home preference, decreasing marginal costs per mile, and increased daily driving range.

First, as in Section 4.5, I eliminate home preference by setting $b = 0$. Since I hold the value of the outside option constant, carriers can still choose not to operate. In the counterfactual, this would capture mechanical breakdowns or maintenance which stop work both at home and on the road[13].

Second, self-driving trucks are expected to experience decreases in marginal costs per mile along several dimensions.[14] Labor costs are eliminated, a decreased risk of

[13]In practice, returning to a home base for maintenance may induce a smaller level of home preference.

[14]Engholm, Pernestal, and Kristoffersson (2020) provides a detailed cost analysis of self-driving trucks.

accidents lowers depreciation and insurance costs, and more efficient driving increases fuel economy. Estimates of the total cost difference from the transportation literature range from 15 percent (Wadud (2017)) to 33 percent (Engholm et al. (2020)) for 40-ton trucks. Following Huang and Kockelman (2020), I consider a moderate scenario where costs-per-mile decrease by 25 percent. In the context of my model, $\gamma_{distance}$ and γ_{diesel} would decrease by 25 percent.

Third, self-driving trucks would no longer be subject to hours-of-service regulation which caps daily driving times for safety reasons. A 2021 Deloitte report suggests that without hours-of-service regulation, daily driving ranges would double (Zarif et al. (2021)). In my model, I set the stochastic arrival rate, λ_{ij}, to double the daily driving range.

I run counterfactual simulations with each of the eliminating home preference, lowering costs, and increasing range counterfactuals, as well as a full counterfactual with all three changes. In Table 10, I present the effects of each counterfactual on three measures: aggregate prices, total quantities, and shipper welfare. Under the full automation counterfactual, prices fall by 28.6 percent and shipper welfare increases by \$16.65B per year. For scale, U.S. Census reported annual expenditure on long-distance general freight trucking at \$111.1B for 2019. On average, eliminating home preference would be responsible for about 25 percent of the total impact of automation on prices, quantities and welfare, comparable to the effect of double daily driving range.

Counterfactual	Price Index	Total Quantity	Shipper Welfare (\$B/year)
No Home preference	-7.7%	+4.1%	+3.91B
Lower Per-Mile Costs	-14.0%	+7.6%	+7.62B
Longer Daily Range	-7.2%	+4.0%	+4.36B
Full Automation	-28.6%	+16.3%	+16.65B

Table 10: Counterfactual: Automation Effects Summary

In Figure 14, I plot the change in average export prices by state. The states that see the largest declines in prices are carrier-poor states with long average length of haul. The elimination of home preference benefits the carrier-poor states, while lower per-mile marginal costs and increased daily driving range benefits the long average length of haul states. These are especially focused in the West and Midwest, including states such as Iowa, Arizona, and Idaho. Among the larger states, Texas sees larger declines in prices relative to California because it has a smaller relative decline in carrier supply.

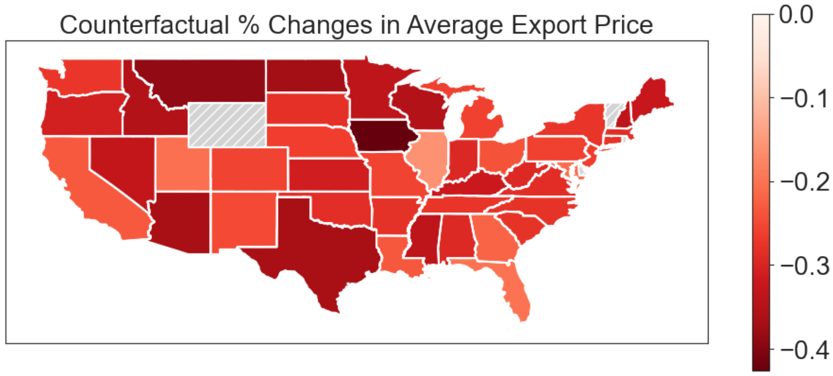

Figure 14: Counterfactual: Effect of Trucking Automation on Prices

Note: In a second self-driving truck counterfactual, I eliminate home preference while lowering per-mile costs by 25 percent and doubling daily driving range. Figure 14 plots the percentage change in export prices from each state as defined by a Laspeyres index which weights prices by quantity in the baseline. States which see rising export prices are blue, while states which see falling export prices are red. Larger magnitudes are indicated by darker colours.

In this counterfactual, I hold demand fixed and ignore externalities from trucking. Longer daily driving ranges would decrease travel times, which would further improve shipper welfare if shippers care about delivery time. In addition, time-sensitive shippers may substitute more from air (the next fastest mode) in ways which are not captured by my demand estimation. Major externalities from trucking include emissions, congestion, and road wear and tear. The increase in quantity under the full counterfactual yields 2.27 billion additional miles driven per year. Using EPA estimates of emissions (65 grams per ton-mile) and gas mileage (7 miles per gallon), this translates into 325 million additional gallons of diesel consumed per year and 3.33 million additional metric tons of CO_2 per year. By comparison, Liu et al. (2018) estimates a 3 percent decrease in CO_2 emissions for self-driving vehicles, suggesting that greater efficiency from self-driving vehicles is not enough to counteract the effect of increased quantity. To study the potential incidence for these externalities, I compute the change in trucking miles travelled by state in Figure 15. The regions that see the greatest increases in trucking mileage are the Northwest and the Northeast. These are regions where long length of haul trips tend to pass through, and also begin or end in a carrier-poor state.

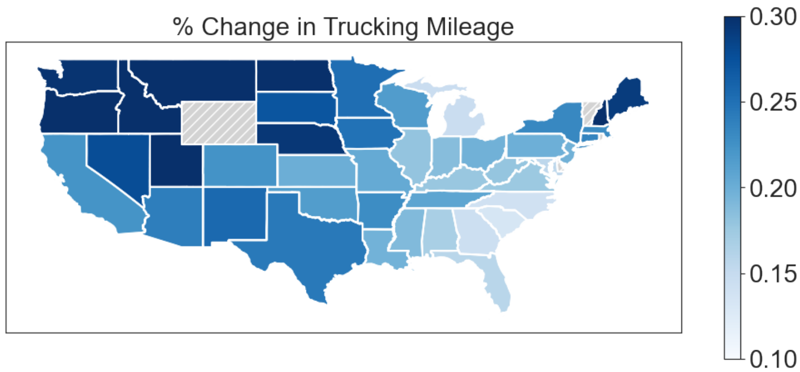

Figure 15: Counterfactual: Effect of Trucking Automation on Trucking Mileage

Note: Figure 15 plots the percentage change in the amount of trucking traffic passing through each U.S. state. Larger magnitudes are indicated by darker colours.

To summarize the effects of my counterfactuals on prices, I run regressions of prices under each counterfactual on distance and home share, and present the results in Table 14. Under the vehicle-miles-tax counterfactual, as expected, the increase in marginal costs causes the relationship between prices and distance to become steeper. This increase is greater than it would be without home preference, as carriers also relocate away from states which ship long distances on average. Under the automation counterfactual, all three effects - elimination of home preference, falling marginal costs per mile, and increased daily driving range - contribute to a flatter relationship between prices and distance. The role of home share also becomes smaller under automation, as I would expect due to the elimination of home preference. For more detailed analysis, I present additional counterfactuals in Appendix C. In my main counterfactuals, I focus on a long-run outcome where all trucks are replaced with self-driving trucks. In counterfactual C.1, I consider a transition path to automation using a set of counterfactuals where a portion of the trucking fleet becomes self-driving while the remainder is human. I find a smooth transition path in terms of overall prices and quantities. In counterfactual C.2, I consider a transition path to automation where self-driving trucks are geographically constrained along the transition path. At first, self-driving trucks gain access to technically simple routes in the sunny Southwest,

Price	Baseline	VMT	Full Automation
Distance	1.168 (0.00653)	1.263 (0.00650)	0.851 (0.00746)
Log Home Share	-0.252 (0.129)	-0.285 (0.124)	-0.130 (0.164)
Constant	0.352 (0.0161)	0.382 (0.0161)	0.388 (0.0187)
Fixed Effects	Yes	Yes	Yes
Observations	1936	1936	1936
R-squared	0.987	0.989	0.975

Table 11: Counterfactual Pricing Equations

Note: This table presents results from a regression of prices in the automation counterfactual on distance and home shares. For consistency, I use empirical home shares rather than re-computing home shares in the simulation.

before expanding North and East. I find that geographic restrictions cause trucks to be misallocated and freight costs can increase even if self-driving trucks have lower per-mile costs and longer driving range. Intermediate levels of automation, combined with geographic restrictions, can lead to price increases and welfare losses for states which do not have yet access to self-driving trucks.

6 Conclusion

In this paper, I have developed a dynamic spatial model of trucking freight, estimated demand, cost, and carrier preference parameters, and conducted a counterfactual simulation where I study the potential effect of self-driving trucks to remove carrier home preference. I have found that human preferences to return home are significant and shape the cost of transportation across the United States today.

Drivers' preferences to return home systematically subsidize driver-rich states and create a divide between short-haul and long-haul shipping. From the perspective of a driver, short-haul routes are much less lucrative than long-haul routes to compensate for being able to return home. From the perspective of a shipper, short-haul routes are attractively cheaper on a per-mile basis than longer routes - unless those long routes are to destinations which can routinely bring a driver home with return traffic. Under automation, the distinction between short- and long-haul routes collapses, and the two markets integrate. My findings suggest a new way that automation and artificial intelligence can affect our economy. When human labor cannot be substituted with capital, worker preferences will shape how and where production occurs. Artificial intelligence can de-link workers from production, and workers' preferences from outcomes.

References

References

Adler, A. (2020): "ATA: Trucking still rules freight movement," *Freightwaves*.

Allen, T. and C. Arkolakis (2019): "The Welfare Effects of Transportation Infrastructure Improvements," 1–47.

Allen, T., D. Atkin, S. Cantillo, and C. Hernandez (2020): "Trucks or The Triple Curse of Remoteness," .

Association, A. T. (2020): "Turnover Rate at Truckload Carriers Rose in Third Quarter," Tech. rep.

Baker, G. P. and T. N. Hubbard (2003): "Make versus buy in trucking: Asset ownership, job design, and information," *American Economic Review*, 93, 551–572.

——— (2004): "Contractibility and asset ownership: On-board computers and governance in U. S. trucking," *Quarterly Journal of Economics*, 119, 1443–1479, iSBN: 9780080453705.

Bao, K. and R. Mundy (2018): "Emerging Freight Truck Technologies: Effects on Relative Freight Costs," Tech. rep., Center for Transportation Studies, University of Missouri - St. Louis, St. Louis, issue: May.

Barla, P., D. Bolduc, N. Boucher, and J. Watters (2010): "Information technology and efficiency in trucking," *Canadian Journal of Economics*, 43, 254–279.

Behrens, K. and P. M. Picard (2011): "Transportation, freight rates, and economic geography," *Journal of International Economics*, 85, 280–291.

Berry, S., J. Levinsohn, A. Pakes, and A. Pakes1 (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Beuthe, M., B. Jourquin, J. F. Geerts, and C. Koul À Ndjang'Ha (2001): "Freight transportation demand elasticities: A geographic multimodal transportation network analysis," *Transportation Research Part E: Logistics and Transportation Review*, 37, 253–266.

Boesen, U. (2020): "Who Will Pay for the Roads?" Tech. Rep. 725, Tax Foundation.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020): "Geography, Transportation, and Endogenous Trade Costs," *Econometrica*, 88, 657–691.

Buchholz, N. (2022): "Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry," *Review of Economic Studies*, 89, 556–591.

Burks, S., M. Belzer, Q. Kwan, S. Pratt, and S. Shackelford (2017): "Trucking 101," .

Burks, S. V., K. Monaco, and A. Kildegaard (2018): "Is the Labor Market for Truck Drivers Broken, and Will Autonomous Trucks Fix It ?" Tech. rep., Annual Transportation Research Forum.

Carrone, A. P., J. Rich, C. A. Vandet, and K. An (2021): *Autonomous vehicles in mixed motorway traffic: capacity utilisation, impact and policy implications*, Springer US, publication Title: Transportation Issue: 0123456789 ISSN: 15729435.

Chen, Y. (2021): "Network Structure and Efficiency Gain from Mergers: Evidence from the U.S. Freight Railroads," .

Chiang, S. J. W. and A. F. Friedlaender (1984): "Output Aggregation, Network Effects, and the Measurement of Trucking Technology," *The Review of Economics and Statistics*, 66, 267.

Clements, L. M. and K. M. Kockelman (2017): "Economic effects of automated vehicles," *Transportation Research Record*, 2606, 106–114.

Corsi, T., C. Grimm, K. Smith, and R. Smith (1992): "The Effects of LTL Motor Carrier Size on Strategy and Performance," *Logistics and Transportation Review*, 28.

Daughety, A. F. and F. D. Nelson (1988): "An econometric analysis of changes in the cost and production structure of the trucking industry, 1953-1982," *Review of Economics & Statistics*, 70, 67–75.

Day, J. and A. Hait (2019): "American Keeps on Truckin," .

Duranton, G., P. M. Morrow, and M. A. Turner (2014): "Roads and trade: Evidence from the US," *Review of Economic Studies*, 81, 681–724.

Duranton, G. and M. A. Turner (2012): "Urban growth and transportation," *Review of Economic Studies*, 79, 1407–1440.

Engholm, A., A. Pernestål, and I. Kristoffersson (2020): "Cost analysis of driverless truck operations," *Transportation Research Record*, 2674, 511–524.

Fajgelbaum, P. D. (2016): "Optimal Transport Networks in Spatial Equilibrium," .

Fréchette, G. R., A. Lizzeri, and T. Salz (2019): "Frictions in a Competitive, Regulated Market: Evidence from Taxis," *American Economic Review*, 109, 2954–2992.

Gagne, R. (1990): "On the Relevant Elasticity Estimates for Cost Structure Analyses of the Trucking Industry," *The Review of Economics and Statistics*, 72, 160.

Ghandriz, T., B. Jacobson, L. Laine, and J. Hellgren (2020): "Impact of automated driving systems on road freight transport and electrified propulsion of heavy vehicles," *Transportation Research Part C: Emerging Technologies*, 115, 102610, publisher: Elsevier.

Gittleman, M. and K. Monaco (2020): "Truck-Driving Jobs: Are They Headed for Rapid Elimination?" *ILR Review*, 73, 3–24.

Glaeser, E. L. (2005): "Reinventing Boston: 1630-2003," *Journal of Economic Geography*, 5, 119–153.

Graham, D. J. and S. Glaister (2004): "Road traffic demand elasticity estimates: A review," *Transport Reviews*, 24, 261–274.

Hayakawa, K., J. Ishikawa, and N. Tarui (2020): "What goes around comes around: Export-enhancing effects of import-tariff reductions," *Journal of International Economics*, 126, 103362, publisher: The Authors.

Holm, S. (2020): "How Big is the Contract Market?" *Freightwaves*.

Huang, Y. and K. M. Kockelman (2020): "What will autonomous trucking do to U.S. trade flows? Application of the random-utility-based multi-regional input–output model," *Transportation*, 47, 2529–2556, publisher: Springer US ISBN: 0123456789.

Hubbard, T. N. (1998): "Governance Structure in the Deregulated Trucking Industry," .

——— (2000): "The Demand for Monitoring Technologies: The Case of Trucking," *The Quarterly Journal of Economics*, 115, 533–560.

——— (2001): "Contractual Form and Market Thickness in Trucking," *The RAND Journal of Economics*, 32, 369.

——— (2003): "Information, Decisions, and Productivity On-Board Computers and Capacity Utilization in Trucking Thomas N. Hubbard University of Chicago and NBER," *THE AMERICAN ECONOMIC REVIEW*.

Ishikawa, J. and N. Tarui (2018): "Backfiring with backhaul problems: Trade and industrial policies with endogenous transport costs," *Journal of International Economics*, 111, 81–98, publisher: Elsevier B.V.

Keeler, T. (1989): "Deregulation and Scale Economies in the U.S. Trucking Industry: An Econometric Extension of the Survivor Principle," *Journal of Law and Economics*, 32.

Lagos, R. (2003): "An analysis of the market for taxicab rides in New York City," *International Economic Review*, 44, 423–434.

Laing, A. and J. Nolan (2009): "Price dynamics and market structure in transportation: For-hire grain trucking along the Alberta-Saskatchewan border," .

Liang, Y. (2021): "Do Safety Inspections Improve Safety? Evidence from the Roadside Inspection Program for Commercial Vehicles," *SSRN Electronic Journal*.

Litman, T. A. (2013): "Understanding transport demands and elasticities: how prices and other factors affect travel behavior," *Victoria Transport Policy Institute*, 1–76.

Liu, J., K. M. Kockelman, and A. Nichols (2018): "ANTICIPATING THE EMISSIONS IMPACTS OF SMOOTHER DRIVING BY CONNECTED AND AUTONOMOUS VEHICLES, USING THE MOVES MODEL," in *Smart Transport for Cities & Nations: The Rise of Self-Driving & Connected Vehicles*, 445p.

Moosavi, S., M. H. Samavatian, A. Nandi, S. Parthasarathy, and R. Ramnath (2019): "Short and long-term pattern discovery over large-scale geo-spatiotemporal data," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2905–2913, arXiv: 1902.06792.

Office, C. B. (2019): "Issues and Options for a Tax on Vehicle Miles Traveled by Commercial Trucks," Tech. rep., Congress of the United States.

Ostrovsky, M. and M. Schwarz (2018): "Carpooling and the Economics of Self-Driving Cars," Tech. rep.

Oum, T. H. (1979): "A Cross Sectional Study of Freight Transport Demand and Rail-Truck Competition in Canada," *The Bell Journal of Economics*, 10, 463.

Rosaia, N. (2020): "Competing platforms and transport equilibrium: Evidence from New York City," .

Savage, I. (1995): "Panzar and Rosse style tests of market structure in the US motor carrier industry," *Logistics & Transportation Review*, 31, 135–144.

Spady, R. H. and A. F. Friedlaender (1978): "Hedonic Cost Functions for the Regulated Trucking Industry," *The Bell Journal of Economics*, 9, 159.

Staff (2020): "How much weight can a big rig carry?" *Freightwaves*.

Viscelli, S. (2018): "Driverless? Autonomous Trucks and the Future of the American Trucker," 73.

——— (2020): "Will Robotic Trucks Be "Sweatshops on Wheels"?" *Issues in Science and Technology*, 79–88.

Vreugdenhil, N. (2018): "Booms , Busts , and Mismatch in Capital Markets : Evidence from the Offshore Oil and Gas Industry," 1–69.

Wadud, Z. (2017): "Fully automated vehicles: A cost of ownership analysis to inform early adoption," *Transportation Research Part A: Policy and Practice*, 101, 163–176.

Waschik, R., C. L. Taylor, D. Friedman, and J. Boatner (2021): *Macroeconomic Impacts of Automated Driving Systems in Long-Haul Trucking*.

Williams, N. and D. Murray (2020): "An analysis of the operational costs of trucking: 2020 update," Tech. rep., American Transportation Research Institute, Arlington, Virginia, issue: November.

Winston, C. and Q. Karpilow (2020): *Autonomous Vehicles: The road to economic growth ?*, publication Title: Brookings Institution Press ISSN: 00118583.

Wong, W. F. (2020): "The Round Trip Effect: Endogenous Transport Costs and International Trade," *Working Paper*, 1–77.

Zarif, R., C. Starks, A. Sussman, A. Kukreja, and S. Abidi (2021): "Autonomous Trucks Lead the Way," Tech. rep., Deloitte Center for Integrated Research The.

Appendix A: Estimation Details

A.1 Computational Details of $\delta^*(b)$

I will compute $\delta^*(b)$ using the following iterative algorithm analogous to the nested fixed point of Berry et al. (1995).

Iterative Algorithm Recall that a carrier which starts traveling from i to j has a probability λ_{ij} of arriving tomorrow.

$$W_{ij}^h = \lambda_{ij} E[V_j^h] + (1 - \lambda_{ij})\beta W_{ij}^h \tag{35}$$

I can rearrange this to find W_{ij}^h as a function of V_j^h.

$$W_{ij}^h = \frac{\lambda_{ij}}{1 - \beta(1 - \lambda_{ij})} V_j^h \tag{36}$$

For notational convenience, let $\tilde{\beta}_{ij} = \frac{\lambda_{ij}}{1-\beta(1-\lambda_{ij})}$ be the effective route-specific "discount factor." This allows me to dispense with tracking the en-route state value functions W_{ij}^h.

Step 1. Given a guess for (b, δ), compute V_i^h.

In a steady-state equilibrium, conditional on δ, a carrier does not need to know what any other carrier is doing in the future. Let T be the following operator defined on a space $B(X)$ of functions $f : X \to \mathbb{R}$ where $X = L \times L$.

$$Tf(h, i) = b \times 1_{h=i} + E_\epsilon[\max\{\max_j\{\delta_{ij} + \tilde{\beta}_{ij} f(j) + \epsilon_j\}, \beta f(i)] \tag{37}$$

T is a contraction mapping, so V_i^h can be computed using value function iteration.

Step 2. Given V_i^h, compute the type-specific choice probabilities s_{ij}^h using Equation 8.

$$s_{ij}^h(p) = \frac{\exp(\delta_{ij} + \tilde{\beta}_{ij} V_j^h)}{\sum_{k \in L} \exp(\delta_{ik} + \tilde{\beta}_{ik} V_k^h) + \exp(\beta V_i^h)} \tag{38}$$

Step 3. Given s_{ij}^h, compute equilibrium carrier locations C_i^h.

The en-route state steady-state condition implies that the population of the en-route carriers is a multiple of the origin location carriers.

$$(1 - \lambda_{ij}) C_i^h s_{ij}^h = \lambda_{ij} C_{ij}^h \implies C_{ij}^h = \frac{(1 - \lambda_{ij}) s_{ij}^h}{\lambda_{ij}} C_i^h \tag{39}$$

Substitute this into the available state steady-state condition to find a relationship between all available carrier locations.

$$\sum_k \lambda_{ki} (C_{ki}^h + C_k^h s_{ki}) = C_i^h \sum_j s_{ij}^h \implies C_i^h = \frac{\sum_j s_{ji}^h C_j^h}{\sum_j s_{ij}^h} \tag{40}$$

Finally, the carriers must sum up.

$$C^h = \sum_i C_i^h + \sum_{i,j} C_{ij}^h = \sum_i C_i^h (1 + \sum_j \frac{(1 - \lambda_{ij}) s_{ij}^h}{\lambda_{ij}}) \tag{41}$$

I can combine these two equations as a linear system

$$\begin{pmatrix} A \\ B \end{pmatrix} (C_1^h \dots C_L^h) = \begin{pmatrix} C_1^h \\ \dots \\ C_L^h \\ C^h \end{pmatrix} \tag{42}$$

where A is an $L \times L$ matrix where $A_{ij} = \frac{s_{ji}^h}{1 - \sum_j s_{ij}^h}$ and B is an $1 \times L$ matrix where $B_i = 1 + \sum_j \frac{(1-\lambda_{ij}) s_{ij}^h}{\lambda_{ij}}$. This system can be solved for C^h. Intuitively, A captures the flows across locations, and B enforces that the distribution of carriers sums to $\bar{C}^h$.

Step 4. Given (C_i^h, s_{ij}^h), compute predicted aggregate shares $s_{ij}(b, \delta)$ using Equation 10.

$$s_{ij} = \sum_h \left(\frac{C_i^h}{\sum_{h'} C_i^{h'}}\right) s_{ij}^h \tag{43}$$

Step 5. Repeat Steps 1-4 and iterate $\delta'_{ij} = \delta_{ij} + \log s_{ij}^O - \log s_{ij}(\delta)$ until convergence.[15]

A.2 Supply-side Estimation Steps

I conduct the supply-side estimation in two steps.

In the first step, I estimate the home preference by maximizing the likelihood of observed inspections. For each guess of the home preference parameter b, I use the δ^*-algorithm to solve for mean payoffs which match observed shares. As part of computing δ^*, I also find equilibrium carrier locations $C(b, \delta)$ which are consistent with b and $\delta^*(b, s^O)$. These imply the probability of inspecting a carrier from h conditional on drawing an inspection in k.

$$\iota_k^h(b, \delta) = \frac{C_k^h(b, \delta) + \sum_{i,j} m_{ijk} C_{ij}^h(b, \delta)}{\sum_{h'}(C_k^{h'}(b, \delta) + \sum_{i,j} m_{ijk} C_{ij}^{h'}(b, \delta))} \tag{44}$$

Combine with the probability of drawing an inspection in k, ρ_k, to form the first stage optimization problem.

$$b^* = \arg\min_{b, \rho_k} \quad \sum_{h,i} I_k^h \times \log \iota_k^h(b, \delta^*(b, s^O)) \times \rho_k \tag{45}$$

where

$$\delta^*(b, s^O) = \delta \quad .s.t \quad s_{ij}(b, \delta) = s_{ij}^O \quad \forall i, j \tag{46}$$

In the second step, I take the mean payoffs corresponding to the optimal home preference, $\delta^*(b^*, s^O)$ and I estimate price and cost coefficients using linear IV. I use the availability of rivers as a price instrument.

[15] For one carrier type, it is straightforward to prove that this δ^*-algorithm is a contraction mapping. For multiple carrier types, the algorithm converges quickly to a solution in practice.

A.3 Observing the Outside Share

To operationalize this estimation, I need to be able to compute empirical aggregate shares, s_{ij}. In my DAT RateView dataset I only observe carriers who choose to sell shipping. To resolve this, I turn to the DAT Trucks in Market dataset, which records search on the DAT marketplace by location. These searches should include both carriers chose to work, as well as carriers who searched and decided not to work, and will therefore be informative about the number of carriers who took the outside option.

Let T_i be the number of searches made by carriers in 2019. I assume that number of available carriers are distributed proportional to the share of searching carriers in the Trucks in Market dataset.

$$C_i = \left(\frac{T_i}{\sum_i T_i}\right) \times \left(\sum_h C^h\right) \tag{47}$$

I can then compute empirical aggregate shares:

$$s_{ij}^O = \frac{Q_{ij}}{C_i} \tag{48}$$

A.4 Calibrated Parameters

The expected arrival time for a trip is $1/\lambda_{ij}$. For across-state trips, I set $\lambda_{ij} = \frac{500}{d_{ij}}$ to fit the expected numbers of days to travel distance d_{ij}, under the assumption that carriers average speed 45 miles per hour and drive the DOT-mandated maximum of 11 hours a day.[16] For within-state trips of distance less than 250 miles, I set λ_{ij} to 1 because a carrier can make a round trip in one day. Otherwise, I set $\lambda_{ii} = \frac{250}{d_{ii}}$. In practice, only California has average within-state trips above 250 miles; this accounts for the lower probability that a carrier taking a within-California trip can return home and receive a home preference.

I calibrate the discount factor β to 0.995. Specifically, since the time periods are short, I set the discount factor to reflect carrier exit rates rather than time preferences.

[16]See Williams and Murray (2020) for survey-based statistics and average moving speed. See Federal Motor Carrier Safety Administration (2020) for the full hours-of-service regulations.

Suppose that carriers are patient but have an exogenous $1 - \beta = 0.005$ probability of exiting the market every day and being replaced by an identical carrier. This generates a discount factor of β, a daily exit probability of 0.5 percent, and an annual attrition rate of 84 percent which is consistent with 2019 industry annual attrition rates of 70-100 percent (e.g. Association (2020)).

Appendix B: Summary Figures

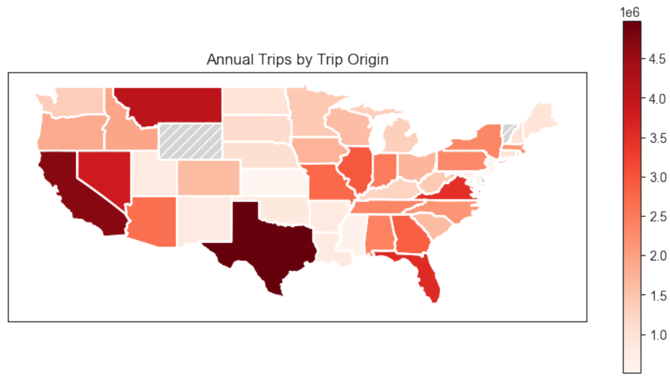

Figure 16: Annual Trips by Trip by Origin / Destination

Note: Panel A plots the total number of annual trips from each state as an origin. Panel B plots the total number of annual trips to each state as a destination. Larger magnitudes are indicated by darker colours. The largest origins and destinations for trucking freight are California and Texas.
Data source: 2019 DAT RateView

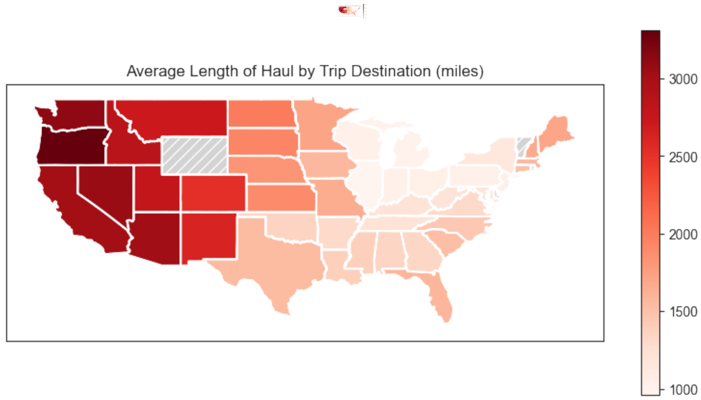

Figure 17: Average Length of Haul by Origin / Destination

Note: Panel A plots the average length of haul (trip distance) for trips from each state as an origin. Panel B plots the average length of haul for trips to each state as a destination. The West Coast has the longest average trip distances, while the Midwest has the shortest average trip distances.
Data source: 2019 DAT RateView

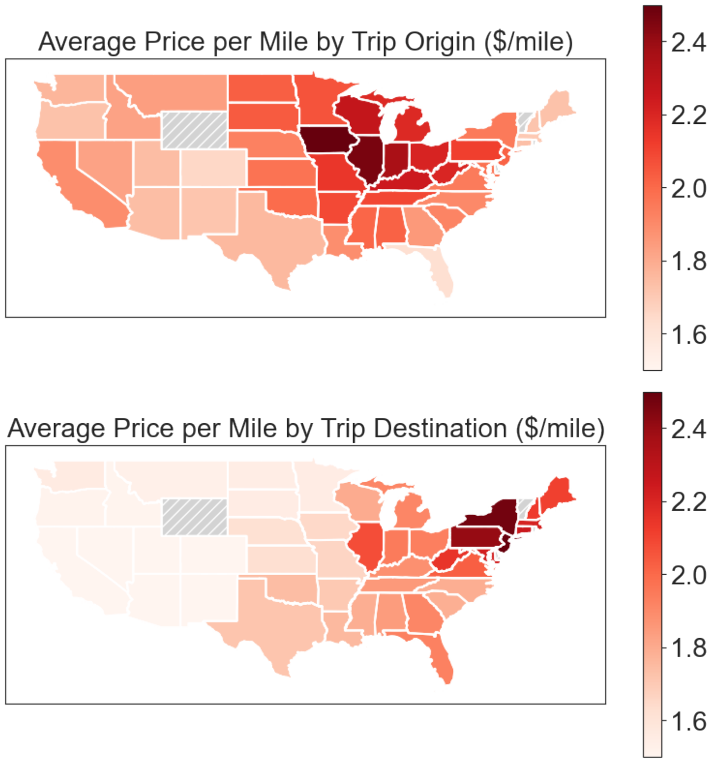

Figure 18: Average Price per Mile by Origin / Destination

Note: Panel A plots the average price per mile of trucking freight for each state as an origin. Panel B plots the average price per mile for each state as a destination. The Midwest has the highest average export prices, while the coastal states have lower average export prices. Destination states which are further east have higher prices per mile for imports.
Data source: 2019 DAT RateView

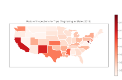

Figure 19: Ratio of Inspections to Trips Originating from Each State

Note: Figure plots the ratio of inspections conducted to number of trips for each state. Consistent with industry interviews, several states appear to be more likely to inspect, including California, Texas, Arizona, and Maryland.
Data source: 2019 DAT RateView, 2019 FMCSA Motor Carrier Inspections

6.0.1 Distance and Prices

In Figure 20, I plot the relationship between distance, prices, and quantity shipped in this market. Panel A presents a bin scatter of price against distance at the lane level. Panel B presents a bin scatter of log quantity against distance at the lane level. Quantities fall with distance, consistent with a gravity-like relationship in trucking flows.

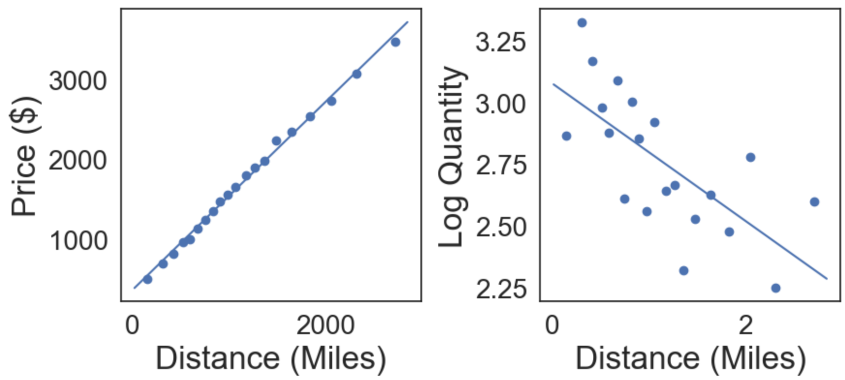

Figure 20: Longer distance trips are correlated with higher prices and lower quantity

6.0.2 Carriers are more likely to work close to home

In Figure 21, I document that carriers are more likely to be inspected in states that are close to their home state. Each highway inspection records the home (registration) state and current state of the inspection. I compute the distance between the centroids of the home and inspection states and plot a density histogram of those distances. Carriers are more likely to be inspected in their home state or in adjacent states than random. For scale, a randomly drawn pair of U.S. states has on average 1,000 miles between their centroids. By comparison, inspected carriers are on average 871 miles away from their home state centroid.

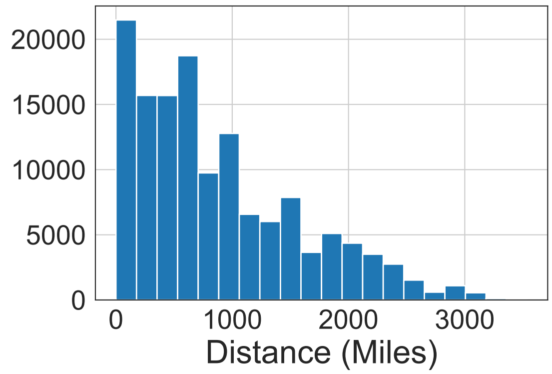

Figure 21: Carriers are more likely to be inspected closer to home

Note: This figure plots a histogram of carrier inspections by the distance from the state of inspection to the home state of the carrier. Distances are measured between state centroids. Data sources: 2019 FMCSA Motor Carrier Inspections

Appendix C: Additional Counterfactuals

Counterfactual C.1: Transition to Automation

In my main counterfactuals, I consider a complete transition from human drivers to self-driving trucks. In this set of counterfactuals, I could consider a transition path where a share of trucks convert to self-driving while the remainder retain their home preference.

In Figure 22, I plot average prices and total quantity for a variety of automation levels between 0 percent and 100 percent. The 0 percent counterfactual is the baseline, and the 100 percent counterfactual is the main counterfactual. I find a relatively smooth transition path from 0 percent automation to 100 percent automation.

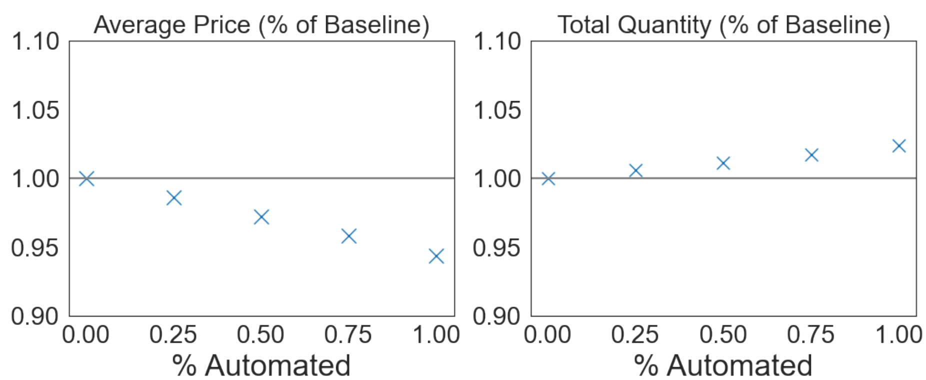

Figure 22: Prices and Quantities on Transition to Automation

Note: Figure plots average prices (Laspeyres Index) and total quantity under transitional counterfactuals where 0 percent, 25 percent, 50 percent, 75 percent, and 100 percent of existing trucks are replaced with self-driving trucks.

Counterfactual C.2: Geographic Transition to Automation

Geographic rollout of self-driving trucks may not be instant across the United States. For example, while highways are much more regular than city streets, snowy northern highways may be a more challenging automation problem than sunny southern highways. A Deloitte report projects that in a first stage, southwestern states from Oregon and California down through Texas may be feasible first. In a second stage, snowier states from Nevada east through Illinois to Virginia may become feasible. Finally, in the third stage automated trucks would be able to travel through any state. In Figure 23, I plot the geographic distribution of these stages.

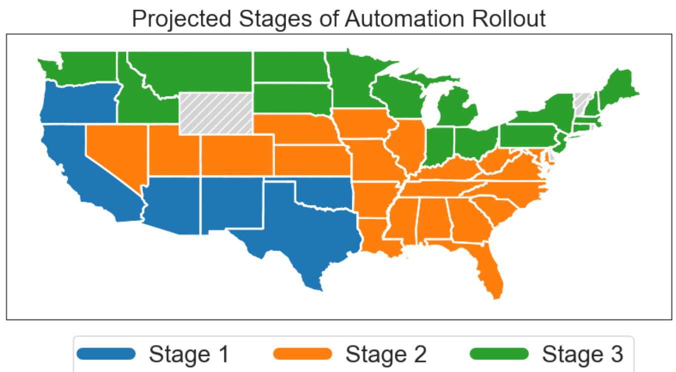

Figure 23: Rollout Stages of Self-Driving Trucks

Note: Figure plots the three stages of automation rollout identified in Zarif et al. (2021). The first stage covers southwestern states from Oregon to Texas. The second stage expands north and east. The third stage covers the northernmost U.S. states.

In this set of counterfactuals, I consider progressive rollout of self-driving trucks in stages.

In the first counterfactual, I convert 25 percent of trucks from Stage I states to self-driving, and I restrict self-driving trucks to only travel to Stage I states. In this counterfactual, self-driving trucks have no home preference, and also benefit from the 25 percent reduced per-mile costs and doubled daily range of the full counterfactual. In the Stage I counterfactual, I find that overall export prices increase everywhere as shown in Figure 24. Although self-driving trucks have lower costs and longer range, constraining some trucks to the Stage I states decreases overall efficiency. These two effects moderate each other for Stage I states such as California and Texas, which see minimal changes in export prices. Meanwhile, Midwestern states see the largest decline in carriers and increase in export prices.

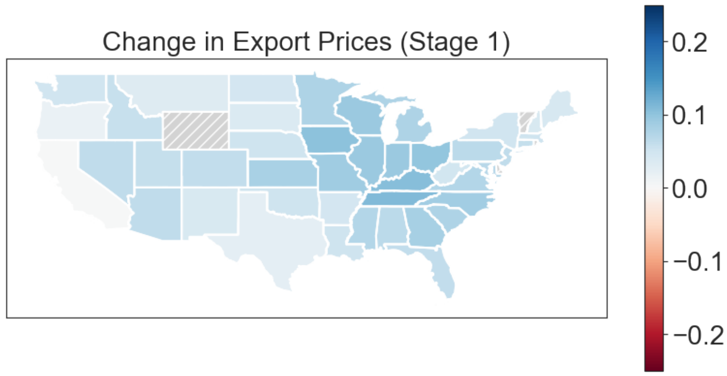

Figure 24: Counterfactual export prices under Stage I rollout

Note: In the Stage I rollout counterfactual, I convert 25 percent of trucks in Stage I states to self-driving and restrict self-driving trucks to Stage I states. Figure 16 plots the percentage change in the average trucking export price from each U.S. state, as defined by a Laspeyres index which weights prices by quantity in the baseline. States which see rising export prices are blue, while states which see falling export prices are red. Larger magnitudes are indicated by darker colours.

In the second counterfactual, I convert 25 percent of trucks from Stage I or Stage II states to self-driving, and I restrict self-driving trucks to only travel to Stage I or Stage II states. In the Stage II counterfactual, I find that export prices fall for Stage I and II states with access to self-driving trucks, but rise for Stage III states without access, as shown in Figure 25. Compared to the Stage I counterfactual, self-driving trucks are less constrained so overall efficiency is higher. However, Stage III states now lose significantly more carrier supply than in the Stage I counterfactual, since more trucks are now automated. This is particularly pronounced in the Midwest, where Illinois is a Stage II state but many of its main export destinations (Michigan, Ohio, Iowa, and Minnesota) are not. This figure suggests that short-run economic consequences of self-driving trucks may depend importantly on the geography of rollout.

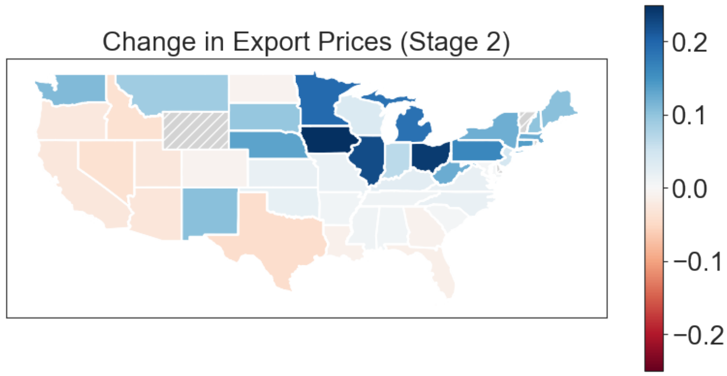

Figure 25: Counterfactual export prices under Stage II rollout

Note: In the Stage II rollout counterfactual, I convert 25 percent of trucks in Stage I and Stage II states to self-driving and restrict self-driving trucks to Stage I or Stage II states. Figure 25 plots the percentage change in the average trucking export price from each U.S. state, as defined by a Laspeyres index which weights prices by quantity in the baseline. States which see rising export prices are blue, while states which see falling export prices are red. Larger magnitudes are indicated by darker colours.

In the third counterfactual, I convert 25 percent of trucks from all states to self-driving, and I allow trucks unrestricted access to all states. Without the geographic restrictions, self-driving trucks efficiently distribute across the economy. Figure 26 shows that export prices fall for all states, and the pattern mirrors that of the "full" counterfactual.

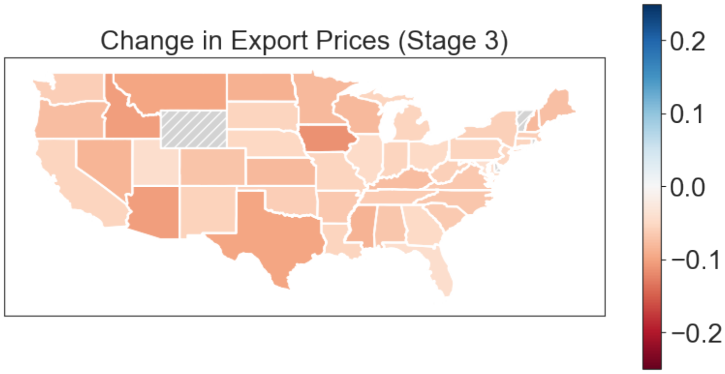

Figure 26: Counterfactual export prices under Stage III rollout

Note: In the Stage III rollout counterfactual, I convert 25 percent of trucks in all states to self-driving and do not restrict self-driving trucks. Figure 26 plots the percentage change in the average trucking export price from each U.S. state, as defined by a Laspeyres index which weights prices by quantity in the baseline. States which see rising export prices are blue, while states which see falling export prices are red. Larger magnitudes are indicated by darker colours.

Appendix D: Industry and Data Details

D.1 Industry Details

In this section, I provide a brief overview of the industry structure.[17] Since deregulation under the Motor Carrier Act of 1980, the trucking freight industry has been characterized by a decentralized market structure. According to Day and Hait (2019), 88 percent of the 396,000 long-distance trucking businesses in 2016 were self-employed truckers. The trucking market is further segmented by (1) freight type, (2) shipment size, (3) geographic scope, and (4) contracting structure. In this paper, I will focus on the general freight, over-the-road, truckload market.

Most truckers haul general freight: commodities that can be shipped in standardized van trailers. By contrast, specialized freight types require specific licenses and skills and include refrigerated trailers ("reefer"), flatbed trailers, tanker trailers for shipping oil, hazardous materials shipping, and specialized trailers for large machinery and other non-standard items. Due to trailer-type-specific skills investments and certifications, carriers often specialize in one trailer type.

Shippers may choose between several shipment sizes. Truckload (TL) services, which are the focus of this paper and the majority of shipments, sell the entire trailer capacity to one client. Less than truckload (LTL) services combine items from multiple clients in one trailer, yielding potentially lower fixed costs in exchange for lower quality service. Parcel services such as UPS and FedEx handle the smallest shipment sizes. Parcel and LTL firms are more labor-intensive than T.L. firms and rely on infrastructure for packing and re-packing trailers.

Shipments can also be segmented by the distance travelled, or "length of haul." Long-haul or "over the road" carriers specialize in shipments over 150-250 miles, and on average, return home once every three weeks. Regional carriers focus on shipments within 150-250 miles, which they can deliver and return home within a 500-mile daily

[17] For additional details, I refer readers to Burks et al. (2017) which provides a detailed industry primer covering the industry, market segments, and ongoing challenges and changes.

driving limit. Local carriers make deliveries of less than 150 miles, such as warehouse-to-final-destination trips. Finally, some consider the low end of the long-haul segment, 150-500 miles, the "tweener" category. Shipments in this region are undesirable for both carriers who want to do longer or shorter days.

The trucking market can be divided into a spot market, a contract market, and private fleets. On the spot market, shippers and carriers match either using word-of-mouth, electronic platforms such as DAT, or centralized matching algorithms. Large shippers and carriers may directly use the spot market, while smaller shippers tend to use brokers. On the contract market, shippers conduct RFP processes where carriers bid with rates. Given bids, shippers construct routing guides that establish a priority among bidding carriers for volume.[18] Finally, carriers may choose to retain drivers and truckers in private fleets.

Since carriers may serve either spot or contract markets, and shippers may flexibly adjust between TL, LTL, and private fleets, the relative market share of these markets is volatile. Holm (2020) estimates a 46 percent market share for full truckload, 10 percent market share for less than truckload, and 45 percent for private fleets. Within full truckload, up to 80 percent of quantity moves through the contract market. Williams and Murray (2020) reports that in 2019 surveys, 26 percent of trips were less than 100 miles, 39 percent of trips were between 100 and 500 miles, 22 percent of trips were between 500 and 1,000 miles, and 13 percent of trips were over 1,000 miles.

[18]Shippers keep the option of going to the spot market rather than to one of their contracted carriers, and carriers keep the option of going to the spot market rather than accepting a load. When spot prices are higher than contract prices, shippers use their contracted carriers more, and carriers are less likely to accept. The reverse holds when spot prices are lesser than contract prices. Recent papers such as Scott, Parker, and Craighead (2016), Aemireddy and Yuan (2019), Acocella, Caplice, and Sheffi (2020), and Harris and Nguyen (2021) investigate the interaction between the spot and contract markets. For my paper, this interaction means that transacted spot and contract prices tend to be very similar: in 2016, average spot and contract prices at the city-city level had a correlation of 98.04 percent, and average prices at the state-state level had a correlation of 99.09 percent.

D.2 Data Details

Since locations in highway inspections are observed at the state level, I aggregate my data to the state-state level and let L be the set of U.S. states[19].

For distance, I use the PC-Miler Effective Distance from i to j, an engineering estimate of the distance between origin and destination along truck-legal routes.

For diesel, I use a route-mileage-weighted average of state average diesel prices from AAA. I scale the dollar-per-gallon price by the distance, and I de-mean diesel prices so I can interpret my distance coefficient as an all-in marginal cost per mile.

For road quality, I filter road segments from the 2019 Highway Performance Monitoring System for interstate and principal arterial roads. For each segment, I observe cracking, faulting, and rutting as continuous measures[20]. I standardize these quality measures to have mean zero and standard deviation one, so that positive values indicate lower quality roads. I take compute state averages for each quality measure, and I weight states by the route mileage shares to compute route road quality measures.

I begin by finding the centroid of the origin and destination state and using OpenStreetMaps to compute the shortest route using U.S. highways. Let m_{ijk} be the share of miles on the shortest route from i to j spent in state k. For each route, I use the mileage shares to weight these state averages.

I measure quantity Q_{ij} using total daily trips in DAT RateView, and the share of shippers who choose trucking $s_{ij}^{shipper}$ using freight shipments in the 2017 Commodity Flow Survey (CFS). I use the total shipping across all modes in the Commodity Flow Survey as my measure of N_{ij}. Since the Commodity Flow Survey is reported in tons, I assume that 22 tons corresponds to a full truckload[21]. I use average 2019 prices in DAT RateView for p_{ij}.

[19] Due to discontinuity from the main U.S. highway system, I omit Alaska, Hawaii, and Puerto Rico from my analysis. In addition, due to data limitations in DAT RateView, I also omit Delaware, Rhode Island, and Vermont. Due to data limitations in DAT Trucks in Market, I omit Wyoming from analyses which depend on Trucks in Market.

[20] Cracking measures the percentage of road area which exhibits fissures and discontinuities. Faulting measures the average vertical misalignment of pavement joints. Rutting measure the average depth of surface depressions.

[21] Staff (2020) reports average payload for a dry van at 44,000 to 45,000 pounds

I use inspections and registrations from the FMCSA. This dataset co-mingles carriers who are involved in freight transportation with other carriers. I filter these for the set of carriers that are (1) engaged in inter-state freight, (2) not engaged in passenger carriage, and (3) not engaged in hazardous materials carriage. Registration address may not be reflective of home address for large carriers. As a result, I drop all carriers which report more than X power units.