In it for the longhaul: Occupational choice for blue-collar workers

Levi Soborowicz

January 20, 2024

UPDATED

Abstract

In 2021 the American Trucking Association claimed the driver shortage reached about 80,000 drivers and predicted a shortage of 160,000 drivers by 2030. The perceived driver shortage is a major concern of both policymakers and firms. The 2021 Infrastructure bill authorized a trial program to allow 18–20-year-olds to drive commercial vehicles. In this paper, I use a Discrete Choice Mixed Logit model to estimate the impact of wage increases, education cost reductions, and the impact of opening truck driving to 18–20-year-olds on the share of truck driving employment in the blue-collar labor market. I use a control function approach to deal with endogeneity in wages and training costs. I use four data sets, the Survey of Consumer Finance March Supplement (ASEC), for individual and market share data. State-specific occupational wage data was sourced from the Occupational Employment Wage Survey (OEWS), Work setting variables were sourced from O*Net, and educational costs from the Integrated Postsecondary Education Data System (IPEDS). Results show that higher wages, reduced training costs, and including 18-20-year-olds increase the number of truck drivers. The largest potential increases come from adding 18–20-year-olds into the market, increasing trucking employment by about 70,000 new drivers nationally.

JEL Codes: L91, J24, J44

Keywords: Truck Driving, Driver Shortage, Mixed Logit

Introduction

Truck driving is a large profession in the United States. About 3.5 million people are employed as truck drivers in any given year. The 2017 Commodity Flow Survey, a representative sample of freight in the US estimates about 65% of all tonnage or 70% of all freight value in the United States was transported by truck only.[1] This figure excludes freight moved using multiple modes, for example, truck to rail. The prices both producers and consumers pay for shipping depend on the input costs of transportation services. Among them are fuel and wage costs. From the perspective of firms buying transportation services, they would like the lowest input costs to ship their freight. And, of course, the drivers are interested in getting the highest price for their services.

A shortage of drivers is a perennial concern of firms in the trucking industry. In 2021 the American Trucking Association claimed the driver shortage reached about 80,000 drivers and predicted a shortage of 160,000 drivers by 2030.[2] The perceived driver shortage is a major concern of both policymakers and firms. However, it is unclear if there is a current shortage of truck drivers or if there will be one in the future due to an aging workforce. The press has fixated on the concept of a shortage; a Google search yields several articles from generally reliable sources such as the New York Times, NPR, CNBC, and Time Magazine discussing the causes and cures of a truck driver shortage. This widespread concern about a shortage led the Biden administration to push for several programs in the 2021 Bipartisan Infrastructure Bill to increase the number of truck drivers. One of the policies is a trial apprenticeship program to allow 18–to 20-year-olds to drive commercial vehicles. Other important policies include expediting the CDL process across states, recruiting veterans, and other policies addressing the working conditions of truck drivers.

As economists, the notion of a shortage in this context implies that the number of drivers supplied in the market at a particular wage rate is insufficient to meet the quantity of drivers demanded. However, as I show later, workers respond to wage increases by shifting into the occupation with the wage increase. The non-economist definition of shortage, I believe, is most useful in this context. That is, the number of drivers supplied in the market does not meet the number of drivers necessary to cover the total logistics load at the equilibrium price. Logistics load, in this context, is the number of loads that need to be carried at the quickest rate possible given the

---

1. Commodity Flow Survey 2017: https://www.census.gov/programs-surveys/cfs.html
2. American Trucking Association: Report

transportation network. Consider the following example: in a classic model of supply and demand, there are potential consumers of the product that are unwilling to purchase the product at the equilibrium price. In the same way, there are firms that would be willing to ship their product if the price were lower, but drivers are unwilling to take loads at that rate. Thus, some portion of the total logistic load goes unserviced, but the equilibrium is a market outcome without a shortage in the economic sense.

Presuming that the current number of drivers is insufficient, there are two margins we can modify to increase the labor supply of drivers. The first is the intensive margin; in principle, we can increase the amount of loads serviced by increasing the number of hours a driver is willing to work. Even though I do not consider any of these policies in the paper since I am singularly focused on the extensive margin, it is worthwhile to discuss some of them here. Currently, drivers are regulated by the number of hours they are allowed to be on duty. This policy is in place for obvious safety reasons, but if there is a mass of drivers bound by this constraint, relaxing it would increase the amount of labor supplied. Alternatively, there has been some discussion of truck parking, and this is related to the time constraint. Parking is a scarce resource and is distributed in discrete intervals. And drivers are constrained by the number of hours they are allowed to drive. If a driver is faced with a choice such as having 4 hours left of allowable hours and the nearest truck stop is immediately available, or 4 hours away and not guaranteed to have an open slot, they must choose whether to stop or risk it. In particular, the drivers will be concerned about the likelihood the next truck stop is at capacity because if it is, they will be stuck on an exit without facilities. Private solutions are now available services, such as Truck Parking Club[3]. These services link landowners with drivers to provide parking in safe locations, mitigating the capacity constraint at truck stops. The Bipartisan Infrastructure Bill allocated funds to increase truck parking in certain states[4]. Improving technology and reducing wait times for drivers, is another way to increase the amount of freight they can carry. The driver's time on the road can be used up by dwell time, the amount of time stuck waiting for freight to be loaded or unloaded. Reducing dwell times is one way to increase the amount of freight drivers are able to carry.

The extensive margin increases the number of drivers in the market, and this is where I put my

---

3. Truck Parking Club Website: https://truckparkingclub.com/
4. Announcement: link

focus in the analysis. I include four counterfactuals that address different policy proposals and a fifth that considers the impact of an aging workforce on the share of drivers. The first three our pecuniary, and they are a wage increase for all potential and incumbent truck drivers, a sign bonus for only new entrants, and thirdly a reduction in training cost of CDL programs. The fourth is allowing 18-20-year-olds to choose truck driving as a profession. Finally, the final counterfactual considers the impact of an aging workforce and what happens to the market share of truck drivers in the blue-collar labor force as the workforce ages.

In this paper, I use a mixed logit model to estimate the impact of wage increases, education cost reductions, an aging labor market, and opening truck driving to 18–20-year-olds on the share of truck driving employment in the blue-collar labor market. I use a control function approach to deal with endogeneity in wages. I use four data sets, the Survey of Consumer Finance March Supplement (ASEC), for individual and market share data. State-specific occupational wage data was sourced from the Occupational Employment Wage Survey (OEWS), Work setting variables were sourced from O*Net, and educational costs from the Integrated Postsecondary Education Data System (IPEDS). Results show that higher wages, reduced training costs, and including 18-20-year-olds increase the number of truck drivers. The largest potential increases come from adding 18–20-year-olds into the market, increasing trucking employment by about 70,000 new drivers nationally.

Literature

Discrete choice models have a long history and many applications. Introduced by McFadden (1973), the conditional logit model proved to be a tractable and empirically useful way to model many different choices. He demonstrated the power of the logit choice model in the context of consumers' shopping decisions. In this paper, I use a product demand model applied to the context of occupational choice. Since I have individual-level data available I model the individual choices in the style of Goolsbee and Petrin (2004). They model a household's choice of television provider using a mixed logit model to compute the welfare gains from the addition of satellite television providers. The convenience of the approach allows the researcher to use individual-level data and aggregate the results to market outcomes. Alternatively, it is possible to estimate the model using aggregate market outcomes with generalized method of moments, Berry, Levinsohn, and Pakes

(1995), Nevo (2001), and Petrin (2002).

Boskin (1974) was one of the earliest papers to model occupational choice as a discrete choice problem. Comparing 11 occupations in the context of a conditional logit model, individuals choose their occupation based on the present value of future earnings in each occupation. Other models have focused on the dynamics of work, education and leisure choices of workers. Keane and Wolpin (1997) use a dynamic discrete choice model to capture the human capital investment and occupational choice of young men. Fuller, Manski, and Wise (1982) model the college choice of students as a function of both the student's characteristics and the economic costs of the school.

Phares and Balthrop (2022) use a mixed logit model to estimate the labor supply of truck drivers in industry occupation pairs. Their methodology allows for rich substitution patterns between occupations, but they restrict the choice set to five possible alternatives, including truck driving. Importantly, they compute the wage elasticity for truck drivers. They show a 1% increase in the wage rate corresponds to a .3% increase in the number of drivers. This result suggests that drivers do respond to wage increases. Other work by Richards and Rutledge (2022) uses a dynamic search model to estimate the impact of the COVID-19 pandemic on truck rates and labor supply. Their results suggest that the trucker market is in disequilibrium since the creation of new jobs exceeds the rate at which the jobs are filled.

Data

The data spans the period from 2010 to 2019. It is a combination of four public data sets the Current Population Survey ASEC March Supplement (CPS), Occupational Employment Wage Statistics (OEWS), Integrated Post Secondary Data System (IPEDS), and O*Net Occupational Characteristics. The CPS records economic and job data for a sample of American households. The CPS data was downloaded through Integrated Public Use Microdata Series (IPUMS) from the University of Minnesota (Flood et al 2023). The CPS allows me to observe the previous and current occupations of a worker, because I observe the worker twice one year apart. O*Net records hundreds of job characteristics for most occupations. OEWS is the source of all employment and wage data used in the analysis. Finally, IPEDS is the source of truck driver training programs by state and year.

The data contains a total of 69 occupations; 47 of the 69 occupations are blue-collar or white-collar with minimal education requirements. I excluded 22 occupations that are either educated white-collar, agriculture occupations, and occupations that have extremely small sample sizes. This restriction removes about one-third of the CPS sample. An exhaustive list of excluded occupations can be found in the appendix. However, some examples listed here are Advance Computer and Math Workers, Advanced Financial Workers, Advanced Medical Workers, Engineers and Architects, and Education Workers. These occupations tend to require advanced degrees such as an M.D. or Masters. I exclude these occupations for two reasons. The first is that my model is insufficient to explain the educational decision-making process of a worker pursuing an advanced degree. Dynamic discrete choice models exist that can capture the human capital accumulation versus work decision of workers Keane and Wolpin (1997). However, since I only observe the person twice, I cannot track them over time and estimate the underlying utility function of a worker. For example, a bachelor's degree can take 3-5 years, and more advanced degrees require more time. Most occupations in my sample take less than one year to train for. These educated occupations are too complicated, given my data set. Secondly, the overlaps between my main occupation of interest and these higher education occupations are relatively small. About 80% of all educated white-collar workers in my sample are either not working or members of the educated white-collar category. The other 20% of educated white-collar occupations are distributed among the other 50 occupations, including agriculture. The other 68 occupations also contribute a small share of workers to the total share of educated white-collar workers.

The occupations consist of two classification schemes: major and minor occupations. The minor occupation codes are broadly based on the Standard Occupational Classification System (SOC). However, each dataset I use uses a slightly different coding system. Both OEWS and O*Net use the Standard Occupational Classification System (SOC) code. O*NET uses the more modern 2019 codes and OEWS uses the 2018 (SOC) codes. Furthermore, the CPS uses a separate system based on the 2010 (SOC) Codes. I created my own occupational coding scheme to harmonize all these occupational systems. And reduce the dimensionality from 997 unique occupations to just 69. This process resulted in the 47 included and the 22 not included occupations mentioned earlier. The major occupation codes consist of 9 broad occupational classes based on the 69 occupations. I use 7 of these codes: white-collar worker, tradesman, manual laborer, service worker, production worker,

transportation worker, and sales worker.

The CPS ASEC March supplement is a survey that asks detailed questions about the workforce. Each person is sampled twice, one year apart, and if a person switches occupations between samples, the survey will reflect that change in occupation. Within a given sample year, half of the sample is incoming, and the other half is outgoing; I treat the observation year as the outgoing year. The incoming sample is used only to record their occupation, and that is my measure of the previous occupation, which is not a choice variable for the worker. A worker is endowed with their previous occupation and maximizes utility conditional on that occupation. The outgoing occupation is the dependent variable, the workers choice of occupation. This takes the repeated panel structure and converts it to a repeated cross-section, where I observe a new set of individuals each year, but I have information on their current and previous occupations in that year. The CPS survey in the cross-sectional format with the updated occupation codes is the main source of data for the analysis. Aside from measuring occupational switching, the CPS data contains multiple demographic characteristics of each worker in particular age and sex. I use age and sex as my main demographics in the analysis since two of the major concerns in the trucking industry are the aging workforce and the heavy male distribution. Recruiting more women is a potential source of new drivers, and understanding how women respond to truck driving's observables can shed light on the effectiveness of recruiting women into truck driving.

Most individuals in the sample do not move occupations (72.5% average retention) Table 1 shows the top and bottom ten occupations by retention in the sample. There is considerable heterogeneity across occupations in the retention rates, ranging from an average high of 80 % and a low of 18 %. Interestingly, truck drivers have one of the highest retention rates of all occupations in the sample. This suggests a policy that increases the number of truck drivers likely has to operate through recruiting new workers rather than retaining the current workers. Hazmat workers have the lowest retention rate in the sample; given the danger of the occupation, this is not necessarily surprising.

O*Net data consists of 877 occupations and corresponding characteristics in six general categories called the Content Model. The six categories are worker characteristics, worker requirements, experience requirements, occupation requirements, occupation-specific information, and occupation characteristics. I use a subset of occupational requirements, denoted as work context. The work

Retention Rate (H-L)	Occupation	Retention Rate (L-H)	Occupation
0.80	Emergency Service/Criminal Justice	0.18	Hazmat Worker
0.74	Personal Appearance Workers	0.28	Information Clerks
0.68	Housekeeping/Janitorial Worker	0.35	Repair Worker
0.67	Bus and Taxi Drivers	0.35	Forestry Worker
0.67	Truck Drivers	0.36	Unemployed
0.65	Electrician	0.38	Gaming Service Worker
0.65	Courier	0.39	Laundry Workers
0.64	Rail Workers	0.40	Industrial Repair Worker
0.64	Food Preparation Worker	0.41	Advanced Electronics/Precision Installer
0.62	Funeral Workers	0.41	Material Movers

*Average retention rate by occupation in the sample. Calculated using the CPS by computing the number of people in the sample that are no longer incumbent in their previous occupation when they are observed in the outgoing rotation.

Table 1: Retention rate by Occupation

context category is defined as the physical and social factors that influence the nature of work. These features describe the working conditions and pressures of an occupation. The characteristics are unchanging over time, and firms have very little ability to change their occupation characteristics. The easiest way to conceptualize these characteristics is to think of them as product characteristics in product choice models. For example, feature of products such as mpg, color, flavor, etc. The eight variables I use to distinguish work conditions are the consequence of error, contact with others, degree of automation, exposed to hazardous conditions, in an enclosed vehicle or equipment, outdoors exposed to weather, time spent using your hands, and time pressure. These variables are time constant and measure the characteristics of a particular occupation. Since I observe the current and previous occupations, I calculate the Euclidean distance in characteristics between the two occupations. The more similar an occupation the closer they are in distance, and the more dissimilar occupations are the larger the measure of distance.

OEWS is the source of market wages and employment. I use the median and the 25th percentile full-time wage income as the measure of the expected compensation. All market wages are converted to 2019 dollars using the Consumer Price Index (CPI). I do not distinguish between part-time and full-time workers since I am modeling the choice of occupation, not the number of hours a worker chooses to work. To capture the differences in wages for incumbent and new workers, I use different measures of occupation wages. For incumbent workers, I use the median full-time annual income as the measure of market wages when they consider their incumbent occupation. However, if a worker considers an occupation that they do not currently occupy, they would observe the 25th percentile of annual income as their expected annual income.

IPEDS is the source of education training data for CDL programs in the United States. While all publicly funded institutions are required to submit their data to IPEDS, many training programs are missing from the data. For example, in 2019 10 states are missing from the dataset.[5]. In any given year for the states with at least one truck driving training program, I compute the mean of the training cost by averaging the observed costs of the training program for each school observed within the state and year. I use the average training cost within each state as the observed training cost in that state.

The final dataset consists of 48 states over nine years, excluding Alaska and Hawaii. There are three categories of variables. The individual-level data from the CPS varies across individuals and includes age, sex, and previous occupations. The State - Time variables that vary across states and time include expected wage and training cost. Finally, the occupation characteristics only vary across occupations.

The summary statistics are below in Table 2.

| | Female | | Age | | Retention Rate | | Distance | | Earnings | | Training |
	Sample	Trucker	Sample	Trucker	Sample	Trucker	Sample	Trucker	25th Pctl	Median	CDL
Mean	0.40	0.04	46.5	49.1	0.61	0.68	2.09	2.87	$36,912.90	$46,611.33	$4,537.27
Std Dev	0.49	0.20	13.4	12.4	0.49	0.47	0.97	1.10	$15,022.12	$20,135.72	$2,959.29
25th Pctl	0	0	36	40	0	0	1.27	2.04	$25,496.75	$31,294.78	$2,461.42
50th Pctl	0	0	47	50	1	1	1.98	2.54	$33,569.24	$42,430.56	$3,588.63
75th Pctl	1	0	57	58	1	1	2.73	3.84	$45,815.06	$58,535.18	$5,364.65
Obs	95,496	3,966	95,496	3,966	95,496	3,966	37,663	1,264	35,744	35,744	380

*Summary statistics computed from the four data sets. Female, Age, Retention Rate and Distance are all calculated from the CPS. Earnings is computed from the OEWS and Training cost comes from IPEDS

Table 2: Summary Stats

Table 2 contains the various demographics of the entire sample and truck drivers, as well as the earnings distribution over occupations. There are 103,399 total individuals in the sample, of which 4,061 are truck drivers. About 3.9% of the sample consists of truck drivers, and the other 96.1% of workers are employed in other occupations. No workers in the sample are unemployed. The age, percent female, and retention rate variables are calculated for the entire sample and truck drivers. The distance columns were calculated only for movers. Since all non-movers all move 0 distance, and about 78% of the sample stay in their previous occupation, computing the summary statistics with the non-movers is less informative. The statistics about how far movers move are more

---

5. The missing states are: Colorado, Connecticut, Louisiana, Maine, Massachusetts, Nevada, New Jersey, Rhode Island, Vermont, and Wyoming

informative because they give some measure of where workers go when they leave their incumbent occupation. The annual earnings statistics are calculated from the state and year variation in real annual earnings across all occupations. Each incumbent worker will see the median annual earnings as their predicted wage for staying in their current occupation, and they will observe the 25th percentile for all other occupations.

The driver sample is both 2.6 years older on average and significantly more male than the average worker. This is a well-documented feature of the truck driving industry; many are concerned about an aging male-dominated workforce. The average age of truck drivers is 49.1, slightly higher than the sample average of 46.5. However, this difference is much less pronounced than the ratio between men and women in the sample. The share of women in truck driving is very small; only about 4% of the trucker sample is female compared to about 40% in the whole sample. The sample, on average, skewed more male, and truck driving is even more skewed. Truck driving has a higher retention rate than the sample average. If we consider this information in the context of a presumed worker shortage, the relatively high retention rate within the occupation is surprising but does not necessarily contraindicate a shortage.

Annual earnings are the real annual earnings converted into 2021 dollars, although the sample ends in March 2020. Each occupation is observed in the OEWS by state and year. To correct for the role of experience and tenure in an occupation, I make the assumption that new entrants are truly new. So, for example, using the numbers from the mean occupation, an incumbent worker would observe annual predicted earnings of about $47,000 if they were to stay in this fictional mean occupation. A potential entrant would observe $37,000 because they are new to the occupation. The occupations in the sample are not exceptionally high-paying occupations; About 75% of occupations have a median earning below $59,000.

The CDL training program costs are observed at the state level across multiple years, although due to data and computational limitations, I only use 2019 to estimate the model at the national level, using cross-sectional variation. This is for two pragmatic reasons. The first reason is due to my use of Google Colab to speed up the estimation process. The model takes time to converge, and using Google Colab helps speed up the process, but they have data limits on allowed RAM, which the full data set exceeds. The second is the MLE process struggles with the unbalanced cross-section since I observe some states in some periods but not others. The median program
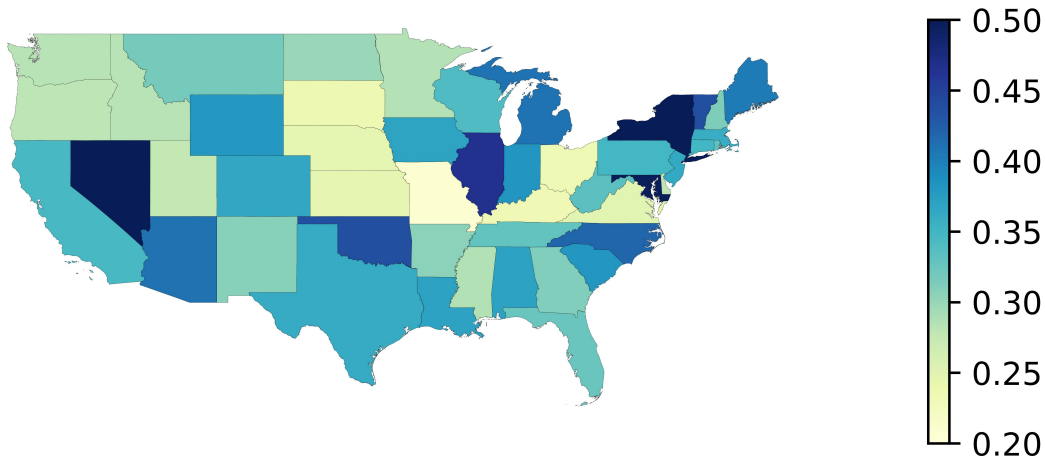
Figure 1: Share of New Truck Drivers by State

*The share of new drivers by a value of 1 would indicate all drivers in the sample in a given state are new. This figure uses all sample years to compute the flow.

costs about $3,600 to complete, with a somewhat large standard deviation across states.

Figure 1 shows the geographic distribution of new truck drivers. The figure is a choropleth showing the share of new drivers within that state. The dark states are states where the largest share of the current truck drivers are recent entrants, and the lighter states are states with fewer entrants. A value of 1 in West Virginia for example, would indicate all drivers in the sample from West Virginia are new entrants. New York, Nevada, Illinois, and Maryland have the largest share of new entrants to incumbent drivers. Meaning these states have the highest rate of entry relative to other states.

Figure 2 shows the geographic distribution of former truck drivers. It is the exit rate by state in the truck driving occupation. Nevada, Maine, North Carolina, Louisiana, Vermont, Boston, Michigan and Wyoming all have the lowest retention rates. For example, a value of .3 means 30% of past truck drivers are not no longer truck drivers. Pairing Figure 1 and Figure 2 we can see that Nevada, New York, and to some degree North Carolina have a large degree of churning in labor market for truck drivers. Within these states, we have a high share of new entrants and a high share of people leaving the occupations. In contrast Illinois an Maryland appear to be a net gainers, having a large share of new drivers, but a low share of exiters relative to the size of their increases. Overall, there appears to be no clear geographic pattern, other than a high degree of churning in some states like Nevada and New York.

11

Figure 2: Share of Former Truck Drivers by State
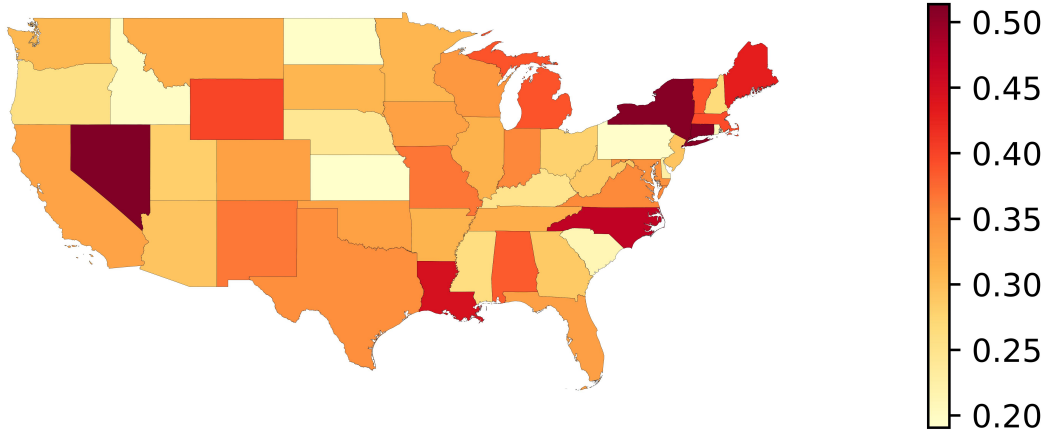
Model

Workers

I model the number of workers in an occupation as the outcome of many individual decisions about which occupation to occupy using a mixed logit model. The mixed logit approach allows more flexibility than conditional or multinomial logit alternatives or modeling the market shares of each occupation directly. Since individual-level data is available, I use individual decisions and aggregate them up to market outcomes. In this case, the mixed logit is particularly attractive because of the heterogeneity in preferences across workers, and the concern about an aging workforce and the potential to recruit younger workers. The mixed logit approach allows me to capture the heterogeneity in preferences of workers in blue-collar occupations, and capturing their preferences in a more sophisticated way allows for complex substitution patterns across occupations. Specifically, my modeling approach allows me to compute several counterfactuals, five of which I report in this paper.

A worker's choice probability of a given occupation depends on the characteristics of the past and current occupations, wages, education cost, worker demographics, and an idiosyncratic Gumbel taste shock. Workers take their past occupation as given and choose among 68 possible occupations. A worker observes the wage rate they would earn w_{js} if they select occupation j. If they are incumbent in occupation j, they observe the median wage rate of occupation j; otherwise, they observe the 25th percentile. The distance between occupations is calculated as the Euclidean distance

between occupations in the characteristics space. Large values of $||x_j, x_k||$ indicate very different occupations and smaller values indicate more similar occupations. And naturally, when the $j = k$ the measure of distance is zero. The training costs are only observed for people considering truck driving as an occupation but are not already an incumbent driver. The training cost is measured as the average cost of a CDL training program in the driver's observed state. I consider two separate utility specifications, one nationally, which includes truck driving training program costs as a component of an individual's choice probability. The other utility specification is estimated state by state, allowing for more heterogeneity and variation in preferences by state. There is a trade-off between these two specifications. The state-by-state specification does not allow for the inclusion of the cost of truck driving training programs do to issues with the training cost variable. However, it allows for more flexibility. The national-level utility function is specified as follows:

$$u_{ijk} = (w_{js} - \text{FC}_s)\alpha_i + ||x_j, x_k||\beta_i + \xi_m + \varepsilon_{ijk} \tag{1}$$

Since the training cost and annual salary are pecuniary I deduct the cost of a truck driving training program from the expected annual salary of a new entrant. The new variable is the annual total compensation for choosing truck driving as a profession. All occupations other than truck driving don't have an observed training cost so the total compensation is just the annual wage rate. The training cost is an unobservable for all occupations other than truck driving, I will discuss the control function approach in a later section.

The second model, which I use for three of the four counterfactuals, allows richer heterogeneity across states, indicated by subscript s. The utility of job j for a worker i given their previous occupation j and state s:

$$u_{ijks} = w_{js}\alpha_{is} + ||x_j, x_k||\beta_{is} + \xi_{ms} + \varepsilon_{ijks} \tag{2}$$

Where
$$\begin{pmatrix} \alpha_{is} \\ \beta_{is} \end{pmatrix} = \begin{pmatrix} \alpha_s \\ \beta_s \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \beta_{Fs} & \beta_{As} \end{pmatrix} \begin{pmatrix} \mathbf{1}(Female = True) \\ Age \end{pmatrix} + N(0, \Sigma_s)$$

Both the national and state-by-state models use this random coefficient structure. The only difference between the state-by-state and national model, is the state specific standard deviation, and

coefficients.

The wage response of worker i depends on the mean wage response in state s and individual specific draw a normal distribution with mean zero, and a standard deviation σ_s specific to state s. I capture all worker-level heterogeneity of wage responses within a state with the estimated standard deviation of the normal distribution. Workers' preferences on distance are more complicated; the preferences depend on a normal draw and the age and sex of a worker. β_{is} can be thought of as the disutility of adapting to a new occupation. For example, if a worker chooses to stay in occupation k, they would experience no disutility from moving. However, if a worker chooses to switch occupations, they would experience disutility in accordance with their value of β_{is}. Presumably, workers vary in their tastes for differences in occupations; it is plausible some people even experience positive utility from differences in occupations. Allowing the model to capture the heterogeneity of people's taste for differences is important. In particular, if women and aging populations are loath to move occupations, this will have important consequences as the workforce ages, and how effective recruiting women into truck driving will be.

The unobservable occupation characteristics are estimated at the major occupation code level. These are the characteristics not controlled for by the wages and differences in occupations. The ξ_{ms} for major code m is the mean utility for choosing an occupation within the major code m. ε_{ijks} is the unobservable factors that influence an individual's occupational choice. I assume ε_{ijks} follows the Gumbel distribution and is independently distributed across choices. I will discuss the unobservables contained in ε_{ijks} later.

The probability a worker i chooses an occupation j is:

$$L_{ij} = \frac{\exp\left(w_{js}\alpha_{is} + ||x_j, x_k||\beta_{is} + \xi_m\right)}{\sum_k^K \exp\left(w_{js}\alpha_{is} + ||x_j, x_k||\beta_{is} + \xi_m\right)} \tag{3}$$

Depending on the draws and characteristics of the worker, the probability a worker chooses a particular occupation changes. Across all individuals in the labor market, the observed share of workers in an occupation is σ_j:

$$\sigma_j = \int L_{ij} dF(\alpha_{is}, \beta_{is}|Age, Sex, \Sigma) \tag{4}$$

The aggregate share of workers in occupation j is the weighted sum of the choice probabilities using the probability density function of the parameters as weights. Using these models I consider four counterfactuals to increase the number of truck drivers: a wage increase for all truck drivers, a sign bonus for new truck drivers, allowing workers aged 18-20 to drive trucks, and a reduction in training costs. I compute a fifth counterfactual to investigate the impact of an aging workforce, and whether there will be a decline in the number of truck drivers as the workforce ages.

The wage and training cost elasticities are specified as follows:

$$e_{\sigma_j,\ell} = -\frac{w_\ell}{\sigma_j} \cdot \int \alpha_i L_{ij} L_{i\ell} \, dF(\alpha_{is}, \beta_{is}|Age, Sex, \Sigma) \tag{5}$$

$$e_{\sigma_j,j} = \frac{w_j}{\sigma_j} \cdot \int \alpha_i L_{ij} \cdot (1 - L_{ij})d \, F(\alpha_{is}, \beta_{is}|Age, Sex, \Sigma) \tag{6}$$

$$e_{\sigma_j,\text{FC}_\ell} = \frac{\text{FC}_\ell}{\sigma_j} \cdot \int \alpha_i L_j(\beta)L_\ell(\beta) \, F(\alpha_{is}, \beta_{is}|Age, Sex, \Sigma) \tag{7}$$

$$e_{\sigma_j,\text{FC}_j} = -\frac{\text{FC}_j}{\sigma_j} \cdot \int \alpha_i L_j(\beta) \cdot (1 - L_j(\beta)) \, F(\alpha_{is}, \beta_{is}|Age, Sex, \Sigma) \tag{8}$$

Control Function

I am assuming that firms have oligopsony power when hiring workers. This varies across occupations and states; some firms can exert more market power than others depending on their industry and geographic unit. For example, 96% of trucking firms operate fewer than 10 trucks, so there are large numbers of buyers for any individual trucker's labor.[6] The market power any individual firm can exert on hiring a commercial driver is lower relative to an occupation with a more centralized market. Since there is a spectrum of markets encompassing a variety of intensities in competition between employers. And firms exert their market power as downward pressure on wages. The levels of markdowns will vary across states and occupations. To estimate the impact of wages on occupational choice, I need to control for the state-specific markdowns as a source of endogeneity.

Firms are aware of their occupation's popularity among workers and leverage that knowledge as a downward pressure on wages. Firms observe all features of their occupation, both those that are observable to a researcher through O*Net and those characteristics that are not observable. Since firms have information about unobserved job characteristics, firms use this information to

---

6. American Trucking Association Industry Data: https://www.trucking.org/economics-and-industry-data

lower wages relative to less popular occupations. Explicitly, I model observed market wages as $w_{js} = \text{mp}_j + \gamma_{js}$, the average marginal productivity as a constant and a markdown γ_{js} based on the market power and unobserved occupation characteristics.

Since I estimate the model using maximum likelihood, I use a control function approach with two sets of instruments to control for endogeneity. The first instrument is a Hausman instrument. I use the mean annual wage rate within each occupation of each bordering state as the instrument. The features that are required for this instrument work are relatively straightforward. The marginal productivities of workers need to be correlated across states, and the state-specific markdowns on unobservables are uncorrelated across states. Consider the case of unobservable training costs since this is a state-specific unobservable; training cost is unobserved to the researcher for all non-truck driving occupations. However, firms observe these costs and account for the training costs when marking down the wage rate. If the relationship between training programs and markdowns are uncorrelated between states, then the Hausman instrument should recover a consistent estimate of the coefficient on total compensation. Since truck drivers, plumbers, and other licensed trades in the sample are licensed at the state level it is more likely that the markets for training programs within states are uncorrelated with outside states. Formally I am assuming all sources of market power between states s, s' satisfy:

$$E(\gamma_{js}\gamma_{js'}|x_j) = 0$$

The second set of instruments are the observed occupation characteristics. These characteristics are assumed to be stable in the sample period, that is firms take the characteristics as given and they are constant across all states. These characteristics are relevant both to the decision of the firm in determining the wage rate and the decision of the worker in choosing that occupation. Controlling for these characteristics is important because all firms observe these characteristics. This causes correlation in the markdowns across states since every firm is using the same information. I use both the Hausman instrument and observed characteristics as the instruments for the endogenous component of wages. Explicitly I model the source of endogeneity as:

$$u_{ijks} = w_{js}\alpha_{is} + ||x_j, x_k||\beta_{is} + \xi_m + \varepsilon^1_{ijks} + \varepsilon^2_{ijks} \tag{9}$$

16

Where ε^1_{ijks} is the additively separable component of the error term correlated with wages. Let z_{js} be the Hausman instrument and x_j be the observed product characteristics. By using $w_{js} = \text{mp}_j + \lambda_1 z_{js} + \lambda_2 x_j + \gamma^1_{js}$ to model wages, the new error term γ^1_{js} is unccorrelated with the endogenous component of wages. The utility model corrected for the endogenous component of wages then becomes:

$$u_{ijks} = w_{js}\alpha_{is} + ||x_j, x_k||\beta_{is} + \xi_m + \lambda\gamma^1_{js} + \varepsilon^2_{ijks} \tag{10}$$

Both the national and state-by-state models are estimated using the Hausman instrument.

Estimation

The two models are estimated using simulation-assisted maximum-likelihood, in Python using the package Xlogit Arteaga et al (2022). The national regression is estimated using 38 of the 48 states in a cross-section. The second model is estimated state by state. I chose to estimate the two models separately due to the following tradeoff. I can use a GPU-assisted estimation using Google Colab, but that requires reducing the sample size to fit the data within the memory constraints Google enforces. GPU-assisted estimation is attractive because it speeds up the estimation process and allows more draws to compute the random coefficients, leading to more accuracy in computing the random coefficients. However, with this approach, I lose the efficiency gained from joint estimation. While in principle, I could estimate the model jointly without GPU assistance, the downside of this approach is my software is not capable of handling a large number of observations and choices. So, I am forced to either estimate the model as a cross-section or a state-by-state repeated cross-section.

The first model is estimated as a cross-section entirely because of data constraints. The IPEDs data has considerable inconsistencies year over year in its reporting of CDL training programs. Because of this, I can't get a consistent sample within states across time. So, I use 2019 the most recent year, and estimate the model in cross-section.

Using the state-by-state approach, I can leverage my 9 years of data, and allow each state to differ in their coefficients. I do not have the same limitations with data consistency using OEWS and the CPS only.

Results

National Response

The regression results from Equation (10) at the national level are below. This model is estimated for 38 out of the 48 states in the sample since I don't observe the training cost data for 10 states in 2019. The data consists of a single cross-section of 38 states in 2019. There are 8,347 workers in the model, and they choose among 42 occupation, based on their initial occupation and observed wage rates.

	Coef	Std Error	z-Statistic
Total Compensation	0.0100	0.0000	302.06
Distance	-1.5404	0.0019	-812.82
Distance × Female	-0.3798	0.0011	-334.04
Distance × Age	-0.0128	0.0000	-332.50
Control Function	-0.0004	0.0000	-13.45
ξ_2	-1.0925	0.0015	-734.28
ξ_3	-0.1449	0.0017	-86.70
ξ_4	0.1877	0.0013	142.87
ξ_5	0.3627	0.0017	212.79
ξ_6	0.1609	0.0013	126.43
ξ_7	1.1010	0.0012	926.75
SD Total Compensation	2.92×10^{-5}	0.0001	-0.43
SD Distance	1.6129	0.0008	2072.32
Obs	8,347		
likelihood	-37,095,913		
First Stage F-stat	15,750		

*National cross-section using 2019 with 38/48 states in the sample

Table 3: National cross-section estimates

The results show that workers are more likely to choose higher-paying occupations ceteris-paribus. More interestingly, the preferences for distance are considerably heterogeneous. On average workers prefer occupations that are more similar to their previous occupations. This preference is larger as workers age and for women. However, not all workers are averse to occupational differences. For example, taking a male with the average age in the sample (46.5), the mean response is a disutility of about 2.1 for a unit increase in distance between their current and past occupation. However, about 10% of the population of men age 46.5, will get positive utility for

occupational differences. This increases as the age of the worker decreases, for male 21-year-olds, about 15% receive positive utility for occupational differences.

The ξ_m are the major occupation unobservables. These are the utilities of choosing an occupation within the major occupation code. The utilities of each major code are normalized to the utility of choosing a non-credentialed white-collar occupation. The other six major codes are trades, manual labor, service, production, and transportation workers.

This set of estimates is used to calculate the reduction of training cost counterfactuals. The other four counterfactuals are calculated using the state-by-state estimates. The mean wage elasticities are reported for truck drivers and the 9 other occupations that have the largest cross-wage elasticities. It is a selection of occupations that experience the largest decrease in market share when the wages of truck drivers rise. That doesn't mean these occupations are the largest source of new truck drivers, because the initial share could be very small, it indicates the sensitivity of workers in these occupations to changes in truck driving wages.

		1	2	3	4	5	6	7	8	9	10
Adminstrator/Manager	1	0.0275	-0.0004	-0.0038	-0.0012	-0.0001	-0.0002	-0.0001	-0.0001	-0.0002	-0.0001
Advanced Construction Workers	2	-0.0005	0.0140	-0.0006	-0.0006	-0.0002	-0.0005	-0.0003	-0.0003	-0.0006	-0.0004
Advanced Sales Worker	3	-0.0054	-0.0007	0.0274	-0.0016	-0.0002	-0.0002	-0.0003	-0.0002	-0.0004	-0.0002
Advanced Transportation Worker	4	-0.0014	-0.0005	-0.0013	0.0170	-0.0001	-0.0002	-0.0002	-0.0001	-0.0003	-0.0002
Bus and Taxi Drivers	5	-0.0002	-0.0003	-0.0004	-0.0002	0.0036	-0.0001	-0.0002	-0.0001	-0.0003	-0.0004
Construction Worker	6	-0.0003	-0.0008	-0.0003	-0.0003	-0.0001	0.0117	-0.0003	-0.0004	-0.0006	-0.0003
Courier	7	-0.0003	-0.0005	-0.0004	-0.0003	-0.0002	-0.0003	0.0061	-0.0002	-0.0004	-0.0014
Forestry Worker	8	-0.0002	-0.0005	-0.0003	-0.0002	-0.0001	-0.0005	-0.0002	0.0050	-0.0003	-0.0004
Rail Workers	9	-0.0003	-0.0006	-0.0004	-0.0003	-0.0002	-0.0004	-0.0003	-0.0002	0.0101	-0.0004
Truck Drivers	10	-0.0003	-0.0007	-0.0004	-0.0003	-0.0004	-0.0003	-0.0016	-0.0004	-0.0006	0.0062

*The ten selected occupations with the highest cross-price elasticities to truck driving. Computed using the formula from equations (5) and (6)

Table 4: Own and Cross Wage Elasiticities

Of the 10 occupations, bus and taxi drivers have the smallest, and management has the highest own wage elasticity; truck drivers' own elasticity is about .0062. Somewhat unsurprisingly couriers have the highest cross-wage elasticity with respect to a wage increase for truck drivers. Couriers deliver goods but do not drive vehicles large enough to qualify as truck drivers. The other occupations are a mix of construction, sales, and transportation-related fields.

The counterfactuals are computed using the 2019 data since this is the final year of data and is most germane to the current market. Each counterfactual modifies some component of 2019, for example, increases wages, and simulates the employment outcome as an increase in the number of workers and a percent change in the number of workers. It reflects how many more truck drivers we

would have had in 2019 if wages were higher, etc. Furthermore, it is not a simulation of the demand side. What employers are willing to pay is not captured by this model. For example, expanding the labor market of truck drivers by allowing 18-20-year-olds to enter the market is a classic supply shifter. All else equal, this will lower wages unless demand is perfectly elastic. However, since I do not measure the employers' demand elasticities, I cannot compensate for the supply shift. Therefore, I assume the demand side elasticities are perfectly elastic, and as a consequence, wages are constant. Therefore, the counterfactual allowing 18-20-year-olds to enter the truck driving market should be an upper bound on the employment increase of that policy.

I simulated three of the five counterfactuals in two ways because the pecuniary counterfactuals can be percent or nominal increases. For the three pecuniary counterfactuals, I compute two cases, a 1% and a $1,000 change. The first counterfactual is a wage bonus to both incumbent and potential truck drivers. This is the largest employment increase of all pecuniary benefits at about 6,600 or 21,300 new drivers, depending on the type of increase. From a firm's perspective, an increase in wages by $1,000 across the board is expensive. With about 3.5 million truck drivers in the United States, the increase in the wage bill of such a policy would be substantial. Another potential tool is to increase the wages of a new entrant. The sign bonus row is an increase in wages for a new entrant, but not incumbents. This is an attractive policy for firms because it does not increase the wage bill for all workers, just new entrants. Simultaneously, it has an effect size about 30% smaller than the previous counterfactual at a much lower cost. This trade-off may be attractive to firms since it gets most of the benefit with a much lower wage bill. The final pecuniary counterfactual is a decrease in training cost as either a 1% or $1,000 transfer. This policy has a relatively small impact compared to the wage and sign bonus. Since training costs are substantially smaller than wages, a 1% decrease in training cost works out to be a $24 scholarship at the mean cost so it is somewhat unsurprising the effect size is so small. However, for a $1,000 increase, the impact is still about half of the impact of a $1,000 signing bonus. The two remaining counterfactuals are non-pecuniary, and adjust the age profile of the 2019 labor market. The first policy is to allow 18-20 year olds to enter the truck driving market. The policy results in an increase of about 100,000 drivers nationally. Two things are driving this result. First, younger people are more likely to switch to different occupations than older groups. Since all 18-20-year-olds by legislation, are not engaged in trucking as an occupation, the higher likelihood of them leaving their incumbent occupation leads them to

	1% Increase		$1000 Increase	
	Employment	% Change	Employment	% Change
Wage Response	6,614	0.17%	21,292	0.54%
Sign Bonus	4,479	0.11%	15,469	0.39%
CDL School Cost	480	0.01%	7,151	0.20%
18-20 Year Olds	96,933	2.44%	96,933	2.44%
Aged 5 Years	18,773	0.48%	18,773	0.47%

*These are the aggregate responses for each of the counterfactuals. All results were computed using the state-by-state estimates except for the CDL School cost reduction

Table 5: Aggregate Responses

be a good source of new drivers. To illustrate this, suppose 18-20-year-olds were extremely averse to new occupations relative to other groups, then it would be much harder to coax them from their incumbent occupation. Secondly, there are a large number of 18-20-year-olds in the blue-collar labor market, with about 3.99 million predicted using the CPS 2019. Since that pool of workers is so large, recruiting a modest share, about 3.2% on average results in big increases in truck drivers. In fact, almost all of the increase is coming from the expanding labor pool since the increase in market share of truck driving only increases by about 0.000013. The final counterfactual simulates what happens when the age distribution increases by five years and workers are required to retire at 70. Aging the workforce decreases the moves between occupations since occupational differences cause more disutility as a worker ages. Aging the workforce results in about 13,400 more truck drivers. This is largely a consequence of higher retention in the truck driving profession; people who were more likely to move occupations when they were younger are less likely as they age.

State Responses

The next set of results focuses on four counterfactuals using the state-specific responses. The progression of results follows the same structure as the national response table that is: wage increase, sign bonus, 18-20-year-olds, and shifting the age distribution. These results encapsulate which states are the biggest contributors to the increase in truck driving employment. All pecuniary counterfactuals are reported for a $1,000 change.

Both a thousand-dollar increase in truck driver wages and a sign bonus result in the following geographic pattern. This figure's legend records the percent increase in truck drivers by state for
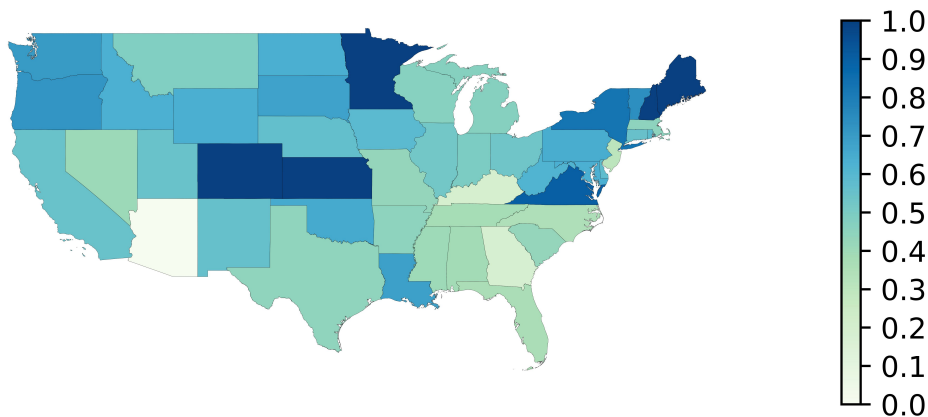
Figure 3: $1,000 Increase in Annual Earnings

*Percent change in the number of truck drivers for a $1,000 annual wage increase. Calculated from recomputing the market share of truck drivers after increasing the wage by $1,000.

a thousand-dollar increase in wages. The geographic pattern is equivalent to the signing bonus counterfactual. The only difference between the two counterfactuals is the size of the impact. The darker blue colors are associated with a larger response and the lighter the color the smaller the employment increase. Minnesota, Maine, Kansas, Colorado, and New Hampshire have the largest percentage increase in drivers within their state. Arizona is unique, with a response very close to 0. The Deep South as a group, with the exception of Louisiana, has reasonably small responses to wage increases. The Upper Midwest and Great Plains regions have higher responses relative to the southern regions. The geographic pattern is generated by the differences in the wage coefficient across states. States differ in the average impact of wage changes and the standard deviation of workers' wage responses. States with the largest average wage response have the largest elasticities in figure 3.

Allowing 18-20 year olds to become truck drivers has an interesting geographic pattern. Figure 4 Shows the share of 18-20-year-olds that choose truck driving as an occupation. This counterfactual demonstrates the utility of a mixed logit approach since the state responses depend on the distribution of prior occupations, younger workers' preferences for different occupations, and the gender and age distributions within people's previous occupations. States where the 18-20-year-old demographic are more likely to be incumbents in occupations that are most similar to truck driving will simultaneously be the states that get the biggest gains from allowing 18-20-year-olds to enter the market. Essentially, this figure demonstrates the willingness of 18-20-year-olds to choose truck
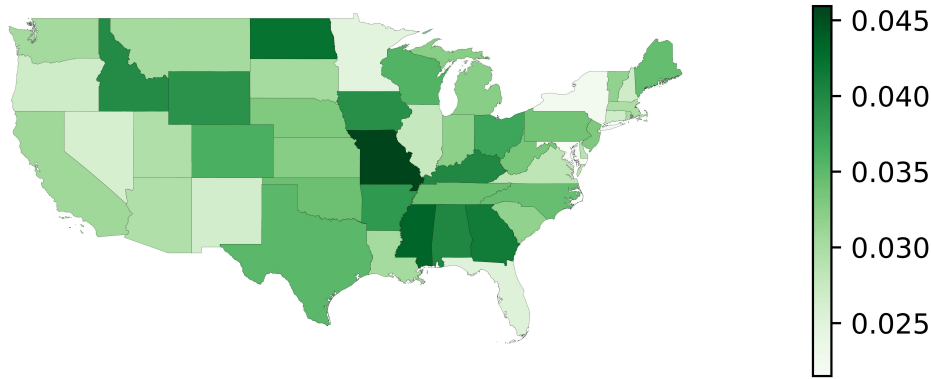
22

Figure 4: Share of 18-20-Year-olds that choose Truck Driving

*Share of 18-20-year-old workers that choose to enter the market for truck drivers. Calculated from computing the market share of 18-20-year-olds that choose truck driving as an occupation by state.

driving as a profession by state. The states where the largest share of 18-20-year-olds choose truck driving are Missouri, Arkansas, Georgia, North Dakota, Kentucky, Idaho, and Wyoming. Unlike the wage counterfactuals, the Southern states get large increases in truck drivers.

However, states also differ in the share of 18-20-year-olds in their workforce; states with the largest share of younger workers will get the largest increases in truck driving by letting them enter the market. Figure 4 ignored the scale of the market by state; for example, a state like Texas has a modest share of 18-20-year-olds predicted to choose truck driving, but has one of the largest increases in the number of new truck drivers as demonstrated in figure 4. Texas and California dominate the geographic pattern because these states have a large number of 18-20-year-olds. Ohio, Florida, and Pennsylvania also see moderately large increases. The low-population states in the Great Plains region see small increases relative to the share of 18-20-year-olds willing to choose truck driving in those states.

The final counterfactual is calculated by increasing the age distribution by five years and retiring workers 70 years and older. Figure 5 Reports the percent change in truck driving employment by state. This counterfactual is interesting because some states experience a net decline in truck drivers, although the plurality of states will see increases. Nevada, New York, Maryland, South Carolina, and Massachusetts see declining numbers of truck drivers as the age distribution rises. This distribution of prior occupations in each state is important for determining whether a state sees an increase or decrease in the share of truck drivers. Individuals are less likely to switch to
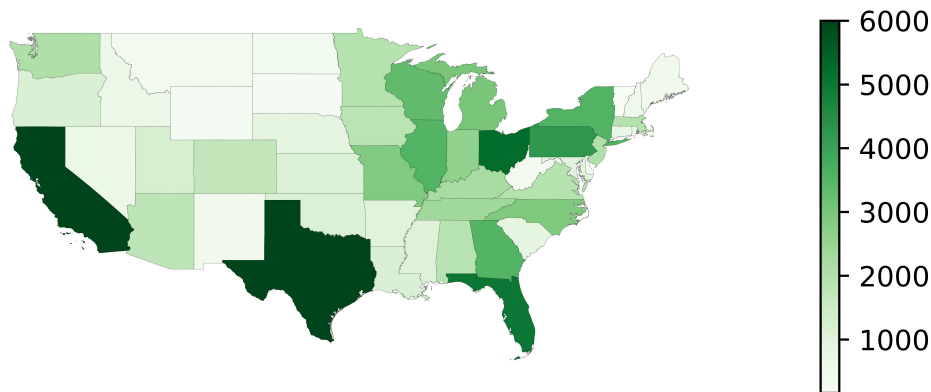
Figure 5: Increase in truck drivers allowing 18-20-year-olds enter the market

*Number of 18-20-year-old workers that choose to enter the market for truck drivers. Calculated from computing the market share of 18-20-year-olds that choose truck driving as an occupation by state and multiplying that by the total number of 18-20-year-olds in that state. The largest value are truncated at 6,000, since Texas and California exceed other state's effect sizes

more distant occupations as they age. As the workforce ages, if a greater share of the occupations in these states are more distant from truck driving, the likelihood of people switching to truck driving falls. To put it another way, the choice probabilities of occupations that are most similar to the incumbent occupation rise as the age distribution rises. Then, if there are a sufficiently large number of people in those occupations that are most similar to truck driving, the share of drivers will increase.

The declines also come from retirement. Since I force workers over the age of 70 to retire, all the drivers that exceed 70 will be dropped from the market. If either the gains from retaining more workers are less than then retirements or the distribution of prior occupations is sufficiently different from truck driving, we will observe a net decline in movers from other occupations. The red states are all states that see increases in the number of truck drivers as they age. New Mexico, Utah, Wyoming, and Indiana are the states that see the biggest increases in the share of drivers.

The final set of results tabulates the top ten states as measured by the employment increases for each of the counterfactuals. These employment increases do not lend themselves well to choropleths since a few states dominate the employment increases due to either their elasticity or large population. TABLE 6 contains the four state counterfactuals, with the two pecuniary counterfactuals computed for a $1,000 increase.

California, Texas, and New York are all very large states in terms of employment; typically,
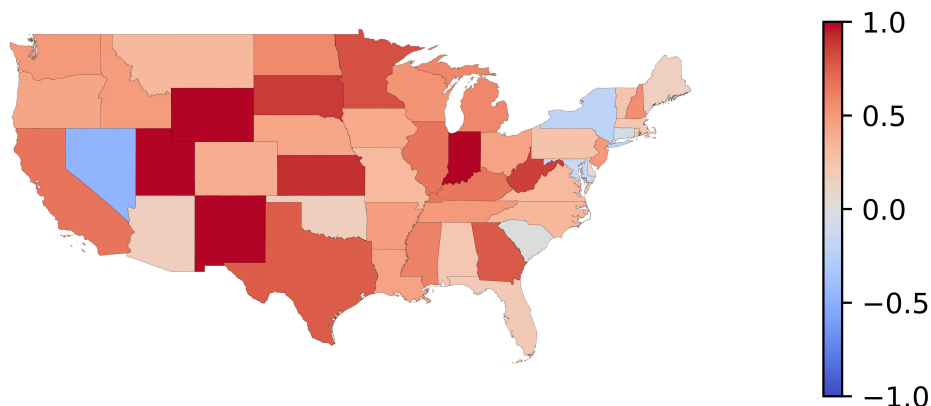
Figure 6: State heterogeneity in the share of Truck Drivers for a five-year increase in the age distribution

*This figure shows the percentage point increase in the number of drivers after aging the workforce by five years. I retire all workers that exceed the age of 70, but I keep the total labor force size the same. That is, the total employment in the blue-collar market is the same before and after the counterfactual change.

State	Wage	Sign Bonus	Share	State	18-20	Share	State	Aged 5 Years	Share
California	2,492	1,768	0.11	Texas	9,214	0.10	California	2,955	0.16
Texas	1,647	1,365	0.09	California	9,074	0.09	Texas	2,860	0.15
New York	1,567	1,191	0.08	Ohio	5,259	0.05	Georgia	1,333	0.07
Colorado	1,048	754	0.05	Florida	5,002	0.05	Indiana	1,175	0.06
Pennsylvania	1,000	677	0.04	Pennsylvania	4,253	0.04	Illinois	971	0.05
Ohio	947	655	0.04	New York	3,561	0.04	Ohio	790	0.04
Minnesota	874	619	0.04	Georgia	3,544	0.04	Michigan	585	0.03
Illinois	797	578	0.04	Illinois	3,541	0.04	New Jersey	557	0.03
Virginia	779	572	0.04	Wisconsin	3,341	0.03	Wisconsin	524	0.03
Florida	738	555	0.04	Michigan	2,993	0.03	Minnesota	499	0.03
National	21,292	15,469			96,933			18,773	height

*This table tabulates the biggest contributors to each counterfactual by state. California, Texas, and New York tend to dominate due to their large populations

Table 6: State Decomposition

they contribute a large number of new truck drivers in each of the counterfactual scenarios. In the case of the wage increase and sign bonus counterfactual, the top 10 states contribute over 50% of the national increase in truck drivers. Allowing 18-20-year-olds to enter the truck driving market leads to the largest increase in truck drivers. Again, Texas, California, and New York are big contributors. However, some smaller states also contribute to the increase in drivers. Wisconsin, Michigan, Pennsylvania, Georgia, and Ohio as a group contributed about 20% of the increase in truck drivers. Finally, the increases from shifting the age distribution follow approximately the same pattern with the loss of New York as a big contributor. California and Texas are the top two states. New Jersey and Indiana are also large contributors which follow from Figure 4.

Consequence for other Occupations

In each counterfactual except allowing 18-20-year-olds into the market, the total number of workers stays constant. Therefore, for 4 of the five counterfactuals, the gains for one occupation means another occupation is losing some percentage of its workforce. As a consequence, any policy to increase the number of drivers will have a negative impact on other occupations, a key result relevant for policymakers. For example, bus drivers are one of the most sensitive occupations to changes in truck driving. By increasing the number of truck drivers we will necessarily decrease the number of bus drivers. Understanding the tradeoffs between having more drivers and the consequences for other occupations are important for deciding whether a policy should be pursued. I report the top 10 occupations by number of jobs lost for each counterfactual except the 18-20 year old counterfactual. Table 7 reports the results for the wage and sign bonus.

Wage Increase ($1,000)	Δ Employment	Sign Bonus ($1,000)	Δ Employment
Courier	-4,593	Courier	-2,172
Bus and Taxi Drivers	-1,171	Rail Workers	-807
Advanced Construction Workers	-1,079	Advanced Construction Workers	-765
Forestry Worker	-1,062	Bus and Taxi Drivers	-745
Rail Workers	-1,035	Housekeeping/Janitorial Worker	-701
Housekeeping/Janitorial Worker	-890	Advanced Sales Worker	-673
Construction Worker	-844	Construction Worker	-648
Advanced Sales Worker	-796	Forestry Worker	-574
Material Movers	-563	Material Movers	-442
Advanced Transportation Worker	-529	Extraction Worker	-433

Table 7: Occupational Impact for Increasing wages in Truck Driving

Table 7 is broken into two parts, the wage increase on the left and the sign bonus on the right. The numbers are reported as the net decline in the number of workers nationally in each occupation as the wage of truck drivers rises. As reported in TABLE 4, Couriers have the highest cross-wage elasticity relative to other occupations in the sample. As such they see the biggest decline in their workforce as the number of truck drivers increases. Most occupations are unsurprising, the linkages from truck drivers to the Bus and Taxi Drivers and Couriers, Construction, Rail, Material Movers, and transportation workers are evident. These occupations are related to truck drivers, either because truck drivers enable these workers to complete their tasks, or they already have a specialized license similar to that of the CDL required for truck drivers. Forestry and Housekeeping/Janitorial workers are at first observation surprising members of this group of occupations. However, there is

a large number of janitorial workers in the sample, so even if a small share of them leave to become truck drivers the impact is fairly large. Forestry workers may see a decline in workers because of their primarily rural status. Truck driving is a fairly good occupation in many rural areas because the job builds in commuting, and while working a driver is on the road to deliver or pick up a load not traveling between home and work. Therefore, it may be an easier transition for forestry workers to become truck drivers, although this claim needs further investigation. The results for a sign bonus are qualitatively similar to that of the wage increase with one major difference. The advanced transportation workers fall from the tenth spot and are replaced by extraction workers. The extraction workers are coal, oil, and natural gas extraction workers. They work in a variety of contexts, extracting fuel for the US.

The second counterfactual I consider is allowing aging of the workforce by five years. Because it doesnt increase the number of workers, not all occupations benefit. Table 8 reports the results for an aging workforce.

Top Occupations (H-L)	Δ Employment	Bottom Occupations (H-L)	Δ Employment
Office Clerk	79,467	Rail Workers	-62,746
Sales Worker	77,676	Advanced Sales Worker	-45,052
Production/Operation Workers	75,829	Advanced Transportation Worker	-39,030
Food Preparation Worker	55,190	Gaming Service Worker	-33,613
Construction Worker	49,198	Printing Workers	-32,481
Housekeeping/Janitorial Worker	47,188	Textile/Shoe Workers	-30,939
Emergency Service/Criminal Justice	38,418	Extraction Worker	-30,274
Transportation Clerk	36,135	Laundry Workers	-29,158
Service Worker	34,952	Energy Installer	-26,120
Material Movers	33,225	Funeral Workers	-24,627

Table 8: Aging Counterfactual

As the workforce ages, there are two components of the mode to consider. First aging makes workers less willing to move to different occupations. The disutility for moving from one occupation to another increases as workers ages, which implies that as the workforce ages, we will observe less moving between occupations. In effect, the labor market is crystallizing. Secondly, occupations differ in their retention and entrance rates. There are occupations where a large share of the workforce comes from new entrants, these are occupations that will experience the biggest declines as the labor market ages. The occupations with the highest retention rates are going to see the biggest increases in the number of workers because workers who would have left had they been

younger are no longer willing to leave the occupation. In essence, high retention occupations see increases as the labor market ages, high entrance rate occupations see the biggest declines. Table 8 records the top and bottom ten occupations by the number of workers the occupation gains or loses. Notably, 61% of occupations see a decline in the number of workers as the labor market ages. The other notable fact is that truck driving is not in the top or bottom ten occupations.

Conclusion

I used four data sets to estimate a mixed logit model of occupational choice to compute five counterfactuals. Given the large concern of a truck driving shortage, my five counterfactual experiments address some of these concerns. The wage and sign bonus counterfactual demonstrates that workers are responsive to wage increases. A $1,000 increase in all wages nets about 21,100 more drivers nationally, and a $1,000 sign bonus nets about 15,500. While these increases are small in the context of the approximately 3.5 million drivers in the United States, it is clear that workers respond to higher wages. An inelastic labor supply does not imply a shortage of workers. Firms may wish that smaller wage increases would coax enough workers into their occupations to cover the perceived shortfall. However, the small responses do not imply a shortage. A sign bonus is an attractive tool for firms to use. A larger bonus will induce more workers to switch occupations at a reasonable cost; however, retaining them becomes a challenge. Presumably, such a policy would lead to a higher wage bill as firms pay more to retain their workers.

The aging of truck drivers has often been cited as a potential cause of the shortage. In my counterfactual, I age the distribution of workers by five years and retire all workers older than 65. Surprisingly, this counterfactual showed increases in the number of truck drivers in most states. Only Nevada, New York, Maryland, and Massachusetts experience declines in the number of drivers as their workers age. Of course, if all drivers get older and no young drivers enter the market, of course, the number of truck drivers will fall.

In the final and largest counterfactual, I allow 18-20-year-olds to enter work as truck drivers. Currently, DOT regulation forbids 18-20-year-olds to drive commercial trucks across state lines. The Bipartisan Infrastructure Bill financed a trial program to authorize young workers to apprentice and work as truck drivers. My counterfactuals predict a nationwide increase of about 97,000 new

drivers if 18-20-year-olds are allowed to enter the market. This result ignores the concerns about safety associated with allowing younger drivers to work as professional drivers. Other research into the unintended consequences of allowing young drivers to enter the market are needed.

Finally, I compare which occupations see the biggest losses in the pecuniary counterfactuals. I find that those occupations that are most similar in proximity or skills to truck driving see the biggest declines in their workforce as the labor market ages.

References

Arteaga, Cristian, JeeWoong Park, Prithvi Bhat Beeramoole, and Alexander Paz. "xlogit: An open-source Python package for GPU-accelerated estimation of Mixed Logit models." *Journal of Choice Modelling* 42 (2022): 100339. ISSN: 1755-5345. https://doi.org/https://doi.org/10.1016/j.jocm.2021.100339.

Berry, Steven T. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 1994, 242–262.

Berry, Steven T, James A Levinsohn, and Ariel Pakes. *Automobile prices in market equilibrium: Part I and II*, 1993.

Boyd, J Hayden, and Robert E Mellman. "The effect of fuel economy standards on the US automotive market: an hedonic demand analysis." *Transportation Research Part A: General* 14, nos. 5-6 (1980): 367–378.

Chandiran, P, M Ramasubramaniam, VG Venkatesh, Venkatesh Mani, and Yangyan Shi. "Can driver supply disruption alleviate driver shortages? A systems approach." *Transport policy* 130 (2023): 116–129.

Goolsbee, Austan, and Amil Petrin. "The consumer gains from direct broadcast satellites and the competition with cable TV." *Econometrica* 72, no. 2 (2004): 351–381.

Hausman, Jerry, Gregory Leonard, and J Douglas Zona. "Competitive analysis with differenciated products." *Annales d'Economie et de Statistique*, 1994, 159–180.

Hensher, David A, and William H Greene. "The mixed logit model: the state of practice." *Transportation* 30 (2003): 133–176.

Keane, Michael P, and Kenneth I Wolpin. "The career decisions of young men." *Journal of political Economy* 105, no. 3 (1997): 473–522.

McFadden, Daniel. "Economic choices." *American economic review* 91, no. 3 (2001): 351–378.

McFadden, Daniel, and Kenneth Train. "Mixed MNL models for discrete response." *Journal of applied Econometrics* 15, no. 5 (2000): 447–470.

Mittal, Neha, Prashanth D Udayakumar, G Raghuram, and Neha Bajaj. "The endemic issue of truck driver shortage-A comparative study between India and the United States." *Research in transportation economics* 71 (2018): 76–84.

Nevo, Aviv. "A practitioner's guide to estimation of random-coefficients logit models of demand." *Journal of economics & management strategy* 9, no. 4 (2000): 513–548.

———. "Measuring market power in the ready-to-eat cereal industry." *Econometrica* 69, no. 2 (2001): 307–342.

Petrin, Amil. "Quantifying the benefits of new products: The case of the minivan." *Journal of political Economy* 110, no. 4 (2002): 705–729.

Phares, Jonathan, and Andrew Balthrop. "Investigating the role of competing wage opportunities in truck driver occupational choice." *Journal of Business Logistics* 43, no. 2 (2022): 265–289.

Richards, Timothy J, and Zachariah Rutledge. "COVID-19, Truck Rates and Trucking Shortages." *Truck Rates and Trucking Shortages (August 23, 2022)*, 2022.

Train, Kenneth E, Daniel L McFadden, and Moshe Ben-Akiva. "The demand for local telephone service: A fully discrete model of residential calling patterns and service choices." *The RAND Journal of Economics*, 1987, 109–123.

Train, Kenneth E, and Clifford Winston. "Vehicle choice behavior and the declining market share of US automakers." *International economic review* 48, no. 4 (2007): 1469–1496.