Effects of the of Autonomous Trucks on the Labor Market for Truck Drivers

Details of Methodology

Charles Hodgson[1]

1. Summary

This research examines the impact of the deployment of self-driving technology in the long-haul trucking industry on the labor market for truck drivers. The objective is to evaluate the effect of different deployment scenarios on truck driver employment. A naive approach to the problem would take the projected number of self-driving trucks deployed and convert this to a number of drivers displaced according to some utilization ratio (for example, one self driving truck might displace two drivers). Our analysis moves beyond this simple accounting exercise by measuring demand and supply elasticities in the market for long haul drivers. We also account for the job creation effect in the short haul sector - additional long haul jobs served by automated trucks will require additional "last mile" short haul drivers. To compute this job creation effect, we take account of the actual distribution of trip lengths.

This document details our methodology. The main results of the analysis are summarized in the accompanying output files.

2. Demand for Long Haul Drivers

We estimate the wage elasticity of demand in the long haul trucking sector using the following demand relationship

$$log(S_{sy}) = \alpha_y + \beta log(wage_{sy}) + \gamma income_{sy} + \epsilon_{sy} \tag{1}$$

Where S_{sy} is the share of the workforce in state s in year y employed in long haul trucking. This is obtained from BLS data. Long haul trucking is given by occupation code 53-3032. On the right hand side of the equation is a year fixed effect, α_y, $wage_{sy}$ the annual wage for long haul drivers in state s, year y, $income_{sy}$, the median income in state s, year y, and an error term. The main coefficient of interest is β, the wage elasticity of demand. Data includes all years from 2001-2016 on all US states excluding Hawaii.

[1] This research, and all proposed work, is carried out in a personal capacity.

Since we want to obtain the wage elasticity of demand, β, we need to instrument wages with a supply shifter that induces exogenous variation in the wage level across states and years. An OLS regression that does not use an IV would be subject to endogneity problems - wages may rise because supply has fallen, and the equilibrium has moved up the demand curve, or because demand has risen and equilibrium has moved up the demand curve. To isolate the slope of the *demand curve*, it is necessary to use an IV. We use the share of all employees in a state-year that are union members as our "supply shifter" instrument. This data comes from the CPS. Conditional on state-year median income and year effects, changes in the fraction of employees in a state that are unionized affect the demand for truck drivers from that state only through its effect on wages, where more unionized state-years will have higher wages than less unionized state-years.

3. Driver Supply

We estimate driver supply using a logit occupation choice model. We suppose that drivers can choose between working in long haul trucking, short haul trucking, other driving, and an "outside option". The "outside option" can be interpreted as the option a driver has to leave driving employment. We do not say anything about what types of jobs former drivers take, or what share of them actually find jobs. For the purposes of this report, we interpret drivers who switch from driving employment to the outside option as "unemployed". For each job type, j, driver i obtains a utility given by:

$$u_{syji} = \alpha_{sy} + \delta_j + \beta log(wage_{syj}) + \xi_{syj} + \epsilon_{syji} \tag{2}$$

Where α_{sy} is a state-year fixed effect, δ_j is a job type fixed effect, $wage_{syj}$ is the wage in state s, year y, for job type j. Wages come from the BLS, and the "other driving" wage is the mean of all occupations with occupation cods beginning in 53-30, excluding long and short haul trucking. ξ_{syj} is a state-yer-occupation level unobservable, and ϵ_{syji} is an individual i specific logit shock. Note that "outside option" utility is fixed at 0, and "other driving" has $\delta_j = 0$ (these are standard normalizations). The logit shock assumption means that the model can be written as follows (after Berry, 1994):

$$log(S_{syj}) - log(S_{sy0}) = \alpha_{sy} + \delta_j + \beta log(wage_{syj}) + \xi_{syj} \tag{3}$$

Where S_{syj} is the share of drivers employed in job j in state s, year y, and S_{sy0} is the share of drivers "unemployed". To obtain these shared, we use employments in each job/sector from the BLS, and divide the number of drivers in each job by the total number of "potential drivers" in a state. Potential drivers are obtained by finding the maximum number of people ever employed in driving occupations in a state over the years 2000-2016 (plus one so the share S_{sy0} is never equal to 0). The main coefficient of interest in this regression is the pseudo-elasticity, β.

Note that we include state-by-year fixed effects and occupation fixed effects. If we did not instrument for wage, there would only be an endogeneity problem here if there is an unobserved factor at the year-by-occupation or state-by-occupation level. That is, something beyond wages is making long haul trucking more/less appealing over time relative to other occupations, or something makes the relative attractiveness of long and short haul different in FL than it is in TX.

Although the fixed effects included here make endogeneity less of a problem, we still instrument for wages to ensure a credible estimate of the wage elasticity of supply. In particular, we need an instrument that shifts wages by shifting the demand curve. We use two instruments - state-year median income and state-year employment in manufacturing. Higher median income in a state increases demand for shipments into that state, and higher manufacturing employment increases demand for shipments out of that state. These instruments should have no effect on the relative supply of workers into different driving jobs, other than through their effect on relative wages.

4. Computing Projected Employment

We use the estimated supply and demand elasticites to measure the labor market response to the deployment of automated vehicles at a per-mile cost below that of a long-haul truck driven by a human. The elasticity of demand given by β in equation 1 is estimated to by -3.66, which means a 1% increase in the price of long haul trucking will lead to a 3.66% reduction in demand. The elasticity of supply from the estimated supply model is given by $\beta(1 - S_{syj})$. Notice that it varies by state, year, and job type. For the purposes of these projections, we use the average elasticity in 2016 for long haul drivers, 2.7. this means a 1% increase in wages will lead to a 2.7% increase in the supply of long haul truckers.

We compute forward labor market conditions for four scenarios. Table 1 below outlines the scenarios. The scenarios assume that costs fall over time and the region of automated truck deployment increases. Each scenario supposes that automated trucks are deployed at a per mile cost that is less than the cost of a human driver. For example, in scenario 1 the per mile cost of an automated truck is 90% of the per-mile cost of a human-driven truck. Automated truck utilization rate refers to the maximum amount of time that an automated truck can be active per year. We convert this into an annual mileage by noting that a truck running for 1 year at 100% utilization and an average speed of 55 mph would cover 481,100 miles.

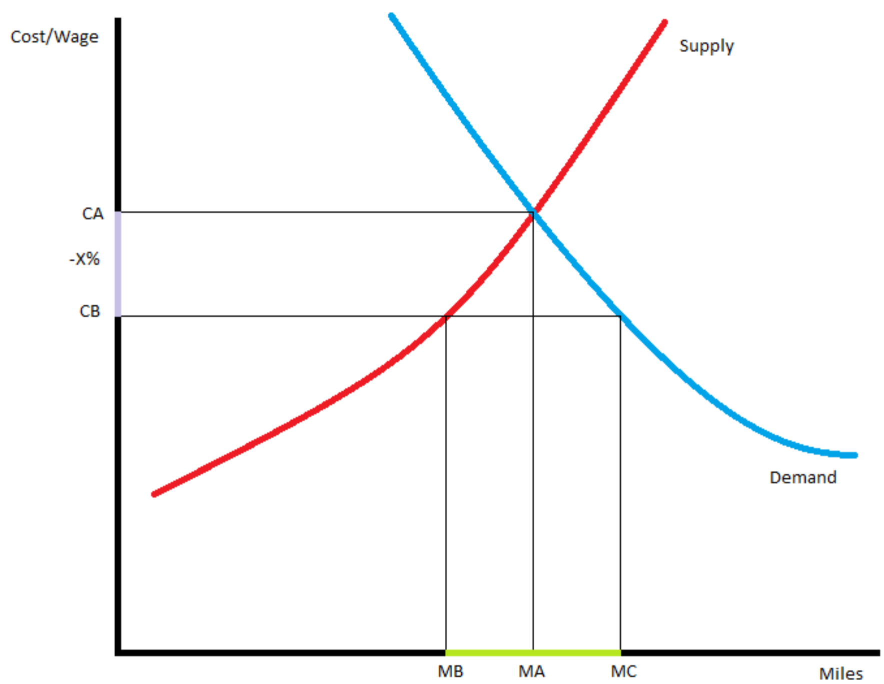
Table 1: Deployment Scenarios

Scenario	Year	Region	Per Mile Cost Relative to Human	Automated Truck Utilization Rate
1	2023	Texas	95%	30%
2	2025	Southwestern US	92%	40%
3	2027	Southern US	88%	50%
4	2029	Whole US	85%	60%

For each scenario, the first step is to forecast the baseline number of long haul miles that would be driven in the absence of automated vehicles. To compute this, we use proprietary data on the total number of long haul miles driven in each of the four regions specified in 2017 and an estimate of the annual growth rate of miles to project expected miles in each of the region-years specified. We convert mileages to long haul jobs by assuming each long haul driver covers 89,804 miles per year.The baseline miles and jobs are recorded in the first two columns of Table 2.

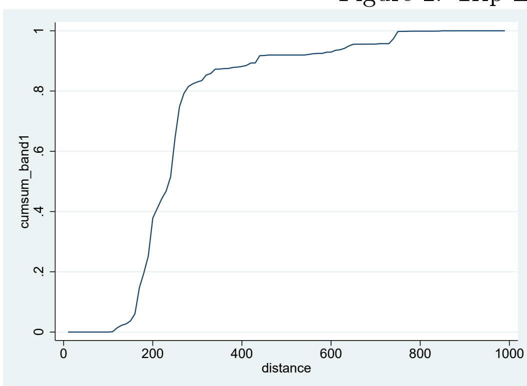
The counterfactual mileage calculations are illustrated by Figure 1. For each scenario, we take the implied percentage reduction in cost from automated trucks, and convert that to an increase in demand and a reduction in human supply. In the figure, the $X\%$ decrease in cost from CA to CB corresponds to an increase in miles demanded from MA to MC and a decrease in long haul miles supplied by humans from MA to MB. In particular, for an $X\%$ decrease in cost, we compute the counterfactual demand as $MC = MA(100 + 3.66X)/100$ and counterfactual supply as $MB = MA(100 - 2.7X)/100$.

Figure 1: Mileage Calculation

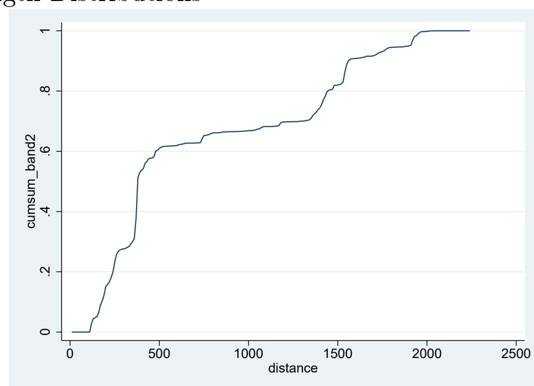


The total long haul miles driven by humans is then given by MB. The mileage $MC - MB$ (the green section in Figure 1) is divided between long haul miles driven by autonomous trucks and “last mile” human drivers. We assume that for every trip driven by an autonomous vehicle, there will be an average of 25 miles of short haul human driving on each end. To compute the breakdown of the green section into automated and human miles, we use 2017 trip length distributions for each of the four regions. These distributions are from a proprietary source, and record the share of all long haul miles that correspond to trips of different lengths, rounded to the nearest 10 miles (for trips longer than 100 miles). For example, in Texas in 2017, 1.4% of long haul miles were driven on trips between 170 and 180 miles in length. The cumulative distribution of mileage by trip length for the four different regions is recorded in Figure 2 below.

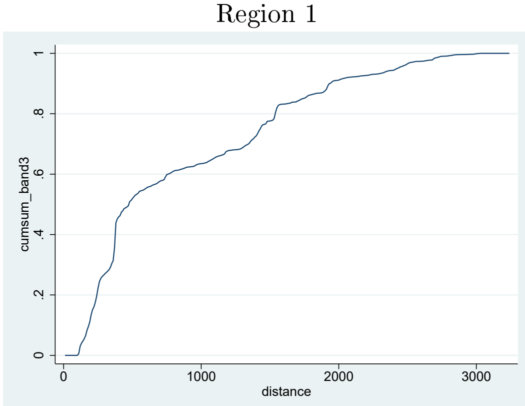
Figure 2: Trip Length Distributions



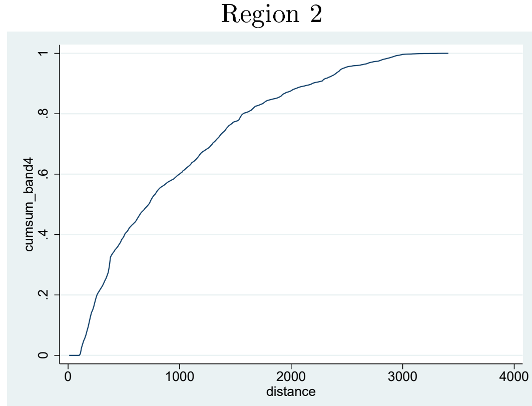
Region 1



Region 2



Region 3



Region 4

Using these distributions, the total mileage given by $MC - MB$ can be converted into a total number of trips. 50 miles of each trip are allocated to short haul drivers. The remaining mileage is allocated to long haul automated vehicles. Finally, we convert short haul miles to short haul jobs by assuming short haul drivers only cover 40% of the miles of long haul drivers each year - urban speeds are slower and a short haul driver spends a greater share of their time loading and unloading etc. Loss in long haul jobs and gain in short haul jobs are recorded in Table 2. Detailed results are available in the accompanying output files.

Table 2: Results

Scenario	Baseline Mileage	Baseline LH Jobs	LH Job Loss	SH Job Gain
1	1.732 bil	19,293	2,605	3,307
2	8.026 bil	89,380	19,306	14,931
3	13.886 bil	154,630	50,100	36,351
4	46.414 bil	516,843	209,321	136,147

Scenario	2017 Total Jobs	Simulation Total Jobs	Job Growth from 2017 (%)
1	16,710	18,011	7.79
2	75,105	76,046	1.25
3	123,598	119,070	-3.66
4	403,844	362,100	-10.34

Driver Retirement

Reduced demand for long haul drivers will leads to lower driver employment, but the loss in jobs is likely to be absorbed by driver retirements, and therefore should not be thought of as unemployment. In order to quantify this effect, we project retirements from long haul trucking each year from 2018 to 2030.

To do this, we obtain the distribution of driver ages each year from 2000 to 2016 using BLS data. We then compute an average retirement rate for each age over 60. For example, the retirement rate at 65 is the difference between the number of 65 year old truck drivers in year y and 66 year old truck drivers in year $y+1$, divided by the number of 65 year old truck drivers in year y, averaged over all the years in the data. We assume that all truck drivers who are still employed at age 70 retire the next year. We then take the 2016 distribution of truck driver ages and project it forward, aging each driver by one year and then applying the retirement rates to all drivers over 60. We repeat this for every year up to 2030, and record the number of retirements for each year.

The projected retirements are therefore a function of average historical retirement rates and the current age distribution of drivers. If, for example, most drivers are over 50 years old, then the projected retirements over 20 years will be higher than they would be if most drivers were under 50 years old.

References

Berry, S. T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics*