

Overland Transport Costs

A Review

A. Kerem Cosar



WORLD BANK GROUP

Transport Global Practice

August 2022

Abstract

Poor infrastructure and high domestic shipping costs are often cited as important impediments to economic activity in developing countries. Domestic shipping being mostly overland, understanding the level and structure of costs in road freight transportation could thus help formulate policies that aim to lower them. This review provides a

summary of overland transport cost estimates with a focus on trucking, the dominant mode of domestic freight. By describing conceptual issues, highlighting sources of data and alternative methodologies with their key findings, it is intended to help practitioners and researchers navigate the literature.

This paper is a product of the Transport Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The author may be contacted at kerem.cosar@virginia.edu.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Overland Transport Costs: a Review*

A. Kerem Coşar**

Keywords: Freight costs, road transportation, gravity regression.

JEL Classification: L9, F1, R4.

*This paper was prepared as a background paper for the Global Transport Practice Flagship Report *Shrinking Economic Distance*, led by Matias Herrera Dappe, Aiga Stokenberga, and Mathilde Lebrand. I thank them for helpful comments and feedback. For their insightful discussions of an early draft, I thank Roman Zarate and Théophile Bougna.

**Department of Economics, University of Virginia; e-mail: kerem.cosar@virginia.edu

1 Introduction

This survey reviews the empirical literature on economic costs of land-based freight transport. In addition to summarizing the measures and estimates of overland transport costs from the literature, it provides an overview of different methods and their data requirements, along with a critical assessment of the strengths and weaknesses of these methodologies.

The importance of road freight transportation is self-evident. While people and firms in developed countries take good roads and well-functioning transportation markets as a given, the situation is very different in low- and middle-income countries. Developed or developing, majority of trade is domestic in all countries, for which overland transport is the dominant form. In international trade, maritime shipping dominates in terms of weight and air shipping has an important share by value. Yet, land-based routes are also of prime importance for regional trade between neighboring countries. Therefore, road transportation plays a key developmental role by facilitating trade in short and intermediate distances.

To put the review in context, first note that the overarching question for economists and policy makers is to understand the magnitude and nature of *trade costs* that affect economic interactions between and within countries. All frictions that firms and consumers face in accessing markets and other economic agents can be considered within the broad definition of trade costs. These include not just the cost of shipping physical goods, but also the cost of finding suppliers and buyers, the cost of contracting, and in the international context, tariffs and non-tariff barriers. Focusing on shipping, all expenses incurred in moving freight from the factory gate to the ultimate retail or consumption point constitute the *cost of logistics*. Such itineraries may entail intermodal switching between modes of transportation, warehousing between different segments, multiple carriers between various nodes and the cost of coordinating all these interlocking activities. This survey focuses on a subset of this chain: *the cost of overland transportation* in a single segment via a single mode, i.e., using either a railroad or a motorized road vehicle.

An overland shipment may entail the crossing of an international border between adjacent countries. When there are border checks and controls, delays and inspections can add to the total cost of shipping. Similarly, to the extent that overland freight transportation regulations differ across administrative units within a country, internal borders could contribute to shipping costs. Although trade-reducing effect of domestic borders are well-documented—see Coughlin and Novy (2021) for a recent estimation using US data and a summary of the related literature—actual evidence that ties these estimates to transport related frictions is scarce. In a rare example, Carballo et al.

(2021) show a large positive impact of an expedited overland transit corridor between Central American countries on export values. In order to address purely transport related issues, the survey abstracts from institutional costs of crossing administrative borders. The exclusive focus is on overland freight costs within countries—or a single market such as the EU—that are due to one of the following factors:

1. **first-nature physical geography:** distance and topography.
2. **second-nature economic geography:** spatial distribution of population and industries.
3. **infrastructure:** roads, railways and bridges.
4. **equipment and supporting structures:** vehicles, gas stations.
5. **fiscal policies and regulations:** tariffs and taxes, licensing requirements.
6. **market structure of the transportation sector:** competition and mark-ups.

Note that these factors interact with each other in complex ways, which makes the measurement and estimation of transportation costs a challenging task. The state of technology and exogenous physical geography determine the production possibility frontier of transport services. Within that frontier, supply is determined by investments in infrastructure, equipment and supporting structures, taxes and tariffs on vehicles and fuel as well as regulations. Infrastructure may be insufficient or misallocated. For example, path dependence can render existing networks within-countries suboptimal as in the case of the mine-to-port structure of the West African railways built during the colonial era (Bonfatti and Poelhekke, 2017). Even within a developed market such as the EU, infrastructure across country borders may be under-supplied due to externalities and coordination issues that persist from the past (Felbermayr and Tarasov, 2019). The economic geography of a country, i.e., its spatial distribution of population and industries, is a key determinant of demand for transportation services through the level and direction of trade flows. Potential trade imbalances give rise to the backhaul problem and may generate asymmetries in shipping costs between origin and destination locations. In turn, transport costs and market access affect endogenous location decisions of firms and the long-run economic geography of a country. Finally, these interactions culminate in an equilibrium market structure where the level of competition and the scale of formal versus informal service providers determine mark-ups over carrier costs and thus the ultimate freight prices faced by shippers.

The objective of this survey is to summarize the extent to which the literature can delineate these components, along with the methodologies and data sources used in that effort. To focus on the costs faced by shippers, the survey abstracts from social costs that arise due to negative externalities such as pollution and noise. Since congestion externalities affect travel times through the interaction of geography and

infrastructure, they are indirectly considered as a determinant of transport costs. To focus on the estimation and measurement of overland transportation costs, the survey also abstains from the theoretical and quantitative literature studying the feedback of transport costs on economic geography and spatial development. For comprehensive reviews on these fronts, see Redding and Turner (2015), Berg et al. (2017) and Roberts et al. (2020).

2 Conceptual Issues

Before proceeding with specific methodologies and the related literature, it helps to enlist key conceptual issues and the properties that one would expect transport cost measures to satisfy. These ideal properties can then help guide the discussion of the widely used methods and sort out which particular application fits which criteria. Following a broader taxonomy suggested by Combes and Lafourcade (2005), there are three properties that an ideal measure of transport costs should meet:

- (a) **Itinerary:** cost measures should reflect the itinerary chosen between the origin and the destination of a shipment, where both road distance and travel time play a role.
- (b) **Mode of shipment:** costs should be specific to the mode of transport used, such as railway or motorized road vehicle.
- (c) **Commodity:** costs should be specific to commodity groups that have distinct unit weight and volume characteristics or special shipping requirements due to their shape or perishability.

The possibility of satisfying any of these criteria depends on the data at hand. The importance of each in turn depends on the setting that policy-makers and researchers are studying. If the objective is to gauge the cost of congestion in certain arteries deemed critical for economic activity in a country, or the benefits of alternative infrastructure projects, measurement and data could focus on (a) and (b). If the evolving commodity structure of a country is of primary interest, one may want to seek or collect data that is also informative on (c).

For practical purposes, the relevant mode of shipment in a developing country context is trucking. With the exception of India, railroads typically have a small share in the transport system of developing countries since their infrastructure investments occurred mostly during the second half of 20th century after the motorized transportation revolution. For example, Gwilliam (2011) reports the dominance of roads in Africa which carry 80-90% of passenger and freight traffic. Therefore, the

main emphasis in this survey is on direct truck shipments.¹

Another important consideration is the choice of units in which transportation costs are denominated. In reality, service providers base their pricing decisions on either the weight or the volume of cargo, or a formula taking both aspects into account. In addition, specialized vehicles may be required depending on the nature of the cargo. Bulk goods that are shipped loosely (e.g., minerals, grains or liquids) can be transported in tanker trucks or dump trucks. Goods that can be shipped in packages or containers can be hauled by dump trucks or semi-trailer trucks. If the cargo itself is wheeled, such as vehicles, a car-carrier trailer may be used. Each case will differ in the way that providers calculate their costs and quote prices.

What shippers care about in assessing their profit margins is the per-unit shipping cost. Given a quote for freight by weight or volume, they can calculate the per-unit cost and determine the delivered cost of a good. Suppose that a transport provider quotes a total charge $T(Q_{od})$ to ship Q_{od} quantities of a product from the origin location o to destination d . Let c_o be the free-on-board (FOB) factory-gate unit cost of the good being shipped at the origin. Regardless of whether the buyer or the seller is paying for transportation, the delivered cost c_d at the destination can be written as

$$c_d = c_o + \frac{T(Q_{od})}{Q_{od}}.$$

This additive cost translates into a multiplicative ad-valorem form by simply taking the ratio

$$\tau_{od} = \frac{c_d}{c_o},$$

such that $\tau_{od} - 1$ is the % shipping cost. Survey-based methods may inquire shippers about either or both of these forms. In the absence of direct information about per-unit costs, estimation-based methods typically assume an ad-valorem percentage cost structure on transportation. In what follows, I will discuss the implications of these choices on making accurate predictions about the effect of transport costs on the volume and composition of trade.

The final conceptual issue to bear in mind is the potential asymmetry in the cost of shipping a certain amount of the same product between two locations depending on the origin. Suppose that direct evidence or inference yields some information about transport cost from location o to d in the $o \rightarrow d$ direction. Does the same cost apply for the reverse flow? That is, are transport costs symmetric such that $T(Q_{od}) = T(Q_{do})$? The well-known *backhaul problem* in transportation suggests that whenever there are

¹A rare recent example of a railroad project with a high potential to reduce transport costs in a developing country is the Addis Ababa-Djibouti Standard Gauge Railway that connects landlocked Ethiopia to the Port of Djibouti.

sizable trade imbalances between regions, trade costs are likely to be asymmetric. This is a simple demand-meets-supply outcome: total trade in one direction constitutes demand for shipping. If a truck is operating on a back-and-forth basis between two places, it has the same capacity (i.e., supply) in both directions. Given uniform supply, the cost should be higher in the direction featuring greater quantities being shipped on total.² In the extreme case, suppose that a truck company transports a certain freight from o to d but expects the vehicle to be empty on the return trip. The pricing of the o -to- d fronthaul will incorporate the additional cost of the return trip. This will be a systematic pattern if there is more freight on total from o to d than in return. Then, any potential shipper from d to o will enjoy lower shipping costs thanks to the competition between truckers that have to go back and will be doing so without cargo otherwise.

In principle, it is possible to circumvent the backhaul problem through trihaul routing ($o \rightarrow d_1 \rightarrow d_2 \rightarrow o$) or more complex itineraries. This is, however, complicated by the cost of obtaining ex-ante information for shipping demand in multiple locations and the uncertainty of finding shippers in spot markets (Brancaccio, Kalouptsidi and Papageorgiou, 2020). Applications of on-demand ride technologies to freight may alleviate some of these costs. In the relevant near future, incorporating potential asymmetries and their causes to transport cost estimates is important for accurately predicting how transport costs affect different regions.

With the conceptual and practical issues in mind, the next three sections introduce the three most common methods to measure and quantify transport costs. These are

- i) survey-based methods,
- ii) imputation-based methods,
- iii) estimation-based methods.

I start with direct measurement from surveys.

3 Survey-based Methods

The best method to measure a phenomenon is to collect direct quantitative data about it. This goes for transportation costs as well. Since a survey can target its sample across different types of firms and entities, it is also the only avenue to potentially obtain information on the price of road transport paid by end users separate from its

²Wilson (1987) models and estimates the price structure using trucking data between a limited number of regions within the U.S., documenting the prevalence of backhaul problem. This same effect has been documented for international maritime shipping by Wong (2022).

cost to the providers of transport services.³

Collection of high-quality survey data requires well-established sampling and stratification procedures on the universe of shippers or carriers in a country. Conducting it in a comparable manner over regular intervals would help inform changes over time. Such capabilities typically fall within the purview of national statistical agencies. Examples from developed countries showcase the variation in the information that is available depending on survey design. In Canada, the Trucking Commodity Origin and Destination Survey collects information on for-hire trucking industry since 1994, targeting trucking companies with at least one establishment and a minimum annual revenue of about 1 million USD. It excludes foreign-based trucking establishments operating in Canada and private fleets of non-trucking establishments. While it contains data on the revenue of carriers on a shipment basis (including information about the commodity being shipped, weight, distance and destination), it does not report the value of goods shipped.⁴ In contrast, the Commodity Flow Survey conducted by the U.S. Census Bureau since 1993 reports the commodity, mode (truck, rail or other modes), weight and travel distance of shipments together with the value of the goods, but does not report freight charges. In Europe, Eurostat’s Road Freight Transport Survey is conducted in EU member countries and Norway since 2012. For around 3 million shipments, it contains information on origin and destination regions, industry of the shipper, the weight of the shipment and the distance covered. While separate vehicle-related, journey-related and goods-related variables inform users about the fleet, itineraries and commodities being shipped, there is no data about freight charges or the value of the goods. For rail shipments, the Carload Waybill Sample is a stratified sample of carload waybills for all U.S. rail traffic collected by the Surface Transportation Board. It includes information about the commodity, origin and destination, weight and freight charges from the sampled shipments. Evidently, each lacking some critical variables, whether any of these surveys is helpful to the user depends on the question at hand. Yet, as will be detailed below, they can be valuable inputs to other methods for imputing or estimate freight costs.

Given the dearth of high-quality surveys conducted and disseminated by national agencies in developing countries, practitioners and researchers combine various datasets on truck fleets and trade flows within countries to get a picture of road transportation—see Allen et al. (2021) for a recent example. An alternative is to conduct a special survey from scratch, possibly with the help of a national statistical agency in sampling from

³In principle, balance sheets of transport providers would distinguish their revenues and costs. In practice, however, most transportation providers in developing countries are not publicly listed or even incorporated. Such owner-operators are unlikely to follow rigorous accounting principles and prepare balance sheets.

⁴Behrens et al. (2018) convert freight costs reported in this survey to a percentage ad-valorem measure by using additional information on the value of cross-border shipments by exporters in one wave of the survey.

the underlying population of shippers or carriers. Notable examples collecting survey-based data on trucking costs in developing countries are Teravaninthorn and Raballand (2009) for Africa, Osborne et al. (2014) for international trade corridors in Central America and Lam et al. (2019) for Vietnam.

Naturally, survey design and sampling presents researchers and practitioners with multiple choices to be made such as whether one is interested in trucking costs throughout the entire country at both urban or intra-city scales, or just in some key corridors. The studies cited above exemplify different choices on this front. To facilitate comparison and guide standardized data collection, survey data is typically collected for a fixed freight scenario that is representative for a typical shipment, such as a 20 ton trailer truck. Many countries have separate fleet inventory surveys that can guide this choice. Similarly, respondents are asked to consider a scenario involving commodities that satisfy certain criteria, such as not being irregularly shaped. The length and detail of the survey is another key design choice: facing a trade-off between detail and response rates, a common approach is to complement questionnaires with in-depth interviews from a small number of respondents.

Most studies in the literature aim to quantify several variables for the case under study: the ton-mile cost of truck shipments to the carrier, the price charged to shippers, the composition of costs to variable versus fixed costs, capacity utilization rates and total vehicle operation costs. As expected, relative and absolute costs differ across countries, but some consistent results also emerge from a range of studies: in developing countries, a larger share of total costs in trucking is borne by variable costs since older vehicles with low fuel efficiency have a larger share in the national fleet. In France and the U.S., two developed economies for which comparable data exists, the cost composition is the reverse, with a higher fraction accounted for by fixed costs. This finding resonates with firm-level models of technology choice that posits a trade-off in operating with a low marginal cost technology at a high fixed cost, or vice versa.

One concern in asking truck operators about their variable and marginal costs is that economic agents do not always explicitly or accurately calculate these economically relevant variables. When asked about them on the spot, they may give noisy guesses. They are, however, more likely to be well-informed about the levels of separate cost components. Therefore, information collected through surveys can be leveraged as an input to imputing transport costs. For example, a survey can simply ask operators about hourly wages of drivers, the cost of vehicle maintenance and the fuel efficiency of their fleet with the goal of precisely measuring each component. Putting all this information together, under assumptions on the importance of each component—which itself can be informed by surveys or interviews—the researcher then imputes freight costs per ton-mile. Since this approach goes one step beyond surveys, the next section

describes its key steps and data requirements in more detail.

4 Imputation-based Methods

Calculating unit costs for any economic activity is challenging when there are fixed and joint costs. Hence, even if surveys collect direct information on carriers' self-reported costs, one can still be skeptical about the degree to which they capture underlying true costs. It is therefore common practice in economics to estimate unit costs assuming that producers minimize costs. As long as there is data on key cost components such as driver wages—which can be proxied by other low-skilled occupational wages—fuel, and vehicle depreciation, a cost minimization routine can be used to impute optimal itineraries and associated costs.

Over the recent past, increased availability and diffusion of Geographic Information System (GIS) data and software enabled researchers to calculate least-cost routing decisions and associated transit times using standard computers. The starting point is a digitized transportation network. Ideally, this should encompass all relevant infrastructure modes (rail vs road) and road types (highway, two-lane, paved, gravel) linking the nodes that represent key centers of economic activity in the economy under study. Information about travel speed along each link helps calculate fastest routes and associated travel times between nodes. A related choice for the determination of the nodes is one of geographic aggregation. Researchers typically use centroids of metropolitan or administrative areas of various size. Another possibility is to impose a high-resolution grid. While doing so at a high geographic detail increases precision, it comes at the cost of steep computational times. Well established algorithms and numerical methods help to resolve the trade-off. Dijkstra's algorithm and fast-marching method are widely used to find shortest or least-cost paths between nodes in a road transport network (Allen and Arkolakis 2014, Donaldson 2018).

Beyond the technical aspects of route optimization, formulating the relevant cost minimization problem requires making assumptions on how carriers operate: do they chose fastest or shortest routes? What is their willingness to pay tolls if that option is available to reduce travel distance and duration? Do they make these decisions independently of shippers or do they offer a menu of shipping times and prices? Do idiosyncratic demand effects on timeliness matter for routing decisions? For each scenario, an appropriate model of minimization can be used to impute times, distance and routes. Additional external information about variable and fixed operation costs would help transform the imputed units from travel distance and time to actual monetary costs.

Instead of calculating routing and travel times as described above, new GPS

technologies allow observing actual itineraries and trip times. A recent example is Hernandez (2021) who uses tracking data from GPS devices located in trucks operating within Colombia. With data on locations and time stamps of 15,000 trucks in 186,000 long-haul trips, he calculates waits and delays in road freight transport. In an application to China, Alder et al. (2021) estimate congestion by road segments using real-time GPS information on 1.8 million (20% of the total) long-haul trucks in 2020. Such GPS data, however, may not be readily available for researchers and practitioners. Therefore, node-to-node least-cost path calculations remain the most viable option for a wide range of applications.

Two examples for imputing domestic transport costs are Allen and Arkolakis (2014) and Donaldson and Hornbeck (2016). Using the fast marching algorithm on a detailed map of the transport network across continental U.S. counties, together with information on trade flows from the Commodity Flow Survey, Allen and Arkolakis (2014) embed a mode-specific transport cost minimization to the estimation of general trade costs. Donaldson and Hornbeck (2016) focus on transportation costs alone. While their goal is to analyze the impact of railroads in 19th century U.S. economy, the methodology they follow in constructing a county-to-county transport cost matrix can be applied to other settings. Using information on ton-mile cost of shipping via all available modes—rail, wagon and waterways—and assumptions on transshipment costs between modes, they calculate lowest-cost freight transportation routes and the associated costs in levels between all U.S. counties before the advent of motorized travel.

In a more recent application, Combes and Lafourcade (2005) calculate transport costs between French economic zones in 1978 and 1998. Focusing attention to road transport, they distinguish various road types (tolled and free highways, secondary roads and urban roads). Additional realism comes from separately incorporating itinerary and distance-related costs (fuel, tolls, vehicle and tire depreciation and maintenance) from time-related costs (driver wages and accommodation expenses, insurance and loading/unloading times). Since both are expressed in the same monetary units, the authors can compute general transport costs in levels between any two nodes within France by minimizing the total of distance- and time-related shipping costs. The information used in this calculation comes from surveys (e.g., maintenance costs), administrative data (e.g., driver wages) and price statistics on intermediate inputs used in trucking (e.g., fuel).

In a similar exercise for Spain, Zofío et al. (2014) impute distance (per km) and time (per hour) costs for road freight transportation in Spain at the NUTS-3 level of geographic aggregation. As Combes and Lafourcade (2005), they consider a representative 40 ton articulated truck, which accounts for about 80% of road freight transport in Europe. They convert time costs (Euro/hr) to the same unit as distance

costs (Euro/km) by dividing it by average speed. Their imputation suggests a per-km cost of 1.23 Euro/km in 1980, which fell to 1.02 Euro/km in 2007. Their methodology helps to decompose the drop to its productivity and infrastructure capital components. Persyn et al. (2020) apply the same methodology to all of Europe starting from a one-square km resolution. They sample centroids for each NUTS-2 European region based on the spatial population distribution.

Imputation-based transport cost calculations lend themselves to practical applications. For example, in an effort to give shippers a benchmark for freight costs, the Colombian Ministry of Transportation developed a simulation tool called *SICE-TAC*. Using this web-based application, shippers can obtain trucking cost estimates for all routes within the country.⁵ The main data used in this application comes from a survey of per-unit freight prices paid to truck owners at the origin-destination-commodity level. The developers used an imputation-based approach to calculate the determinants of unit costs based on the characteristics (such as distance and slope) of sampled routes. Through a GIS interface that contains information about these characteristics in the entire road network of the country, the simulation tool then yields predictions on monetary costs on any hypothetical route. As in this example, data collection efforts can provide useful inputs to imputation-based calculations that generalize cost calculations to a wider geography.

While imputation-based methods can predict monetary shipping costs across hypothetical routes, they cannot inform how these costs affect shipping demand and trade between locations. Doing so requires additional conceptual tools discussed in the subsequent section.

5 Estimation-based Methods

Before delving into the estimation of trade costs, it is in order to note that the methodology discussed in the previous section provides some of the key variables to be used. In what follows, it is assumed that the travel distance and travel time variables are the output of a network-based optimal itinerary/route calculation along the lines described above.

5.1 Gravity Approach: Estimation Using Trade Flows

Empirical studies of trade have long established that bilateral exports between countries or regions within countries are proportional to their economic size and inversely

⁵Information in Spanish available in <https://www.mintransporte.gov.co/publicaciones/4462/sice-tac/>

proportional to distance. Its form being reminiscent of Newton’s law of universal gravitation, this empirical relationship has been labeled as the ‘gravity equation’ in trade. More recently, economists have shown that a wide range of trade models generate the gravity equation as an equilibrium outcome. The structural approach to parameterize these models through the estimated distance coefficient allows researchers to quantify the effect of trade costs on real incomes of the economic units under study. Head and Mayer (2014) provide an extensive overview and summary of this literature.

The fundamental equation for estimating a gravity model to explain trade flows X_{od} is

$$\ln(X_{od}) = \gamma_o + \gamma_d + \sigma \cdot \ln(\tau_{od}) + \epsilon_{od}, \quad (1)$$

where σ is the elasticity of trade to trade costs $\tau_{od} \geq 1$. Origin and destination fixed effects γ control for size and productivity differences across locations. If trade costs are already specified in monetary terms, then one can directly use $\tau_{od} = TC_{od}$. Otherwise, if one is to use shortest travel distance or time, a functional form should be used to transform travel distance $DDist_{od}$ or travel time $TTime_{od}$ to trade costs. Suppose $DDist_{od}$ is being used. There are two common specifications:

$$\tau = \exp(\theta \cdot DDist), \quad (a)$$

or,

$$\tau = DDist^\theta, \quad (b)$$

which means that when τ is substituted into the gravity equation (1), distance as an explanatory variable may appear either in levels or in logarithms, depending on whether functional form (a) or (b) is used.

If the data is spatially aggregated so that there are observations with $o = d$ capturing trade flows within geographic units, one needs to make an appropriate normalization for travel time in both specifications. For the *log – linear* specification in (a), this would be $DDist_{oo} = 0$, which is invariant to time unit being used. For the *log – log* specification in (b), however, one cannot use $DDist_{oo} = 1$ since it is not invariant to the choice of units. Therefore, the preferred method is to first calculate a measure of internal travel time for each region o , and then to normalize the entire $DDist_{od}$ matrix by the smallest value $\min_o\{DDist_{oo}\}$. This can be done by using one of the within-unit distance measures for $DDist_{oo}$ described in Head and Mayer (2010).⁶

Regardless of the specification of τ , an important result is the ensuing identification problem for estimating the distance elasticity θ . To demonstrate this without loss of

⁶In practice, depending on the geographic aggregation, it is possible that there may be $o \neq d$ pairs with a shorter distance than the smallest $DDist_{oo}$. In this case, one can preserve the consistency condition $DDist_{od} \geq 1$ of the normalization by simply truncating the distance measure at one.

generality, substitute τ from (b) to gravity equation (1):

$$\ln(X_{od}) = \gamma_o + \gamma_d + \underbrace{\sigma\theta}_{=\delta} \cdot \ln(DDist_{od}) + \epsilon_{od}. \quad (2)$$

Estimating equation (2) identifies δ , the product of two parameters σ and θ . To recover the parameter θ , one has to assume a value of σ to calculate $\theta = \delta/\sigma$. For aggregate trade flows, typical values from the literature range between 5 and 10 (Anderson and van Wincoop 2003, 2004). Note that the impact of varying travel distance or travel time on trade flows is still informed by the combined elasticity δ . However, if one is seeking to quantify the elasticity of trade costs to travel distance or time, the correct value is θ .

The minimal data needed for this estimation is trade flows between locations within a country. Ideally, these locations should be spatially disaggregated so that most flows are across regions for which one can construct fine measures of distance or travel time. Otherwise if data is spatially not disaggregated enough and locations are fairly large, one has to take a stance on distance for within-regional shipments. One common approach is to use a proxy for internal distance (Head and Mayer, 2010). However, this ignores the distribution of economic activity within the region, and would be biased if it is concentrated in a few locations.

High quality data on domestic trade flows typically exists in developed or middle income countries. These are compiled from various sources. Some are due to specialized surveys, such as the *Commodity Flow Surveys* in the U.S. or in Colombia (*Encuesta Origen - Destino a Vehiculos de Carga*). While there is a voluminous literature making use of the former, few studies analyzed road transport costs using the Colombian data. Estimating a gravity framework, Duranton (2015) finds that a 10% increase in the travel time between Colombian cities reduces the value of trade between them by about 7%. To gauge whether within-city roads affect trade, he constructs a municipality-level road index summing up the log of the mileage of principal roads and the log of the number of exists from these roads. A 10% increase in a city's road index increases the value of its exports by about 4%, suggesting that congestion may be an important factor for transport costs.⁷

In some developing countries, comprehensive domestic trade flow data may not be available. All countries, however, collect customs level international trade data. In many cases, these records contain information about exporting/importing firms, their locations in the country, and the port of export/import. Linking these two provides researchers with a snapshot of trade flows within the country, despite the restricted and

⁷A similar data source from India—*Inter-State Movement/Flows of Goods by Rail, River and Air*—does not contain information about motorized shipments on roads.

biased nature of the data. It is restricted in the sense that it doesn't cover trade between all the bilateral region pairs. Rather, it only contains flows from regions where exporters and importers are situated to regions where ports are. Being based on shipments of exporting firms could bias estimates of trade costs as these firms are more productive than non-exporting firms. Under the assumption that regional distribution of exporters is not endogenous, one can still identify relative trade costs across regions. Since this may be the only data source available for many countries, it is still a useful venue for researchers to pursue. Recent examples are Van Leemput (2021) for India, Cosar and Demir (2016) for Turkey, Baldomero-Quintana (2020) for Colombia and Fan, Lu and Luo (2021) for China.

More recently, a number of countries have made available administrative databases that contain information about firm-to-firm transactions. While such transactions are not direct evidence for cargo shipments—even if they were related to freight, they do not contain information about the characteristics of the goods involved—they inform researchers about trade flows between locations. Such links can be very granular if the addresses of firms or establishments are also observed. Otherwise, they can be aggregated to a regional level to create an internal trade flow matrix. In a recent application using Turkish data, Coşar et al. (2021) estimate the impact of road capacity (i.e., lanes) and quality improvements. According to their results, a one-hour reduction in travel time between two districts increases bilateral trade between those districts by around 8.2 percent. This effect translates into an almost 1 million USD increase in trade flows for a typical supplier district over 10 years. If two districts were not trading prior to the road investment program, a one-hour travel time reduction increases the probability that they will start doing so by 11 percent.

In estimating the elasticity of transport costs to travel distances, a threat to identification arises from the fact that travel distances emanate from infrastructure placement, which may itself be endogenous to expected trade potential. Endogeneity bias may affect the ordinary least squares (OLS) estimates in either direction. If selective road investments prioritize routes with the highest trade potential, the OLS estimate in a regression of trade flows to driving distance will be biased upward. Since endogenous placement lowers driving distance where trade flows are higher, this bias generates the impression that trade flows are very sensitive to transport costs. On the other hand, if road investments follow distributional concerns and prioritize disadvantaged locations whose exports may still lag behind, the downward bias generates the impression that trade flows are not sensitive to transport costs.

What can be done to overcome this threat? To some extent, geodesic straight line distances capture the exogenous variation in travel distance due to geography but are not directly relevant for estimating how trade flows respond to actual transport

networks. Finding an exogenous source of variation is challenging. A good example is Martincus and Blyde (2013) who exploit the partial destruction of the Chilean road network in the 2010 earthquake as a natural experiment. Using establishment-level export flows from geo-referenced origins to ports, they first impute distance and time components of road travel costs before and after the earthquake with a methodology similar to the one described in the previous subsection. In this imputation, they utilize a survey on the operational transport costs of land cargo services (*Encuesta de Servicio de Transporte de Carga por Carretera*) conducted annually by the Chilean statistical office INE. The exogenous variation in these imputed costs induced by the earthquake provides a basis for estimating the gravity equation (2) in differences.

An alternative in addressing endogeneity concerns related to road placement is to use the historical routes as an instrument for present day investments. Under the exclusion restriction assumption that historical networks affects current trade only through the determination of current roads, this approach helps correct the bias that may plague OLS estimation. In an example from the U.S. context, Duranton et al. (2014) use exploration routes between 1528 and 1850 and railroad routes in 1898 as instruments for the Interstate Highway System. Banerjee et al. (2020) use historical transportation corridors connecting large old cities and 19th century Treaty Ports in China as instruments for modern paved roads.

To sum up, the gravity estimation relies on the presence of high quality domestic trade data which may be hard to obtain in developing countries. An alternative method is to infer trade costs from interregional gaps in prices, which is more feasible since national statistical agencies typically collect prices of various commodities with the primary intent of calculating the inflation rate. I describe this methodology next.

5.2 Estimation Using Price Differentials

Spatial price gaps are informative about trade costs. Suppose the price of a good in location o is p_o , and its price in location d is p_d . Assuming that there is a competitive trading sector, researchers invoke the following condition to estimate transport costs t :

$$p_d = p_o + t_{od}.$$

This equation implies that at the prevailing observed prices, a trader who buys a good at the origin and incurs transport costs to sell it at the destination makes zero economic profits (no-arbitrage condition). Specifying t as a function of distance, one can obtain an estimating equation similar to equation (2) with the dependent variable being price differentials rather than trade flows. Price differentials can be in absolute value $|p_o - p_d|$ or a log difference $\ln(p_o/p_d)$. The implicit benchmark in this approach is the law-of-

one-price (LOOP): in the absence of trade costs, prices of identical goods should be equalized across locations. Due to differences in the cost of non-tradables (e.g., rents) that affect retail prices, one would typically not expect the strong form of LOOP (equal price levels) to hold. The weak form of LOOP allows prices levels to differ but contends that prices should co-move. Using a panel of prices, the typical estimating equation for this method is

$$V_{od} = \gamma_o + \gamma_d + \beta \cdot \ln(dist_{od}) + \epsilon_{od}, \quad (3)$$

where the dependent variable is the variance (or standard deviation) of log price differences $\ln(p_{ot}/p_{dt})$ over time. In using this specification, the main focus of the literature has been explaining cross-country price differences and real exchange rates (Engel and Rogers, 1996). In these applications, researchers use a panel of prices from various regions/cities in two countries, and include a border dummy to distinguish location pairs within the same country from those that are in separate countries. The same method can also be used to gauge the so-called ‘border effect’ across administrative regions within a country as Borraz et al. (2016) do in the context of Uruguay. In a recent contribution,

Before introducing in detail a notable exception that estimates intra-country trade costs from spatial price differentials (Atkin and Donaldson, 2015), let me first discuss methodological issues that researchers have to be mindful of while using this approach. The first challenge is the high data requirement to obtaining prices of *identical goods*. In recent years, prevalence of bar-code level scanner price datasets have enabled researchers to circumvent this issue and use price dataset containing information for a high number of consumer goods. Such data, however, is typically available for a handful of developed markets. Another recent innovation is to scrape online prices, but this also remains limited to countries where online shopping has a non-trivial market share (Cavallo and Rigobon, 2016).

Second, even with a competitive trading sector, the absence of arbitrage or perfect price comovement does not imply this condition to hold with equality. Rather, the true no-arbitrage condition consistent with competitive trading is $p_d \leq p_o + t_{od}$; that is, buying a good at o , and incurring transport cost t_{od} should not enable the trader to profit. For instance, in a competitive market, there may be no trade between two locations and the condition may hold with inequality. Since estimation requires equality, while the true condition may be a strict inequality, price differentials only provide a lower-bound for trade costs. The problem can be alleviated if the researcher knows that there is trade, which would unambiguously be the case if one of the locations is where production takes place. The third and perhaps the main challenge is the perfect competition assumption itself. The presence of market power and spatial variation in market structure implies that observed price gaps will reflect not just trade costs but

also mark-up differences (Cosar, Grieco and Tintelnot, 2015).

A rare attempt to overcome these challenges by Atkin and Donaldson (2015) applies the price-gaps methodology to estimate trade costs in Ethiopia and Nigeria. The authors use data on prices of a number of staple consumer goods collected by statistical agencies for constructing the consumer price index, addressing the first challenge. Complementing the price data with the origin information for the goods, they overcome the second challenge.⁸ Moving away from perfect competition and featuring oligopolistic intermediaries, their model addresses the third challenge. Finally, to allow a comparison to a developed country, they also estimate trade costs using U.S. data.

The results suggest an economically large difference between trade costs in the two African countries relative to the US. The costs of trading goods to the most remote compared to the least remote location is 9 cents in Ethiopia and 13 cents in Nigeria. In the US, the same distance costs only 2 cents. In terms of the marginal trade cost with respect to geodesic distance, the cost is 3.53 times higher in Ethiopia and 5.26 times higher in Nigeria than in the US, respectively.

Noting that these estimates encompass not just transportation but other costs of trading, a natural question for the purposes of this review is how to unpack the contribution of freight costs to this inter-country difference. In one exercise, the authors replace their geodesic distance metric with the fastest travel time measure. Since the U.S. is better endowed with higher quality roads, a change in relative costs is informative about the role of infrastructure. As a result, the estimated marginal cost difference between the two African countries and the U.S. drops to 2.46 for Ethiopia and to 4 for Nigeria. Comparing with full costs at face value, infrastructure explains about 24-30% of these countries' cost gap with the U.S. While this is not a trivial magnitude itself, it is likely to be a lower bound for the impact of infrastructure since omitted factors such as the cost and quality of transportation equipment, and fuel costs, are likely to interact with road quality.

Although the countries under study differ, it is noteworthy to compare the estimates from Atkin and Donaldson (2015) with the direct evidence from trucker surveys by Teravaninthorn and Raballand (2009), as the authors also do. The latter is reported in terms of per km cost for one truckload shipment, but in comparison, both the estimates and survey-based actual costs are expressed relative to the U.S. Estimated marginal cost of shipping relative to the U.S. mentioned above (2.56 and 4 in Ethiopia and Nigeria, respectively) are in the ballpark with the results from Teravaninthorn and Raballand

⁸Origins are either identifiable production plants within each country or the main international port for imports. Asturias et al. (2019) follow a similar strategy and use prices charged by monopolists in different destinations to estimate transport costs in India from price differentials. Reduced price gaps after the construction of the Golden Quadrilateral highway are informative about the pro-competitive effect of reduced transport costs. Similarly, Donaldson (2018) uses price gaps in salt, a commodity produced in unique locations, to infer transport cost reducing effect of railroad construction in 19th century India.

(2009) for the Mombasa-Nairobi corridor in East Africa (1.88 times that in the U.S.) and for the Bamako-Accra corridor in West Africa (3.28 times that in the U.S.).

6 Bridging the Methodologies

I finish the review with a comparison of methodologies laid out so far and a discussion of how they can be used in tandem. Table 1 provides a summary of data requirements, types of transport costs inferred and challenges for each.

As described above, estimation based methods recover trade costs at large, including frictions other than transportation costs. In their review of international trade costs, Head and Mayer (2013) make an accounting exercise using observable international shipping costs,⁹ and a range of reasonable parameter values for the elasticity of trade to trade costs. Their results suggest that "dark trade costs," i.e., what cannot be accounted for and is therefore a residual, explains between 50-85% of the effect. Similar to the literature trying to explain per capita income differences across countries using observable capital and labor stocks, a large unexplained residual remains as the measure of our ignorance. In the context of domestic trade, this residual component is expected to be smaller since some of the border-related frictions are absent. Still, an expected gap is an important point to keep in mind for practitioners when they compare estimated trade costs with survey-based or imputed transport costs.

Such comparisons, however, are valuable in that they provide benchmarks across studies. Similar to the practice by Atkin and Donaldson (2015) described above, Porteous (2019) leverages information generated by different methodologies. After estimating trade cost for agricultural commodities in Africa, he compares the values for the corresponding corridors with the trucking costs from Teravaninthorn and Raballand (2009). His trade cost estimates are 50 to 100% higher than survey-based freight costs, which, per the discussion above, is an expected outcome since price-differentials capture more than transport costs. The fraction accounted for freight is informative about what policy makers can hope to achieve in terms of boosting trade and economic activity with policies and investments that can affect transport costs in the short run.

Despite its shortcomings listed above, it is the structural gravity model that enables predicting the GDP impact of particular transportation projects. A data challenge in linking gravity-based and survey-based measures in a developing country context is the fact that data on within-country trade flows is generally lacking in such environments. It is precisely for this reason that notable attempts to estimate trade costs in Africa cited above (e.g., Atkin and Donaldson, 2015; Porteous, 2019) use price data. Prices,

⁹To estimate transport and insurance costs, they use the difference between reported cost-insurance-freight (CIF) and free-on-board (FOB) prices in customs records.

especially for agricultural commodities, are collected systematically in these countries. While this provides a snapshot of costs for widely produced commodities, it is restrictive in quantifying the potential implications of transport cost reductions on trade flows and welfare. One way to bridge this gap and generate simulated trade flows is to start by surveying firms from various sectors of interest in order to estimate their freight demand, as Herrera Dappe et al. (2019) do for Bangladesh. In the absence of a comprehensive nationwide commodity flow survey, the authors sample four thousand economic establishments across the six largest freight-intensive sectors and across the country in order to understand where freight is likely to be originated. As a secondary data source informative about flows, they use truck traffic measurement from key arteries. Fitting observed traffic to the Freight Origin-Destination Synthesis (FODS) model of Holguin-Veras and Patil (2008), they estimate origin-destination flows. They complement this with links between ports and exporting or importing firms obtained from customs data, as discussed in Section 5.1. The fitted model facilitates the construction of a domestic trade flow matrix for Bangladesh.

A common limitation of all the approaches discussed so far is their inability to separate transport costs incurred in the first- and last-mile from those incurred in the main haul. The sizable share of first- and last-mile costs are well-known in transportation economics. However, unless explicitly asked, most surveys do not differentiate them from total shipment costs, thereby missing a potential source of non-linearity in the transport cost function. A recent example overcoming this shortcoming by synthesizing multiple data sources and methodologies is Kebede (2019). Studying of the impact of the Universal Rural Road Access Program in Ethiopia, he uses survey-based information from the Ethiopian Socioeconomic Survey on costs incurred by farmers while transporting recorded agricultural commodities from the farm-gate to the nearest market town. For an average distance of 12 km traveled, farmers report ad-valorem transport costs of 11.4% on average, with the median at 6.5%. Kebede then uses responses at the village level to estimate the cost of transport by road types (paved, gravel, cobbled or earth roads). Additional information comes from standard price data collected from farms as well as main market towns, which is informative about the potential share of transportation in overall trade costs estimated from price differentials. These costs are then used to calculate least-cost itineraries and shipping cost measures throughout the entire road network, which are in turn used to impute a model-based market access measure for a quantitative welfare exercise.

In Section 2, I discussed the important question of whether transport costs are per-unit (additive) or ad-valorem (multiplicative). Both anecdotal evidence as well as direct measurement from surveys suggest that the former is a better characterization of transport costs. Imputation based methods are designed to deliver additive per-

unit costs. Estimating trade costs from trade flows with the standard iceberg cost specification, however, yields multiplicative costs. Then, a natural question is how much this discrepancy matters for correctly predicting the impact of transport cost changes. Bergquist et al. (2019) provide an insight on this question. After estimating additive transport costs in levels from price gaps, they use a model-based approach to show the bias in average and distributional effects had they used an ad-valorem specification. Another source of potential bias from such misspecification comes from the well-known Alchian-Allen effect, or “shipping the good apples out” which implies that the quality and thus the unit cost c_o of a traded product may itself be a function of per-unit transport cost. The reason is that given a per-unit cost t_{od} , selling a high-quality version of a product with higher c_o lowers the percentage cost τ_{od} .¹⁰ A change in per-unit transport costs will impact not only the quantity but also the quality and thus the unit price of shipments. A framework that assumes ad-valorem costs will then yield biased estimates for the responses of trade values and real incomes to transport cost changes.

Another important consideration in using existing data sources is to better incorporate the market structure and endogenous determination of mark-ups in trucking. This is critical for separating prices from costs and gauging the role for policies that could increase competition. Trucking sector is known to display a dual structure in many developing countries, with informal owner-operators forming a competitive fringe and formal, modern carriers operating large fleets.¹¹ In the segments that they serve based on routes and service quality, the latter group typically commands market power. Recent work by Allen et al. (2021) presents novel stylized facts on market concentration in trucking from Colombia, which inform a spatial imperfect competition model of carriers. When transport costs are high, they could be further magnified through endogenous entry decisions of carriers and high mark-ups. In a more general setting, high trade and transport costs could also affect market structure in downstream markets. Using market and farm-gate agricultural prices in India, Chatterjee (2020) estimates the role of spatial frictions in increasing the market power of intermediaries. Bergquist and Dinerstein (2020) provide similar evidence from Kenya using a randomized controlled trial (RCT) design.

Given the importance of data collection for the subject matter of this review, it would be befitting to conclude with a note on what data governments should aim to collect. The overall message is clear: there is need for high quality micro data on road freight transportation that is more direct and detailed than those usually available.

¹⁰Existing evidence on the presence and implications of the Alchian-Allen effect comes from studies of international trade (Hummels and Skiba, 2004; Irarrazabal et al., 2015).

¹¹Also, the high level of informality and prevalence of owner-operators, as well as the general lack of list pricing in trucking, pose a challenge for using digital technologies such as web scraping to collect data.

Systematically collecting and disseminating such data should follow best practices. Balancing reporting costs against obvious benefits, surveys could be updated to incorporate critical information missing in existing questionnaires. Closer collaboration between transport ministries or departments and statistical agencies would leverage existing knowledge base, inform data needs and capabilities, and help avoid replication of efforts. As alluded to in Table 1, governmental agencies can increase the use and benefit of their existing products by providing transparent procedures to facilitate access to micro-data and sharing aggregated statistics in their websites—and making these available in English for global reach. Moreover, low-cost improvements to existing efforts could make them more informative for measuring overland transportation costs. For example, to aide estimation from price gaps, data that is regularly collected for compiling the consumer price index (CPI) could cover additional locations so as to generate sufficient geographic and spatial variation for this methodology to be more reliable. Another consideration described above is the comparability of products for which prices are collected. To the extent possible, targeting the same brands for any given product would be a good practice for not just precisely measuring transport costs but also the CPI. Other potentially helpful data products are consistent and frequent measures of road quality and traffic volumes. Once again, such data are often collected in order to monitor maintenance needs and congestion. There are well-established methods to define and quantify road quality such as the International Road Roughness Index. Appropriately designing the spatial features of these collection efforts would come at minimal additional cost but increase their usefulness for the purpose of measuring and lowering overland transportation costs.

References

- Alder, S., Song, Z. and Zhu, Z. (2021). Unequal returns to China’s intercity road network. Working Paper.
- Allen, T., and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3), 1085-1140.
- Allen, T., Atkin, D., Cantillo, S. and Hernandez, C. E. (2021) Trucks. Working Paper Slides.
- Anderson, J. E., and Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1), 170-192.
- Anderson, J. E., and Van Wincoop, E. (2004). Trade costs. *Journal of Economic literature*, 42(3), 691-75.

- Asturias, J., García-Santana, M., and Ramos, R. (2019). Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India. *Journal of the European Economic Association*, 17(6), 1881-1940.
- Atkin, D., and Donaldson, D. (2015). Who's getting globalized? The size and implications of intra-national trade costs (No. w21439). National Bureau of Economic Research.
- Baldomero-Quintana, L. (2020). How Infrastructure Shapes Comparative Advantage. Mimeo.
- Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145, 102442.
- Behrens, K., Brown, W. M., and Bougna, T. (2018). The world is not yet flat: Transport costs matter!. *Review of Economics and Statistics*, 100(4), 712-724.
- Berg, C. N., Deichmann, U., Liu, Y., and Selod, H. (2017). Transport policies and development. *Journal of Development Studies*, 53(4), 465-480
- Bergquist, L. F., and Dinerstein, M. (2020). Competition and entry in agricultural markets: Experimental evidence from Kenya. *American Economic Review*, 110(12), 3705-47.
- Bergquist, L., Faber, B., Fally, T., Hoelzlein, M., Miguel, E., and Rodriguez-Clare, A. (2019). Scaling agricultural policy interventions: Theory and evidence from Uganda. Unpublished manuscript, University of California at Berkeley.
- Bonfatti, R. and Poelhekke, S. (2017). From mine to coast: Transport infrastructure and the direction of trade in developing countries. *Journal of Development Economics*, 127, 91-108.
- Borraz, F., Cavallo, A., Rigobon, R., and Zipitria, L. (2016). Distance and political boundaries: Estimating border effects under inequality constraints. *International Journal of Finance & Economics*, 21(1), 3-35.
- Brancaccio, G., Kalouptsi, M., and Papageorgiou, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, 88(2), 657-691.
- Carballo, J., Graziano, A., Schaur, G., and Volpe Martincus, C. (2021). The Effects of Transit Systems on International Trade. Forthcoming at the Review of Economics and Statistics.
- Cavallo, A. and Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, 30(2), 151-78.

- Chatterjee, S. (2020). Market power and spatial competition in rural India. Working paper.
- Coşar, A. K., and Demir, B. (2016). Domestic road infrastructure and international trade: Evidence from Turkey. *Journal of Development Economics*, 118, 232-244.
- Coşar, A. K., Demir B., Ghose D. and Young N. (2021), Road Capacity, Domestic Trade and Regional Outcomes. Forthcoming at the *Journal of Economic Geography*.
- Coşar, A. K., Grieco, P. L., and Tintelnot, F. (2015). Bias in estimating border-and distance-related trade costs: Insights from an oligopoly model. *Economics Letters*, 126, 147-149.
- Combes, P. P., and Lafourcade, M. (2005). Transport costs: measures, determinants, and regional policy implications for France. *Journal of economic geography*, 5(3), 319-349.
- Coughlin, C. C. and Novy, D. (2021). Estimating border effects: The impact of spatial aggregation. *International Economic Review*, 62(4), 1453-1487.
- Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899-934
- Donaldson, D., and Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2), 799-858.
- Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and Trade: Evidence from the US. *Review of Economic Studies*, 81(2), 681-724.
- Duranton, G. (2015). Roads and trade in Colombia. *Economics of Transportation*, 4(1-2), 16-36.
- Engel, C. and Rogers, J. (1996). How Wide Is the Border? *American Economic Review*, 86(5), 1112-25.
- Fan, J., Lu, Y., and Luo, W. (2021). Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters. Forthcoming at *The Review of Economics and Statistics*.
- Felbermayr, G. J. and Tarasov, A. (2019). Trade and the spatial distribution of transport infrastructure. Forthcoming at the *Journal of Urban Economics*.
- Gwilliam, K. (2011). Africa's Transport Infrastructure : Mainstreaming Maintenance and Management. *Directions in Development ; infrastructure*. World Bank. <https://openknowledge.worldbank.org/handle/10986/2275>. License: CC BY 3.0 IGO.
- Head, K. and Mayer, T. (2010). Illusory border effects: distance mismeasurement inflates estimates of home bias in trade. In "The Gravity Model in International Trade:

Advances and Applications," Pages: 165-192. Cambridge University Press. Editors: Peter A. G. van Bergeijk and Steven Brakman.

Head, K. and Mayer, T. (2013). What separates us? Sources of resistance to globalization. *Canadian Journal of Economics/Revue Canadienne d'économie*, 46(4), 1196-1231.

Head, K. and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics* (Vol. 4, pp. 131-195). Elsevier.

Hernandez, C. E. (2021). Waits and Delays in Road Freight Transport. Mimeo.

Herrera Dappe, M., Kunaka, C., Lebrand, M., and Weisskopf, N. (2019). Moving forward: Connectivity and logistics to sustain Bangladesh's success. World Bank Publications.

Holguin-Veras, J. and Patil, G. R. (2008). A multicommodity integrated freight origin-destination synthesis model. *Networks and Spatial Economics*, 8(2), 309-326.

Hummels, D. and Skiba, A. (2004) Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture. *Journal of Political Economy* 112(6), 1384-1402.

Irrarrazabal, A., Moxnes, A., and Oromolla, L. D. (2015). The tip of the iceberg: a quantitative framework for estimating trade costs. *Review of Economics and Statistics*, 97(4), 777-792.

Kebede, H. A. (2021). The Gains from Market Integration the Welfare Effects of New Rural Roads in Ethiopia. Mimeo.

Lam Y., Sriram K., and Khera N. (2019). Strengthening Vietnam's Trucking Sector. World Bank Publications.

Martincus, C. V., and Blyde, J. (2013). Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment. *Journal of International Economics*, 90(1), 148-161.

Osborne, T., Pachón, M. C., and Araya, G. E. (2014). What drives the high price of road freight transport in Central America?. World Bank Policy Research Working Paper, (6844).

Persyn, D., Díaz-Lanchas, J., and Barbero, J. (2020). Estimating road transport costs between and within European Union regions. *Transport Policy*.

Porteous, O. (2019). High trade costs and their consequences: an estimated dynamic model of African agricultural storage and trade. *American Economic Journal: Applied Economics*, 11(4), 327-66.

- Redding, S. J., and Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of regional and urban economics*, 5, 1339-1398.
- Roberts, M., Melecky, M., Bougna, T., and Xu, Y. (2020). Transport corridors and their wider economic benefits: A quantitative review of the literature. *Journal of Regional Science*, 60(2), 207-248.
- Teravaninthorn, S., and Raballand, G. (2009). Transport prices and costs in Africa: a review of the main international corridors. World Bank Publications.
- Van Leemput, E. (2021). A passage to India: Quantifying internal and external barriers to trade. *Journal of International Economics*, 131.
- Wilson, W. W. (1987). Transport markets and firm behavior: the backhaul problem. *Journal of the Transportation Research Forum* (Vol. 28, No. 1, pp. 325-333).
- Wong, W. F. (2022). The round trip effect: The Round Trip Effect: Endogenous Transport Costs and International Trade. *American Economic Journal: Applied Economics*. Forthcoming.
- Zofío, J. L., Condeço-Melhorado, A. M., Maroto-Sánchez, A., and Gutiérrez, J. (2014). Generalized transport costs and index numbers: A geographical analysis of economic and infrastructure fundamentals. *Transportation Research Part A: Policy and Practice*, 67, 141-157.

Table 1: Summary of Methodologies

METHODOLOGY	DATA REQUIREMENTS	OUTCOMES/TYPE OF TRANSPORT COST INFERRED	CHALLENGES	EXAMPLES
Using data from an official survey	Availability from national statistical agencies	- Monetary costs in levels - May require computation of unit (ton-km) costs by dividing over road distances between reported origin and destination	Obtaining micro data from official agencies	- Behrens et al. (2018) - Allen et al (2021)
Collecting data through a special survey	Sampling from firm and carrier registries. Carrier surveyed on the cost of per ton-km shipment with a standard truck. Shipper surveyed on the price paid for cargo with its origin, destination, weight, value	- Monetary costs in levels	- Expensive data collection - Sampling from carrier or firm registries	- Teravaninthorn and Raballand (2009) - Herrera Dappe et al. (2019)
Imputation	Two types of data needed. Per ton-mile and time costs: vehicle operation costs, driver wages, tolls, fuel price. For routing decisions: GIS analysis on a digitized transport network with slope and congestion & speed of segments	- Monetary costs in levels - Enables simulation of per ton cost for any route	- Obtaining cost data to be used as inputs in computations - Complex GIS analysis	- Combes and Lafourcade (2005) - Zofio et al. (2014)
Estimation from trade flows	Domestic trade flows (value, weight) between regions. Could be exports or imports between ports and regions. Road distances from GIS based routing analysis or from an online map API	- Trade costs including transportation and other frictions - Relative costs by distance - Can be commodity specific if trade flows reported by commodity - Enables simulation of relative costs between routes	- Estimation is standard but applications require a structural gravity model	- Duranton (2015) - Coşar et al. (2021)
Estimation from price gaps	Prices of the same goods in various locations in the country. Road distances as above.	- Trade costs including transportation and other frictions - Relative costs by distance - Commodity specific by design	Obtaining micro data from official agencies	- Atkin and Donaldson (2015) - Chatterjee (2020)