The Traveling Trucker Problem *

Treb Allen
Dartmouth and NBER

David Atkin
MIT and NBER

Santiago Cantillo Cleves
University of California San Diego

Carlos Eduardo Hernández
Universidad de los Andes

December 2023

Abstract

This paper documents three new stylized facts showing that truckers in Colombia frequently choose to make complex chains of shipments in a single trip before returning home. It then provides a new model of optimal trucker trip-chaining with a general geography that is consistent with these facts.

1 Introduction

Workhorse trade models typically assume that trade costs are of an "iceberg" variety, causing goods (or, equivalently, workers' labor endowments in the origin) to "melt" en route, see e.g. Samuelson (1954). In reality, however, the transportation of goods is accomplished by workers who may not reside in either the origin or the destination of the good. Moreover, these transportation workers—henceforth "truckers"—may be engaging in a complicated logistical problem of how to maximize

*Allen: treb@dartmouth.edu; Atkin: atkin@mit.edu; Cantillo Cleves: santiago@ucsd.edu; Hernández: carlose-hernandez@uniandes.edu.co. Allen is grateful to the hospitality of CEMFI, where this paper was completed. All errors are our own.

their own profits by choosing an optimal sequence of shipments. How do truckers solve this "traveling trucker problem"? And how does the solution to this problem affect our understanding of trade flows and the spatial distribution of economic activity?

This short paper makes two contributions to answering these questions. First, it uses a novel database on the universe of truckers in Colombia to document that reality departs substantially from the typical iceberg view: in particular, (1) truckers home locations play an important role in determining trade flows; (2) truckers frequently chain together multiple shipments on the same trip; and (3) truckers appear to be optimizing these trip chains to maximize revenue and minimize costs. Second, the paper presents a new model of the traveling trucker problem that delivers tractable expressions governing trucker trip choice and its influence on the pattern of shipments. The predictions of the model are consistent with our stylized facts.

This paper is related to recent work examining the economics of the transportation sector. Like Brancaccio, Kalouptsidi, and Papageorgiou (2020), we use highly detailed data on carrier itineraries to document new empirical regularities; like Yang (2022), our facts and theory emphasize the important role that a trucker's home location plays in their trip choice. This paper contributes to this literature by also offering an analytical solution to both the traveling trucker problem and the resulting distribution of shipments in a setting with many locations and arbitrary costs and payoffs to transiting between these locations. To do so, the paper leverages the techniques developed in Allen and Arkolakis (2022) whereby truckers choose across a countably infinite set of possible trips. Our analytical solution highlights how complex transportation networks and endogenous trade costs yields richer patterns for trade flows than a typical trade model.

2 Trucking in Colombia

We begin by describing the empirical context before describing our novel data set.

2.1 Empirical context

Colombia has a mountainous terrain, large heterogeneity in road quality, and large variation in population density. Unlike in other Latin American countries, the population is not heavily concentrated in a single city. Given its mountainous terrain, almost all intranational trade in Colombia is done by truck, accounting for 96% of non-coal and oil tonnage transported within Colombia in 2019 (Ministerio de Transporte, 2018).

The trucking industry is subject to multiple government regulations. This paper focuses on the transportation of non-agricultural products. Most relevant to what follows is that firms moving these goods must hire truckers through intermediaries (who guarantee that other regulations are respected) and that these intermediaries must report to the government the date, product, origin, destination, owner, and driver of the truck that completed the shipment.

2.2 A Unique Truck Dataset

We assemble a unique data-set on the trucking industry in Colombia, ideally suited to examine how truckers supply various routes. The data-set comes from three sources. First, we assemble shipment-level data on the universe of legally registered non-agricultural shipments in Colombia between December 26, 2017 and December 24, 2018, which we collected by scraping the complete history of shipments made by every truck in the country from the Government web portal. For each individual shipment, we observe the origin municipality, destination municipality, and date, as well as the truck's license plate.

Second, we merge the shipment level data with a data set containing information on truck owners from the National Registry of Trucks, matching using the license plate. These data provide us information about the owner of the truck, most notably his or her municipality of residence.

Third, for every pair of municipalities, we calculate travel times along the optimal route between the pair given the existing road network, with the road network drawn from Open Street Maps in July 2018. In particular, we construct a speed image of the network accounting for the terrain and road quality and use the Fast Marching Method (see Sethian (1996)), as popularized in the economic geography literature by Allen and Arkolakis (2014).

For the stylized facts that follow, we focus on the 1,744,293 shipments made between Christmas 2017 and Christmas 2018 by 30,955 truckers who a) solely own only one truck (74% of truckers), and b) whose residence is reported in the National Registry of Trucks (68% of truckers with one truck).

2.3 Three New Stylized Facts about Trucker Trip-Chaining

We now describe three new facts about the route choices of truckers in Colombia.

Stylized Fact 1: Truckers are more likely to supply routes close to their home

Table 1: Truckers are more likely to ship near their homes

Sample:	$Shipments_{hodi} > 0$		$Shipments_{hod} > 0$	
	(1)	(2)	(3)	(4)
Log travel time, o to d	-0.136***	-0.162***	-0.300***	-0.310***
	(0.007)	(0.006)	(0.010)	(0.009)
Log travel time, h to o	-0.157***	-0.149***	-0.521***	-0.570***
	(0.005)	(0.005)	(0.009)	(0.009)
Log travel time, d to h	-0.178***	-0.142***	-0.499***	-0.550***
	(0.006)	(0.006)	(0.009)	(0.009)
Dummy variable, $o = d$		-0.610***		-3.360***
		(0.152)		(0.186)
Dummy variable, $h = o$		0.192***		1.043***
		(0.017)		(0.026)
Dummy variable, $d = h$		0.253***		0.758***
		(0.022)		(0.027)
Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Trucker FE	Yes	Yes	Yes	Yes
Observations	364718	486491	134926390	175702769

Notes: Poisson pseudo maximum likelihood estimation. The dependent variable is the number of shipments of a trucker between December 26, 2017 and December 24, 2018 between an origin municipality and a destination municipality. Standard errors clustered at the trucker level reported in parentheses. Stars indicate statistical significance: * p<0.10 ** p<0.05 *** p<0.01.

Table 1 runs a gravity-like specification that regresses the number of shipments trucker i

supplied on route od on the distance between od, measured by the log travel time described above. In addition, we include the log travel time from both the origin and destination to the truckers home h (and trucker, origin, and destination fixed effects):

$$\ln\left(\#Shipments_{hodi}\right) = \beta_1 \ln traveltime_{od} + \beta_2 \ln traveltime_{ho} + \beta_3 \ln traveltime_{dh} + \delta_i + \delta_o + \delta_d + \varepsilon_{odi} \tag{1}$$

While the gravity literature has long noted that trade in goods and services declines with distance, we also find that, conditional on od travel time, truckers are more likely to supply routes whose starting or end points are closer to their homes. Column (1) focuses on the intensive margin, i.e. restricting the analysis to observations for which $Shipments_{hodi} > 0$, using a Poisson Pseudo-Maximum Likelihood (PPML) estimator well suited to count data. A 10% reduction in ho or dh distance increases the annual number of shipments by 0.16 and 0.18 respectively (with truckers in our sample completing 54 trips a year, on average). Each of these are larger than the semi-elasticity with respect to od distance. Column 2 includes intra-muncipality shipments and shipments that begin or end at home (when $o = d$, $h = o$, or $d = h$) and includes dummies for each of these cases. Annual shipments are 0.19–0.25 greater if either the origin or destination is the trucker's home, but are lower if the origin and destination are in the same municipality.

Columns 3 and 4 consider both intensive and extensive margins by including all $\#Shipments_{hodi} = 0$ when at least one trucker from h has served that od. Including the extensive margin, the bias towards trips close to home become nearly four times larger.[1] The routes truckers serve are strongly influenced by where truckers live.

Stylized Fact 2: Truckers frequently transport multiple shipments prior to returning home

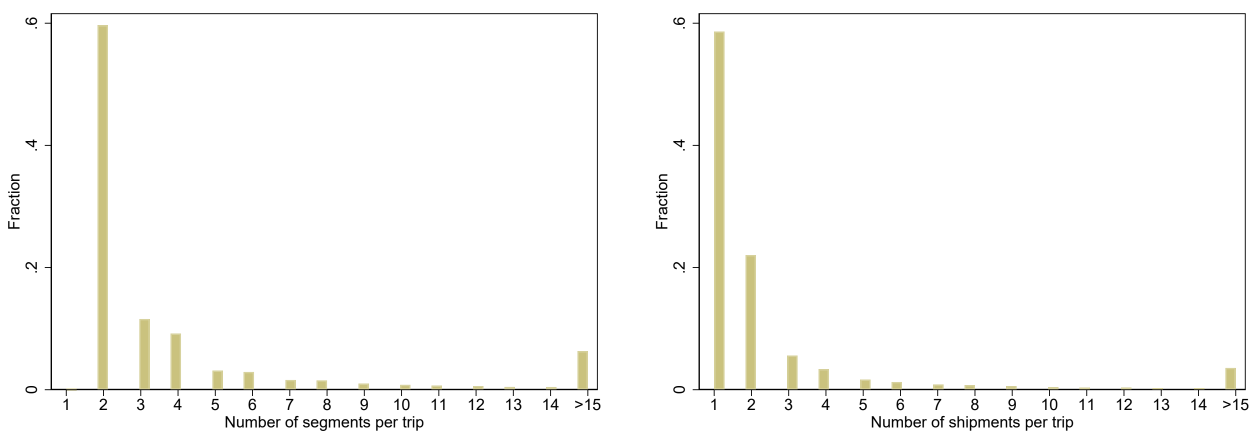
Figure 1 plots a histogram of the number of segments a trucker completes (panel a) and shipments a trucker delivers (panel b) before returning to their home. To do so, we first define a trip as a sequence of segments that begins and ends in the trucker's home municipality, where a segment is either an

[1]If anything, this is an underestimate of the true effect of distance to home on the extensive margin, as the sample excludes od pairs which no trucker from h serves.

(observed) shipment or an (unobserved) transit from the destination of the prior observed shipment to the origin of the subsequent observed shipment.[2] By this definition, a trucker completes 7 trips per year, on average. As Figure 1 shows, while the modal trip goes from home to a destination and returns empty, 40 percent of trips involve visiting an additional location and/or making multiple shipments. Truckers frequently embark on complex trips involving multiple shipments and locations before returning home.

Figure 1: Distribution of trip length

(a) Number of segments (b) Number of shipments

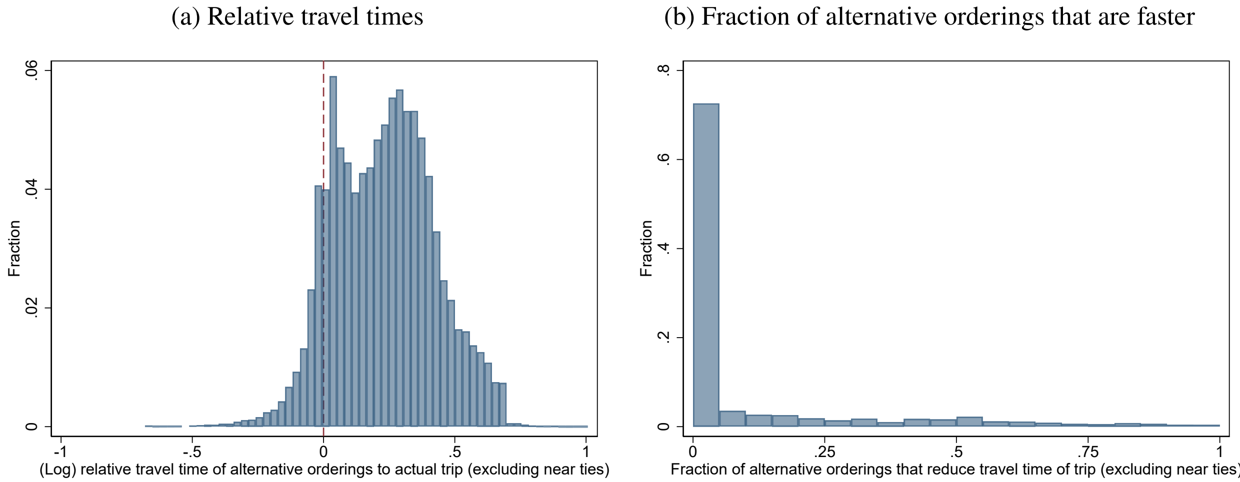


Notes: These figures depict the distribution of trip lengths (panel a) and observed shipments (panel b) across all trips made by truckers between December 26, 2017 and December 24, 2018. We exclude trucker who complete more than 1 shipment per day as in those cases the date stamps do not reveal the ordering of routes. A trip is defined as beginning in the home municipality and ending when the trucker returns to the home municipality. The trip length is the number of segments made during the trip, where a segment is either an (observed) shipment or an (unobserved) transit from the destination of the prior observed shipment to the origin of the subsequent observed shipment. Truckers are assumed to be at home on Christmas 2017 and 2018, thus, all trips either begin with a transit segment from h to the origin o of the first shipment after Christmas 2017, or with a shipment segment that originates at h (and all trips either end with a transit from the last shipment destination d to h, or with a shipment whose destination is h).

[2]This definition of a trip minimizes both the number of segments and the travel time spent by drivers on unobserved transits but may introduce measurement error if, for example, drivers actually return home empty in between some of the observed shipments rather than going directly to the next origin. We assume that all truckers spend Christmas 2017 and 2018 at home so there is at least one trip per trucker; 93% of truckers complete more than one trip a year.

Stylized Fact 3: The choice of itinerary appears to maximize payoffs

Figure 2: Efficiency of chosen trips, keeping revenue constant



Notes: These histograms compare the travel time of the actual trips made by truckers relative to the travel times of 100 alternative orderings that include the same observed shipments as the actual trip but in which these shipments are completed in a randomly chosen order. As the observed shipments are the same, revenue is the same, so minimizing travel time minimizes travel costs. The sample includes all trips made between December 26, 2017 and December 24, 2018 by truckers who did not complete more than 1 segment per day. A trip is defined as beginning in the home municipality and ending when the trucker returns to the home municipality. The travel time is the sum of the travel time along all segments, where a segment is either an observed shipment or an unobserved transit from the destination of the prior observed shipment to the origin of the subsequent observed shipment. In panel (a), we report the distribution of relative travel times, the ratio of the travel time along each alternative trip to the actual travel time. In panel (b), we report the distribution of how frequently a randomly chosen alternative ordering of the observed shipments would result in a lower travel time than the actual trips made by truckers. In both cases, the histogram excludes alternative trips that would have yielded travel times that are nearly identical (within 1%).

Figure 2 compares the travel costs along trips actually taken to equally-lucrative alternative routes. To do so, we hold fixed the shipments the trucker chooses to transport on a given trip (and hence, revenue), and ask whether there are alternative sequences of those same shipments that would have incurred lower travel times (and, hence, costs). As before, transit segments are added when the previous shipment's destination is different from the next shipment's origin. Panel (a) shows the histogram of travel times under a random sample of 100 alternative trips relative to the actual travel time; and (b) shows the fraction of alternative trips that were faster. Truckers itinerary choice usually minimizes total transit costs, with only one third of possible sequences faster. Truckers appear to be optimizing their itineraries for a given set of shipments.

3 A tractable model of the traveling trucker

We now develop a tractable model of the traveling trucker in a general geography. The model both is consistent with the stylized facts above and highlights how how incorporating the micro-foundations of the traveling trucker changes the implications of a standard trade model.

Setup

Consider a world comprising N locations. In each location, there are workers who produce goods and truckers who ship the goods between locations. A trucker i residing in location $h \in \{1, ..., N\} \equiv \mathcal{N}$ chooses to make a *trip*, which is a sequence of locations that begins and ends in h, so that a trip r of length K can be written as:

$$r = \{o_0 \equiv h, o_1, o_2, ..., o_{K-1}, o_K \equiv h\}. \tag{2}$$

A *segment* is a sequential pair of elements in a trip. Suppose that segment $o \in \mathcal{N}$ to $d \in \mathcal{N}$ yields a trucker a payoff p_{od} but incurs a cost c_{od}. In addition to these segment payoffs, a trucker incurs a dis-amenity "homesickness" cost ρK proportional to the length of trip. Each trucker i residing in h has an idiosyncratic preference for trip r, $\varepsilon_{ih}(r)$, that is assumed to be independently and identically distributed extreme value (Gumbel) with shape parameter θ. The value of trucker i from h choosing trip r of length K is then:

$$V_{ih}(r) = \sum_{k=1}^{K} \left(p_{o_{k-1}o_k} - c_{o_{k-1}o_k}\right) - \rho K + \varepsilon_{ih}(r). \tag{3}$$

Trucker trip choice

Given her realization of $\varepsilon_{ih}(r)$, a trucker will choose the trip that maximizes her payoff from equation (3). The expected utility U_h of truckers from h (taken over possible realizations of the idiosyncratic preferences across truckers) is:

$$U_h \equiv \mathbb{E}\left[\max_{r\in\mathcal{R}_h} V_{ih}(r)\right], \tag{4}$$

where $\mathcal{R}_h$ is the (countably infinite) set of all routes that begin and end in h. Given the assumed distribution of the idiosyncratic trip preferences and explicitly enumerating all possible trips, we can write this as:

$$U_h = \frac{1}{\theta}\ln\sum_{K\geq 0}\sum_{o_1=1}^{N}\cdots\sum_{o_{K-1}=1}^{N}\left(\prod_{k=1}^{K}\left(W_{o_{k-1}o_k}/\exp\rho\right)^{\theta}\right), \tag{5}$$

where $W_{od} \equiv \exp\left(p_{o_{k-1}o_k} - c_{o_{k-1}o_k}\right)$ is the exponential of the trucker profits for a particular segment.

Let $\mathbf{A}$ be defined by the $N \times N$ matrix with ij^{th} element $a_{ij} \equiv \left[\left(W_{ij}/\exp\rho\right)^{\theta}\right]$. Define the $N \times N$ matrix $\mathbf{B} \equiv (\mathbf{I}-\mathbf{A})^{-1}$ as its Leontieff inverse, which is well defined as long as ρ is sufficiently large. Let b_{ij} be the ij^{th} element of $\mathbf{B}$. Intuitively, b_{ij} is the expected profits of the optimal trip from i to j; in contrast, a_{ij} is the profit from a shipment directly from i to j. Thus, the expected utility of a trucker residing in h can be rewritten as:

$$U_h = \frac{1}{\theta}\ln b_{hh}. \tag{6}$$

Similarly, the share of truckers from h that choose trip r is:

$$\pi_h(r) = \frac{1}{b_{hh}\exp(K\theta\rho)}\prod_{k=1}^{K}\left(W_{o_{k-1}o_k}\right)^{\theta}. \tag{7}$$

Equation (7) says that truckers are more likely to supply trips with greater expected payoffs.

Equation (7) allows us to calculate the *intensity* (number of trips) with which truckers from h serve any segment od, which we define as $S^h_{od} \equiv \sum_{r\in\mathcal{R}_h} \pi_h(r)\,\iota_{od}(r)$,where $\iota_{od}(r)$ is the number of times trip $r \in \mathcal{R}_h$ uses segment od. Hence, S^h_{od} is the supply of truckers from h to segment od. We can then calculate this intensity by fully enumerating all possible sequences where the B^{th} element in the sequence is o and the $B+1^{th}$ element is d, which yields:

$$S^h_{od} = \frac{1}{b_{hh}} \sum_{K\geq 0} \sum_{B=1}^{K-1} \prod_{k=1}^{B-1} \left(W_{o_{k-1}o_k}/C\right)^{\theta} \times \left(W_{ic}/C\right)^{\theta} \times \prod_{k=1}^{K-B} \left(W_{o_{k-1}o_k}/C\right)^{\theta} \iff$$

$$S^h_{od} = \frac{b_{ho} a_{od} b_{dh}}{b_{hh}}. \tag{8}$$

Equation (8) takes the form of a modified gravity equation. It says that the supply of truckers from h to od is increasing in (i) the profitability of shipping directly from o to d (i.e., a_{od}) and (ii) the expected profitability of the optimal trips to and from od (i.e., b_{ho} and b_{dh}). This second component—absent in standard trade models—highlights the role that truckers' home locations play in determining shipment patterns.

Explaining the stylized facts

We now show how this framework can explain the three stylized facts.

Stylized Fact 1 Regression (1) is the empirical analog (in logs) of the equilibrium trip intensity equation (8), where a_{od} is proxied by $traveltime^{\beta_1}_{od}\delta_o\delta_d$, where b_{ho} is proxied by $traveltime^{\beta_2}_{ho}$, where b_{dh} is proxied by $traveltime^{\beta_3}_{dh}$, and where b_{hh} is absorbed by the trucker fixed effect, implying $\beta_1, \beta_2, \beta_3 < 0$, as in Stylized Fact 1. Note that the empirical analog does not capture any variation in the profitability of the trips to/from od beyond the simple travel time from h to o and from d to h. However, the fact that we find β_2 and β_3 are larger in magnitude than β_1 is consistent with the expected profitability of the optimal trip to and from od being more responsive to travel time than the profitability of shipping directly from o to d.

Stylized Fact 2 From equation (7), the probability a trucker chooses any trip of length K is:

$$\sum_{r \in \mathcal{R}_{h,K}} \pi_h(r) = \frac{1}{\exp(K\theta\rho)} \frac{1}{b_{hh}} \left(\sum_{o_1=1}^{N} \cdots \sum_{o_{K-1}=1}^{N} \prod_{k=1}^{K} \left(W_{o_{k-1}o_k}\right)^{\theta} \right),$$

where $\mathcal{R}_{h,K}$ is the set of all trips of length K. Hence, as long as ρ is sufficiently large, the probability of choosing a trip will be declining in its trip length, as in Stylized Fact 2.

Stylized Fact 3 Consider two trips r and r' which both include the same segments, but in different orders, with r' incurring greater costs, $\sum_{k=1}^{K} \left(c_{o_{k-1}o_k}(r)\right) < \sum_{k=1}^{K} \left(c_{o_{k-1}o_k}(r')\right)$. From equation (7), we have that $\pi_r(r)/\pi_r(r') = \left(\exp\left(\sum_{k=1}^{K} \left(c_{o_{k-1}o_k}(r')\right) - \sum_{k=1}^{K} \left(c_{o_{k-1}o_k}(r)\right)\right)\right)^{\theta} > 1$. In other words, truckers are more likely to choose the lower cost ordering that serves the same segments, as in Stylized Fact 3.

Conclusion

The stylized facts presented above suggest that the transportation sector is much richer than is typically modeled, with truckers engaging in a complex optimization over possible trips they could take. The model presented above shows that, despite its complexity, the solution to this problem delivers analytical expressions for patterns of shipments. What remains to be shown is how these rich patterns of shipments interact with other economic forces to determine equilibrium trade flows and the equilibrium distribution of economic activity, a question tackled by the companion paper Allen, Atkin, Cantillo Cleves, and Hernández (2023).

References

ALLEN, T., AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129(3), 1085–1140.

——— (2022): "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 89(6), 2911–2957.

ALLEN, T., D. ATKIN, S. CANTILLO CLEVES, AND C. HERNÁNDEZ (2023): "Trucks," Working paper.

BRANCACCIO, G., M. KALOUPTSIDI, AND T. PAPAGEORGIOU (2020): "Geography, transportation, and endogenous trade costs," *Econometrica*, 88(2), 657–691.

MINISTERIO DE TRANSPORTE (2018): "Transporte en Cifras: Estadísticas 2018," Discussion paper.

SAMUELSON, P. A. (1954): "The transfer problem and transport costs, II: Analysis of effects of trade impediments," *The Economic Journal*, 64(254), 264–289.

SETHIAN, J. (1996): "A Fast Marching Level Set Method for Monotonically Advancing Fronts," *Proceedings of the National Academy of Sciences*, 93(4), 1591–1595.

YANG, R. (2022): "(Don't) Take Me Home: Home Preference and the Effect of Self-Driving Trucks on Interstate Trade," Discussion paper, National Bureau of Economic Research.