Urban and infrastructure determinants of freight cost in India

Alejandro Molnar*
The World Bank

Forhad Shilpi
The World Bank

March 14, 2025

Abstract

This paper describes variation in the cost of road freight across India, using transaction data from a digital freight matching platform. The cost of moving freight is higher in states with less infrastructure and slower traffic, and for shipments to more remote areas. Controlling for market conditions at origin and destination, the cost of road freight is lower when freight moves over accessed controlled expressways, Golden Quadrilateral or NS/EW corridor roads and Public-Private Partnership (PPP) roads operated under an annuity mode. Freight cost is higher over tolled PPP roads, and when it intersects cities with low indices of traffic mobility and high congestion. Truck rates on long distance shipments are lower on routes where the placement of the railway network makes rail more competitive.

Keywords: Transportation. Infrastructure. JEL: D22, D62, L91, L92, R41, R42

*We are grateful to Sebastián Gómez Roca, Matías Harari, Mariana Santi and Francisco Tuñon for extraordinary research assistance, and to Muneeza Alam, Gilles Duranton, Matías Herrera-Dappe, Theodore Papageorgiou and Mansha Sehgal for helpful comments and discussions. Correspondence to aimolnar@worldbank.org. The use of maps in this document is preliminary, and has not been subject to cartographic review. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgement on the part of the World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

1 Introduction

Expanding highway infrastructure is a major development policy for the Government of India: the length of national highways increased by 60% in the decade that ended in 2023 (The Economic Times, 2024). In 2022 the government announced a National Logistics Policy aimed at reducing the cost of logistics in the economy.

This paper employs freight transaction data from a digital trucking platform in India to study the relationship between freight rates and a wide range of determinants of logistics cost, including road infrastructure. Identification in the paper is cross sectional, and leverages the variation in the paths that freight shipments take over India's road network. The movement of freight intersects urban areas and roads of varying characteristics, allowing us to estimate how the attributes and congestion of road infrastructure contribute to freight cost. Controlling for market conditions at origin and destination, we find that the cost of road freight is lower when freight moves over accessed controlled expressways, or highways that are part of the Golden Quadrilateral or NS/EW corridors.

Highways in India are largely delivered by means of Public-Private Partnership (PPP) projects, a policy advocated by government officials (Dash, 2023) and facilitated globally by the Public–Private Infrastructure Advisory Facility (PPIAF), hosted at the World Bank.[1] PPPs play a major role in financing infrastructure in low- and middle-income countries, but empirical studies on the effects of infrastructure delivered by PPPs is scarce (Fabre and Straub, 2023).

We find that conditional on road class, freight cost is lower on PPP roads where the private operator is remunerated through an annuity payment and is, on average, higher when freight moves over other classes of PPP roads (e.g., where the private operator recovers earnings through a toll). We also find that: i) freight that intersects cities with low indices of traffic mobility (or high congestion) pays an urban congestion premium and ii) freight rates are lower on trucking routes where the placement of India's rail network makes rail a more cost-competitive alternative.

A common approach to studying local roads or transport corridors is to evaluate their impact on local economic outcomes (e.g., Faber, 2014; Asturias *et al.*, 2019; Alam *et al.*, 2022; Asher and Novosad, 2020). Allen and Arkolakis (2022) use a general equilibrium model to measure the contribution of highway infrastructure to the economy as a whole. Studies of logistics cost are scarce, and are hindered by lack of data on user cost in trucking, particularly in the context of developing countries. Fan *et al.* (2021) use a revealed-preference approach to estimate the user cost of expressways in China. The closest research in this area of study is Lall *et al.* (2022), who employ the same data on freight transactions. Their focus, however, is on the relationship between freight cost and internal trade flows, whereas this paper studies a range of spatial and institutional determinants of freight cost. Our attribution of cost to road infrastructure also relies on a novel spatial database of highway projects in India, which we developed for this paper and

[1] Toll roads in India are the fourth largest project category (by count or total real investment) that is tracked by the World Bank's Private Participation in Infrastructure (PPI) Database, trailing only tolled roads in China and power generation in India and Brazil.

discuss in Section 2.2 below.

Investment in a particular segment of road infrastructure (whether a new road, new lanes, other upgrades or maintenance) may have effects on two margins. On the intensive margin, the investment in infrastructure may determine the cost of using the segment to freight users (say, through capacity and congestion, roughness and maintenance, or changes in toll fees). On the extensive margin, a road investment may cause freight flows to reallocate over the modified transportation network, resulting in shorter paths across the road network, diversions to lower cost paths and spillover effects. The paper does not directly address this second extensive margin, and modifying network connectivity may be a major potential effect of road supply on freight cost. That said, the paper does provide: i) an empirical measurement of road use cost as a function of road attributes, which can be used to ground network analysis when applied to counterfactual transportation networks, and ii) a back-of-the envelope calculation of how connectivity differences across states impact logistics cost.

2 Background and data

2.1 Freight transactions

Our main source of data consists of the transaction records of a digital freight matching platform in India. The primary business of the platform is to make a spot market and settle transactions between freight service providers (referred to as carriers) and freight service users (shippers), but it also manages a fleet of trucks that supply service through the platform. Our sample goes from March 1, 2017 to March 9, 2020; we select the end date of our sample one week prior to any noticeable impact on mobility on Google mobility trend data. The data contains 479,187 transactions over this period, accounting for 8.7 million metric tons of cargo.

Each transaction record includes the date of transaction, total rate in rupees (Rs.) charged to the shipper, cargo tonnage and type (described with 13 categories), three major truck types (container, open bed or trailer trucks) and 18 additional vehicle categories (mutually exclusive indicators for vehicle length and axle number) and an indicator for transactions provided by the platform's own fleet (4.6% of transactions). The data includes origin and destination locations for the cargo with sufficient precision to coordinate pickup and drop-off. Locations in the data are coded as Google Place IDs: the coordinates of these places usually refer to warehouses and other industrial sites both within cities and at specific rural locations, but in smaller towns and villages may be a central location within the village. The transactions involve 70,300 unique origin-destination pairs, with 6,889 unique origins and 8,932 unique destinations across all Indian states and union territories (except for the island territories). The data does not contain actual truck trajectories or performed shipment times. Summary statistics on the freight transaction data are provided in Table 1. Figure 6 uses binned scatterplots to describe the conditional mean relationship between rates, distance and tonnage.

Our sample of freight transactions can be compared to a sample of 2014-15 rates on 28 city pairs routes reported by Transport Corporation of India (TCI) and IIM Calcutta (2015). Caveats to the TCI data include that it is sourced from only four trips per city pair and that it predates India's GST reform, which started in 2017. The average rate across the 28 city pairs where both samples can be compared was Rs. 3.42 per ton-km on the digital freight platform and Rs. 2.75 in the TCI data, both adjusted through 2019 by the wholesale price index (diesel fuel inflation outpaced the general index by 6.7 p.p. over the period). Rate dispersion was identical at 0.72 and 0.70, respectively, and the correlation between both data sources is 0.63. A comparison of route distances and freight rates is displayed in Figure 2.

2.2 Road inventory data

To study the relationship between road freight and the attributes of individual road segments (such as their mode of governance) requires spatial data on the road network at a *routable* level of spatial resolution. Routable here means that the spatial layer can be made to overlap (under a common-sense definition) with a realistic description of a truck's trajectory or path over real-life road networks.

We construct a spatial and routable inventory of major road projects in India that focuses on three major categories of roads: i) National Highways Development Project (NHDP) roads under the Golden Quadrilateral and NS/EW corridor programs, ii) public-private partnership (PPP) roads and iii) access controlled expressways. To our knowledge there is no such prior, publicly available spatial database for India, either for the Golden Quadrilateral roads[2] or the other road categories.

We create a spatial database of PPP roads in India by merging the World Bank's Private Participation in Infrastructure (PPI) database to a database of PPP projects provided by Department of Economic Affairs in the Ministry of Finance of the Government of India. While these databases provide some attributes of PPP contracts, their location descriptions are narrative and brief. To merge the two data sources and digitize precise road locations we research the routing of each project using project definitions from current and archived versions of the National Highway Authority of India website, as well as annual reports for national and state road authorities and corporations, MOUs with private partners and other primary sources. For recent projects we verify the complete construction and operation by January 1, 2018 using historical imagery from the Google Earth desktop application. When road projects are only partially inaugurated by this date we subdivide the projects and digitize and date their sub-components. Our merge of the two data sources results in 667 road infrastructure PPPs constructed through January 1, 2018, of which 343 projects (45,777 km) were present in both databases, 245 projects (24,058 km) were present in the GoI database only, 64 projects (6,422 km) were present in the WB PPI database only and 15 projects (1,712 km) were present in neither and were identified through supplementary research of government documentation. We also digitize

[2]Ghani *et al.* (2016) provide a spatial layer for the Golden Quadrilateral and NS/EW corridors, but it is not routable or precise to the level of census districts.

the location of NHDP roads and accessed controlled expressways, which we identify though state and national sources. Further details on digitization are described in Appendix C. We also build a comprehensive dataset of the locations and fees for toll plazas, digitizing the location of 1,328 plazas. The construction of the toll plaza data is described in Appendix A.2.

Figure 1a displays the road layer for NHDP program roads, as well as all other National Highways identified in OSM data. Figure 1b displays the PPP segments in our database, using arbitrary colors to highlight the location and extent of each distinct project. Each of these segments is attached to a road identifier that we can match to information from the GoI PPP database, the World Bank PPI database or our own supplemental data.

2.3 Urban footprints and congestion

We rely on the definition of urban areas from Akbar *et al.*, 2023 (hereafter, ACDS), who define a footprint or boundary for 180 cities in India. For each urban footprint ACDS calculate an index of car traffic mobility, which they break down into congestion and uncongested mobility.

3 Empirical strategy and estimation results

The main empirical strategy in this paper is to describe the cross-sectional variation in freight rates in terms of attributes of the space intersected by freight routes. A first challenge is that the shipment-level data does not contain information on the actual route taken by the truck, only the address and coordinates of pick-up and delivery. We use the Valhalla open-source routing engine to build a network representation of India's roads using Open Street Map data as it was on January 1st, 2020. For each trip we construct routed paths from each coordinate-level origin to destination.[3] We intersect each shipment's route with spatial layers for administrative units (states and Census 2011 districts), road classes and urban footprints. The result of the intersection is that every meter on each route is "tagged" as belonging to one or more of these categories. Routed paths may employ any car-accessible road in OSM, and are not restricted to move along these spatial layers outlined above. Route segments that do not correspond to any of the tagged categories are measured and accounted for in a baseline road category. These roads are mostly composed of state highways and major district roads that are not operated under a PPP, and other rural or urban roads that are not contained within a defined urban footprint.

State-level variation in freight costs. We are interested in describing the variation in freight rates across India, and examining whether this variation is related to physical, infrastructure or other attributes of space. As a benchmark, we use a simple regression to decompose the market

[3]As a robustness check we also retrieve distances from a Google web service. The correlation between network algorithm routed distances and distances obtained from the web service in 2022 is 0.998. Google distances are 2.1% longer on average, despite a larger stock of highways in 2022 than are available to the routing algorithm, possibly reflecting a trade-off of time over distance. The absolute percentage difference between a Google distance and a routed distance is less than 8.6% for 90% of the sample (11.3% at the 95th, 19.1% at the 99th percentile of the sample).

price of a freight transaction into two spatial factors, which we estimate at the state level: i) the distance that each trip traverses through each state, and ii) the net change in a truck's position. Specifically, we regress

$$rate_i = \sum_s \beta_s \text{state km}_{si} + \text{controls} - \eta_o + \eta_d + \delta_{my} + \varepsilon_i \tag{1}$$

where each $\beta_s$state km$_{si}$ term attributes cost to the trip distance within each state times an estimated average cost per km per state. We also estimate a single η f.e. per state, subtracting it for trip origins and adding it to destinations so that $-\eta_o + \eta_d$ is the net effect from the repositioning of a truck from state o to d. We control only for the incidental attributes of each shipment that are unrelated to a place-based cost: a quadratic for tonnage, and fixed effects for 13 categories for cargo, three main truck types (container, open bed or trailer trucks), 18 additional vehicle categories (mutually exclusive indicators for vehicle length and axle number), month & year effects as well as an indicator for whether the transaction was provided by a platform-owned vehicle. We include position effects for all states and union territories, but only estimate a state-specific distance cost for states that account for at least 1% of all freight km in the sample: this excludes most union territories (except Delhi and Jammu & Kashmir), Goa, and the northeastern states of Arunachal Pradesh, Manipur, Mizoram, Nagaland, Sikkim and Tripura. Figure 3 maps the results of the decomposition in Eq. 1. Figure 3a shows that there is a stark spatial pattern in the market premium required to drive to particular locations in the country. Shipments to Eastern and Northeastern states require premia above Rs. 9,000, and five states require premia above Rs. 27,000.[4] Shipments to the largest state economies, which are in the west, allow for rate *discounts* with the premium to reposition a truck rising concentrically with distance from this center of economic activity.

State premia are identified from shipment direction within a state pair, and shipments that remain within state do not contribute to identification. Our decomposition allows for shipments in opposite directions of the same pair of states to have markedly different freight rates: e.g., a one-way shipment that travels east from Maharashtra (an estimated discount of Rs. 11,300) to West Bengal (Rs. 11,000 premium) is estimated to cost Rs. 44,600 more than an equivalent shipment in the opposite direction. This estimate compares to the raw data: there are 5,140 shipments within this state pair going in either direction; rates per ton-km average Rs. 3.27 eastward and Rs. 1.83 westward. Given the average distance (1,858 km) and tonnage (19.1 tons) across both directions, the Rs. 1.44 price gap per ton-km accounts for a directional price gap of Rs. 51,115 on the average shipment, slightly above the gap estimated from Eq. 1.[5] The

[4]The average rate in the sample is Rs. 43,125, but shipments to Northeastern states are, on average, longer than the average shipment. Results include controls for shipment distance (and allow for a state-varying distance coefficient), tonnage and other freight attributes.

[5]There is also a "trade imbalance" between these states, with 3.7 times as much tonnage moving east as west. This imbalance may be understood to rationalize the directional price gap under theories of market thickness and information frictions (Brancaccio *et al.*, 2022). Although data from the digital freight platform is representative of prices on the spot market for one-way shipments, we refrain from interpreting "trade quantities" because we the quantities transacted on the platform may not be representative of the market as a whole.

premia to relocate a truck in Figure 3a can be large; in the Maharashtra to West Bengal example, eastbound rates were 1.79 times westbound rates, indicating that the direction of freight is a substantial factor in the domestic freight market in India. In Figure 3b we control for the effect of repositioning a truck to display the estimated average per-km cost of moving freight through each state. Average per-km state costs are highest in the mountainous territories and states of Jammu & Kashmir, Himachal Pradesh and Meghalaya, and in the National Capital Territory of Delhi, all of which are above 75 Rs./km. Outside of these the highest rates are again in the Eastern states as well as in Madhya Pradesh, Gujarat, Punjab and Maharashtra (rates above Rs. 40 per km). The lowest rates are in Haryana, Rajasthan, Tamil Nadu, Andhra Pradesh and Telangana (below Rs. 30). A key lesson from this decomposition is that per-km rates can vary by more than 33%, even after excluding mountainous areas and the capital territory. The factors behind this variation are subtle, and are examined in detail throughout the rest of the paper. In the remainder of this section we continue to use the lens of state-level average costs to describe the incidence of three major classes of factors: diesel prices, traffic speed and the attributes of road infrastructure. It should be noted, however, that our decomposition of total rates into a positional effect and a distance through each state estimates an "intensive margin" cost, i.e. a cost over the network of roads that is in place in each state.

The total cost of moving through a state also depends on the availability of roads, allowing trucks to take more or less direct routes. As an example, consider freight over two city pairs: i) Ludhiana, Punjab to Jaipur, Rajasthan and ii) Ludhiana to Srinagar, Jammu & Kashmir (JK). Both city pairs require a drive of approximately 530 km,[6] with shipments northbound to Srinagar 1.9 times more expensive per ton-km than shipments south to Jaipur. However, Ludhiana and Jaipur are 446 km apart as the crow flies (the ratio of road to crow flies distance is 1.18, below the median in our sample across India), while Ludhiana and Srinagar are only 389 km apart (1.36 ratio) so the average per-km rate in this transaction understates the cost of moving north through JK. The distinction over distances in a "crow flies" vs. a road metric is essential to interpreting the role of infrastructure covariates on freight cost: an addition to infrastructure that decreases the road distance of freight may lower cost in a manner that is not captured by the average per-km state costs displayed in Figure 3b.

We quantify the contribution of road supply to routed distance across each state with the following procedure: let c_i and r_i be the crow-flies and routed distances on a given freight transaction, and r_{is} the distance sub-segments of r_i over each state s. We regress the "circuity" (i.e., surplus distance expressed as a share) $(c_i - r_i)/r_i$ for each transaction on a vector of state share weights r_{is}/r_i. The coefficient on each state share weight is consistent for the average circuity contributed by each state in the freight sample. For example, the estimate for Jammu & Kashmir is 0.41, which implies that across our sample each crow-flies km through JK averages $1/(1 - 0.41) = 1.69$ routed kms.

We then extrapolate state cost per km using the estimated state circuity averages, displaying in

[6] Prior to the inauguration of the Banihal Qazigund Road Tunnel in April 2022.

Figure 3c the cost per routed km (in orange) next to a cost per crow-flies km (grey). As expected, the incremental cost due to path routing is highest in mountainous states and territories (JK, Meghalaya, Himachal Pradesh, Assam and Kerala), and is lowest in Delhi, where road density is highest. The median circuity in sample is .26, and excluding the previous examples is above 0.3 in Chhattisgarh, Telangana and West Bengal and below 0.2 in Uttar Pradesh and Punjab.

We now examine state-level variation in the incidence of three major classes of freight cost determinants: diesel prices, traffic speed and the attributes of road infrastructure. In all of these cases we examine the sensitivity of a baseline average state cost per km to the inclusion of a determinant of cost by estimating an equation of the form

$$rate_i = \sum_s \beta_s \text{state km}_{si} + \gamma_\mathbf{i} \mathrm{G}_\mathbf{i}^\mathbf{f} + \text{controls} - \eta_o + \eta_d + \delta_{my} + \varepsilon_i \qquad (2)$$

where $\mathrm{G}_\mathbf{i}^\mathbf{f}$ are covariates intended to control for cost factor f. Crucially, the $\mathrm{G}_\mathbf{i}^\mathbf{f}$ are not required to be nested within each state and may overlap with different states. In Figure 4 we plot $\hat{\beta}_s^{(1)} - \hat{\beta}_s^{(2)}$ for each state, i.e., the difference between the average state cost (Eq. 1) and the same after absorbing a cost factor (Eq. 2). Details on the construction of each cost factor are provided in the shipment-level regression analysis below. Here we summarize these determinants as: (a) a set of interactions measuring the share of each trip route that is spent at different intervals of **traffic speed**, (b) the average **price of diesel** along the route, and (c) a set of interactions measuring the distance that is routed along different classes of **road infrastructure**. Each of these cost determinants varies at the trip level, and with the exception of diesel prices the sources of the underlying variation do not strictly overlap with state boundaries.

Figure 4a shows by how much average state cost decreases when we separately control for traffic speed. Cost in mountainous states is explained by traffic speed, likely due to road grade in addition to any effect due to traffic congestion. A second spatial pattern displayed in this map is that traffic speed explains freight cost in Eastern states, which have slower highway traffic and less road infrastructure. Figure 4b shows that the variation in state cost that can be explained by differences in diesel prices is more muted: although the price of diesel is a major cost factor, it does not vary substantially across states.[7] Pass-through of state-level diesel prices to freight cost may also be moderated if drivers can time some fueling across state borders. Figure 4c maps changes in freight cost after controlling for a set of interactions that measure the distance within each trip that is routed along different classes of roads: Golden Quadrilateral and NS/EW roads, access-controlled expressways, national highways and public-private-partnership roads under tolled, annuity or other modes of governance. Unlike the previous cost factors, this exercise

[7]The procedure to estimate diesel prices in each state and month is outlined in Appendix A.1. Our procedure to assess variation in diesel prices differs from Eq. 2 above because the price of diesel can be expected to enter freight cost linearly with distance, at a rate given by a vehicle's fuel efficiency. We model the effect of diesel cost as distance times a distance-weighted average state diesel price. We normalize a lower bound for the effect of diesel price on freight cost by adding back to each state cost the value of the fuel price interaction when diesel prices are set at the 5^{th} percentile in the sample. The estimates in Fig. 4b do not reflect the total attribution of freight cost to the cost of diesel, the magnitude of which would be larger than pictured, but the extent to which this attribution varies across states relative to a benchmark where fuel prices are set at their 5^{th} percentile.

reveals net reductions and increases in state cost. For a given state, finding a reduction in cost here means that freight cargo through the state is routed over road classes that are found on average to have lower cost per km. Anticipating results in the next section, these include access controlled expressways and national highways improved under the Golden Quadrilateral and NS/EW programs. Conversely, the use of infrastructure results in a higher average freight cost in states where cargo is routed over road classes where user cost is high. These include PPP roads under the Build-Operate-Transfer (tolled) mode and other, non-annuity PPP modes. The maps in figure 5 (a)-(e) display the fraction of freight km through each state that are carried over roads in each of five road classes, while figure (f) shows the average frequency of encounters with toll plazas. A caveat to the preceding results is that they reflect variation in user cost over existing roads, and do not quantify the benefits of infrastructure from the extension of the road network and potential improvements in connectivity. Since they are derived from market price charged to freight users, the benefit estimates in Fig. 4c are inclusive of both any speed or productivity improvements and the cost of toll fees that may be passed on to users. To disentangle potential productivity benefits from toll costs, in Fig. 4d we reproduce the analysis from the previous figure but where the outcome (market freight rates) has been netted of predicted toll fees for each transaction. Our transaction data does not include toll fees, and platform prices are settled with toll liabilities due to the driver. Appendix A.2 outlines our procedure to estimate the toll fee bill for each transaction. Net of toll costs, most states see reductions in freight cost where infrastructure is available. Taken together, the estimates in Figs. 4c and 4d reveal that a large share of the freight productivity improvements due to infrastructure are offset by charges to users in the form of toll fees, and that the incidence of infrastructure on freight cost varies across states with the prevailing form of infrastructure that is available and that is used by the freight sector (see Fig. 5).

Shipment-specific determinants of freight cost. We now focus on the determinants of freight rates at the transaction level, rather than state-level descriptive statistics. To set a benchmark, we first describe freight rates in terms of shipment attributes, excluding the spatial components that are of main interest. We estimate

$$rate_i = \beta_d \text{total km}_i + \text{controls} + \delta_o + \delta_d + \delta_{my} + \varepsilon_i \tag{3}$$

where the total rate in rupees (Rs.) of each freight transaction is regressed on the total route distance (in km) as well as a set of controls that do not depend primarily on a shipment's path over the road network. Table 2, Col. 1 reports the simple regression of rates on route distance and tonnage, and serves as a contrast with Col. 2 which adds F.E. by the Census 2011 district of origin and of destination (δ_o and δ_d) as well as by month and year (δ_{my}). Time effects are added for precision, and the specification is cross sectional: the sample spans 36 months and the primary spatial attributes of interest do not vary substantially over the period. Opening dates and construction periods for infrastructure are not dated with sufficient precision to leverage panel variation within this time frame.

The specification in Column 1 of Table 2 reports a marginal cost to the freight user from an additional km of transportation on the digital platform market, while controlling for tonnage. As a comparison with the next specification shows, however, the estimate of this marginal cost depends on the origin and destination markets for freight. In Col. 2 we use fixed effects at the level of Census 2011 districts at each endpoint of the trip. Unlike cities, districts are defined unambiguously throughout India and the larger area of a district (there are 596 in the data) constitutes a sensible definition of a market area for long-haul freight. Estimating these effects separately for each origin or destination allows for flexibility in absorbing any market discount or premium for moving to or away from each district. Controlling for origin and destination we find that the marginal cost of an additional km at Rs. 35.8, or Rs. 1.96 per ton-km at average tonnage. The remaining specifications (as well as specifications with spatial attribution that will be reported on below) consistently find a baseline marginal cost in the range of Rs. 33 to 35 per additional km of transportation, estimated with precision.[8] Col. 3 adds cargo and vehicle effects for 13 categories for cargo, three main truck types (container, open bed or trailer trucks), and 18 additional vehicle categories (mutually exclusive indicators for vehicle length and number of axles), as well as an indicator for whether the transaction was provided by a platform-owned vehicle. Col. 4. adds a quadratic term for tonnage, with negligible impact. Col. 5 adds controls that, while spatial in nature, are not primarily determined by a shipment's trajectory over space: indicators for shipments that remain within state or within district, the number of states visited and indicators for whether the trip starts or ends within an urban footprint defined by Akbar *et al.* (2023) for a sample of 180 cities. The urban endpoint indicators are common across cities and, in this specification, are intended to capture an additional average cost from the specifically urban component of freight beyond the effect that is absorbed by district-level fixed effects. Across these and subsequent specifications we consistently find a rate discount for within-state transactions and a rate premium for the number of states visited (controlling for distance) and for shipments that originate in urban settings. The cost estimates for within-district and urban destination shipment attributes are not precisely estimated or robust, particularly to very flexible functional forms for the relationship between distance and tonnage (discussed below).[9]

The baseline specification for the spatial attribution analysis to be performed below is reported in Col. 5. We use a baseline linear specification of rates on distance to allow for the coefficients on sub-segment distances that will be reported on below to be interpreted as per-km premia. Cols. 6-8 present additional robustness checks: Col. 6 includes a quadratic of

[8]Standard errors are clustered at the level of the origin-destination district pair (28,141 clusters). The sample of 479,160 shipments consists of 70,279 unique origin-destination pairs at the address level, and these would make a natural unit for clustering. We more conservatively group the variation of shipments in the same origin-destination district pair, as these may share a substantial component of the routes (the average shipment moves 890.7 km and crosses 3.1 states).

[9]In unreported regressions, this lack of robustness is related to the fixed cost component of freight rates: flexible functional forms for distance (i.e., kinks and additional polynomial flexibility below a threshold distance) estimate a discount for within-district trips and a premium for urban destinations, although not statistically significant.

distance, and Col. 7 interacts distance with tiers of tonnage, allowing for differential cost per km as suggested by Figure 6a.[10] Col. 8 measures distance with routing paths queried from a Google web service in 2022: the distance coefficient is 4.8% lower than in the baseline, about half of which can be attributed to the fact that routes are 2.1% longer according to Google. The remaining attenuation may result from measurement error over the roads that are available in a 2022 Google response but were not available during the 2017-2020 sample. Subject to the same caveat, Col. 9 includes travel time obtained from Google as a regressor. Since trip distance and duration are jointly determined (and highly correlated, $\rho = 0.9899$), the interpretation for the coefficient on distance changes in this specification, and the coefficient can be interpreted as the contribution to cost of moving through space (e.g., due to fuel) after partialling out the time cost of that movement. Coefficients are significant, sensible and consistent with prior results despite the high correlation between the variables: the cost due to time and distance of a trip at the sample means of 890.7 km and 16.4 (web) hours is Rs. 29,829.5 ($= 17.6 \times 890.7 + 863.0 \times 16.4$), or Rs. 33.5 per km, consistent with estimates from alternative specifications. This specification is also suggestive of an estimate that an hour of driver and truck time costs Rs. 863. However, there are at least two hurdles to making such an interpretation. First, technological and speed limit differences between motorcycles, cars and trucks mean that Google travel time predictions are likely measured with some error for our purposes. A related issue is that travel time predictions over long distances also do not account for the stops required in long-haul trucking. Second, just as the coefficient on distance in Col. 9 cannot be interpreted as the cost of moving freight by 1 km because it partials out the time cost of such a move, the coefficient on time has to be interpreted as conditioning on the return to trucking from moving cargo across space. In other words, this estimate cannot be used in isolation as a measure of freight cost to the user or the freight carrier.

Cost of time. We develop an alternative estimation strategy that allows us to learn the marginal cost of time from our freight transaction data while addressing the previous concerns. For each route in the data we download travel *directions*, which contain an average of 62 "maneuver"-level segments and travel times per routing query. Maneuver-level data adds up to the usual Google travel time predictions, but has a spatial resolution that allows us to attach it to specific route sub-segments and to create a histogram of the distance that each trip spends at each speed.

We estimate a specification identical to that in Table 2, Col. 5 but where we estimate a coefficient on distance separately for each of eight traffic speed intervals. This allows us to estimate how movement at each of these speeds adds to freight cost. We report the speed interval coefficients in Table 3. Point estimates show that freight rates are highly responsive to Google's predictions of traffic speed: the freight cost per km declines monotonically as speed increases, and levels out after a speed of 50 km/h. If we assume that truck speed reaches free flow at the speed of 50 km/h at which cost flattens out, we can calculate a cost of ₹28.8 per km at free flow. For speed intervals below 50 km/h we subtract 28.8 from the point estimate of cost to

[10] We provide results on alternative outcome specifications (log rates, rates per ton and per ton-km) in Online Appendix B, which are consistent with Table 2.

obtain a surplus or incremental time cost per km. The speed at the midpoint of each interval implies an incremental time beyond free-flow speed, and from our estimate of surplus cost and the incremental time we obtain an implied cost of time, which is reported in the last row of Table 3. Estimates in the range from 10 to 40 km/h are consistent for a time cost of Rs. 2,000-2,500 per hour.[11] These estimates of the cost of time represent the cost of time incurred in excess of that required to move cargo under traffic free flow conditions, as opposed to average time. However, these results indicate that excess travel time is valued at almost the same rate as average time: the cost range given above means that the cost of excess travel time is in the range of 76.5% to 95.6% of the average per hour freight rate.

Road-class determinants of freight cost. Freight routes take place over different classes of road, and the nature of the road may account for freight cost on the route. Figure 8 plots the density of freight rates (per ton-km) according to the share of the route that is covered by PPP roads. Routes with a PPP share in the top three quintiles (3, 4 and 5) have the lowest rates, and the distribution is substantially shifted toward higher rates for roads with the lowest coverage by PPP roads. While suggestive, the figure does not provide evidence on the role of PPP roads on freight cost because PPP roads may be located along corridors that connect markets or that go through cities that incur high freight rates.

We extend the previous estimation strategy to decompose the contribution of such factors on rates. We use specifications of freight cost that are linear in both rates and distance (rather than log-log specifications, for example). Linear specifications give equal weight to every kilometer within a route, while allowing for the assignment of more than one spatial attribute to a given km of road (i.e., when spatial attribution layers to overlap), or for some attribute to take on a value of zero (e.g., when a trip does not cross any urban boundary, or use any PPP road.). Specifically, we estimate equations of the form

$$rate_i = \beta_d \text{total km}_i + \textstyle\sum_s \beta_s \text{segment km}_{si} + \text{controls} + \delta_o + \delta_d + \delta_{my} + \varepsilon_i \tag{4}$$

where segment km_{si} is the length of the route on transaction i that is contained within road segment s, of which there may be a large set. Segment s could represent an individual road or sub-project on a road, but in the following specifications will represent one of several potentially overlapping road classes. For example, a route from Delhi to Ahmedabad may include segments on one of several PPPs roads, on the Golden Quadrilateral, on both, or neither. The lengths attributed to each of these potentially disjoint spatial segments is summed up to a length per road class. A particular interval on the road may "turn on" the distance count for more than one road class, and on any observation the value of a given segment length may be zero (if the

[11] Estimates for the 40-50 range are somewhat higher, but are highly sensitive to the assumption that free-flow speed is at 50 km/h. The implied time cost for this range would be in line with the remaining estimates at 2,500 Rs./h if we had instead assumed free-flow at 51.7 km/h. Estimates in the 0 to 10 km/h range are an order of magnitude higher, which may indicate the additional cost of uncertainty for what are predicted to be the most congested road segments in the sample. Estimates from this range should be interpreted cautiously as they are estimated with the least precision and account for less than 1/10,000 of the route km in the sample.

route does not intersect a segment of the class) or some continuous value up to the total segment length. All specifications include the total length of the route, so the estimated coefficients on road class lengths identify empirical premia (or discounts) per km, relative to a baseline.[12]

Table 4, Col. 1. adds to the baseline specification of shipment-level covariates a count for the route km that take place on any PPP road. Consistent with Figure 8, km on a PPP road are Rs. 2.2 cheaper on average, even after absorbing the effect of origin and destination districts.

We next include segment travel lengths for a broad set of spatial layers: access-controlled expressways, urban footprints, the GQ and NS/EW highway programs and all other km on the National Highway system. Travel across these classes of roads may incur varying costs, such as: i) a direct monetary cost, as some roads are tolled, ii) indirect monetary costs, due to road conditions and maintenance or the effect of congestion on fuel use, iii) the cost of driver and vehicle time, and perhaps, iv) compensating differentials to the driver due to risk, comfort or roadside amenities. The premia or discounts that we estimate in Table 4 include all of the above factors (we will address netting toll fees from the above premia below). In Col. 2 we find that travel along expressways and the GQ/NSEW has a substantial rate discount (Rs. 6.5 and 5.8 respectively, or 19.2% and 17.4% relative to the baseline cost of distance). Statistically significant estimates of approximately these magnitudes are robust across the specifications displayed in Table 2 above. The freight rate premium on km traversed across cities is positive, but it is not estimated precisely or consistently across specifications.[13] All together, when attributing freight cost to a broader set of road classes we find a premium of Rs. 1.7 per km on a PPP road. This premium for PPP governance on a road might be attributed to maintenance and tolls on these roads, but needs to be interpreted jointly with the fact that some expressways and a fraction of the GQ/NSEW highways were delivered through PPP contracts: the net estimated effects on such roads is still a discount (e.g., Rs. -4.1 for a PPP segment of the GQ). Col. 3 further decomposes the length of freight routes attributed to PPPs into three main modes: Build-Operate-Transfer (BOT) roads financed by tolls, Build-Operate-Transfer (BOT) roads financed by a government annuity and other, less common modes. We find substantial heterogeneity across modes: freight premia for travel on BOT tolled roads are Rs. 7.53 higher per km than on BOT annuity roads,[14]while the highest road premia are on routes that make use of other modes.[15] Travel along national highways (NH) is not statistically significant from the baseline,

[12]The estimating equation is the simplest case of the method developed in Mangrum and Molnar (2020) to use data on sums or integrals (such as total travel time, or total trip cost) over arbitrary endpoints in a line to identify their integrand pace or marginal cost function.

[13]Section 2.3 below focuses on urban travel, finding substantial and city-specific heterogeneity in urban travel cost and its determinants. Additionally, the coefficient on the length of urban travel is the only one of the per-km regressors that is estimated jointly with two intercepts, as the standard set of spatial controls include indicators for starting and ending within a city.

[14]BOT annuity roads are sometimes, but not always, toll-free. On some roads the government separately concessions a toll operator, generating revenue that funds the annuity. See Fabre and Straub (2023) as well as Rajan *et al.* (2014) and Patel and Bhattacharya (2010) for background on the supply of infrastructure PPPs, and for roads in India in particular.

[15]These include Special Purpose Vehicles (SPV), which are frequently used in port connectivity projects or other situations where infrastructure has a single main beneficiary, PPPs granted or renewed under an Operation and

which consists of travel along state highways, major district roads and any other urban or rural road not classified. Road governance estimates are not substantially changed after including average travel time (Col. 4) or a flexible specification of baseline distance (Col. 5, where distance is interacted with nine mutually-exclusive vehicle type and tonnage-tier categories). Estimates for an average urban premium are large but imprecisely estimated and not significant in any of these these specifications; we address the heterogeneity behind the average urban km below.

To separately identify the monetary costs from the potential benefits of road categories we need data on toll fees. We do not observe sub-components of the freight rate, such as spending on toll fee payments. Transacted rates are gross of toll fees, and fees are levied on the truck driver. We address this issue by predicting the toll fees for each trip, given its route through India and the date at which it took place. To this end, we construct a novel dataset of toll plazas throughout India and predict the rate at each plaza at each point in time. In the case of access-controlled expressways we do this for each journey within the expressway. Details of this procedure are described in Appendix A.2. We then intersect routes through plazas (or entry and exit plaza pairs, for access-controlled expressways) to predict a total toll fee bill for each trip.

In Table 5 we re-estimate the specifications from Table 4 but where we have netted the toll fee bill from the total freight rate. We do not separately include toll fees as additional regressors because fees are highly correlated to the set of segment km measured under all of our road categories. Comparing Col. 2 across both tables we see that the Rs. 1.7 premium on PPP roads becomes a Rs. 2.7 discount on revenue net of tolls. This suggests that while the attributes of these roads provide net benefits to freight, the benefits are more than offset by toll fees which are passed on to users. In an unreported version of the baseline regression in which we include the total predicted toll fee for each trip we find a pass through rate of 0.30 per tolled rupee. A rate that is less than one-for-one is consistent with some forms of market power, with non-monetary benefits to truckers that result in a shared incidence of the toll, or with attenuation due to measurement error, as we predict toll fees rather than obtain them from transaction data. Columns 3-4 shows that the results repeat for BOT tolled roads, rates over which are Rs. 2.6 lower net of fees, and for BOT Annuity and other PPP modes which result in cost discounts that are Rs. 5.0 and 4.9 greater, respectively, net of toll fees. The comparison across both BOT modes suggests that travel across BOT annuity roads provides a substantial cost discount to users. Whereas BOT annuity roads are in practice often tolled by the NHAI, the user cost of tolling does not entirely offset the cost benefit. Even conditioning on origin and destination markets and a large set of spatial controls, the benefit from the BOT annuity mode and its difference with the BOT tolled mode may owe to the frequent deployment of this mode outside of major inter-state trunks, where alternatives may be of relatively lower quality.

We describe heterogeneity in the impact of tolled roads on freight users by means of Table 6. Specifications in this table include alternative counts for toll plazas crossings, a single linear

Maintenance (OM) contract and a more recent mode (Hybrid Annuity Model, HAM) for which there have been few project completions through March 2020.

distance (i.e., no distances by road class), and the spatial and cargo controls employed earlier. In Panel A we allow the reduced form coefficient for crossing a toll plaza to vary by quintile of the plaza's toll fee, and we include access-controlled expressway tolls in a separate row. Column (RF) displays reduced-form coefficient estimates for each quintile of the toll fees: conditional on distance among other controls, crossing a toll plaza is associated with lower net freight rates for plazas in the 1st, 4th and 5th quintile of toll rates, as well as for A.C. expressways. Since crossing a toll plaza mechanically incurs a toll fee on the truck, in the next column we display the average fee cost for tolls in each quintile. We subtract coefficient estimates from this cost and display both a point estimate and a 95% confidence interval for a "net benefit" of each toll that is inclusive of payments to the toll operator (note that this is not a social benefit, as it is not net of the costs of building or maintaining the road). Tolls in the top 40% of fees (as well as A.C. expressways) have the highest benefits, explained both by the higher fees paid and a negative reduced-form impact on freight rates. Roads in the bottom 60% of fees have comparable benefits of around Rs. 200 per toll, but the toll fee paid by users exceeds this benefit for all but the cheapest 20% of tolls.

Toll fees vary for a number of factors, including the nature of the road improvement project, the PPP mode and bidding in the awarding stage. A first order factor, however, is project length. In Panel B we interact terciles of toll fee rates with quintiles of project length. After controlling for the length of the projects that give rise to toll fees, we find that cheap tolls are consistently associated with lower freight rates, as benefits to the user seem to exceed the toll fee. Results for intermediate fee tolls are mixed, while expensive tolls are associated with freight cost reductions only for the shortest projects - the most common project of this type are urban ring roads.

Urban congestion. Our previous estimate of the freight rate premium for the average urban km was imprecise and inconsistent across specifications. We find that this is accounted for by heterogeneity across and within cities: Table 7 reports on additional regressions based on Eq. 4. We include road class layers (PPP modes combined) and the usual spatial and shipment-level controls, but we decompose the urban intersections into quintiles under alternative measures of traffic mobility or population density. Col. 1. reports on a decomposition by quintiles of population density, which is calculated at the level of up to five segments per city.[16] Here, a segment class s accounts for the sum of lengths over a set of equivalent layers (e.g., the ring-level footprints for all city ring-segments that are in the top quintile of population density nationally). Estimated population density premia are roughly monotone, with a rate discount for the two quintiles of lowest population density, and a substantial premium for the top three quintiles by density. Cols. 2-5 use mobility data estimated by ACDS, and the attribution of a city to a quintile varies at the city- (as opposed to city-ring) level. The effect of urban premia is monotone in the overall mobility index (Col. 2), with large and statistically significant results for the top and bottom quintile of mobility: the per-km rate across a city in the lowest quintile of mobility

[16] Akbar *et al.* (2023) provide population and areas for up to five concentric rings per city (up to 2 km from city center, 2-5 km, 5-10 km, 10-15 km, 15-100 km). Areas are not constant because each city's urban footprint is contained within one or more of these rings.

is 55% higher than in the baseline rural km, or 2.77 times higher than the average urban km (= 19.203/6.928, employing the cross-city estimate from Table 4, Col. 2). ACDS decompose mobility into indices for congestion and uncongested mobility. Rates vary strongly and with statistical significance across quintiles of the congestion index, although estimates are imprecise and not strictly monotone. Results for uncongested mobility and average speed are again close to monotone, but only large and statistically significant for the lowest or lowest two quintiles (for uncongested mobility and speed, respectively).

Table 8 reports on a version of Table 7, but where the outcome of interest are the web estimates of travel time and the unit of observation is a routed origin-destination pair rather than a freight transaction. The regressors of interest are route attributes and ACDS congestion measures, but the truck paths used in this analysis may differ from those in Table 7 because they follow the recommended route and navigation instructions provided by the the web platform. The table does not employ freight platform data, except to define the sample of route paths that are of interest.

Results in Col. 1 of the table indicate that, weighted over our sample of realistic freight paths, the web-predicted travel time for the baseline km across India in October 2022 was 0.017 hours per km (i.e., 58.8 kph). A baseline km in this specification takes place outside of cities or the major road categories of interest. This speed is higher than average truck speed in India, and reflects the web-platform's model of average conditions for an average vehicle (which may be a car or a two- or three-wheeler). Our interest lies in how the platform's estimate of travel time varies with road attributes: we find that the platform predicts a travel time of 0.22 hours per km (45.5 kph) within cities, faster than baseline speeds (in the range of 62 to 71 kph) for PPP roads, AC expressways and GQ/NSEW roads, and slower than baseline (50 kph) speeds for National Highways not included in the above categories.

Columns 2 through 6 again break out travel within urban areas into quintiles of population density and the ACDS congestion indices. Results are roughly consistent with Table 7 and with common sense, with ACDS measures of higher mobility and lower congestion indices associated with faster travel. The results are not always monotone, however, indicating some discrepancy between our sample of freight paths and the sample paths queried by Akbar *et al.* (2023). We also find that movement through city rings that have the lowest population density have substantially slower traffic (in our sample of freight paths), which may reflect infrastructure availability though low density urban areas.

Geography. Table 9 reports on some further geographic determinants of freight cost. We measure elevation gain or loss along each route by tracing the path of each route over a digital elevation model (we use the SRTM 90m Digital Elevation Database v4.1). We query elevation changes every 100m along each path and construct a histogram of the distance that each truck covers at every grade. We find that at this level of spatial resolution the routes that have large elevation gains (ascents) also have large descents, regardless of whether a route involves a substantial net ascent or descent. Because we cannot estimate the effect of net elevation gain (it is mostly absorbed by fixed effects at the district of origin and destination) we aggregate grades by absolute value to estimate the effect of driving at grade conditional on total elevation change. Col 1. reports coefficient estimates on cost per km covered at a range of grades. Travel at grades of up to 1% are estimated to have costs comparable to the baseline specification. The point estimate for travel in the range of 1 to 2% grades is three times as high, but is not estimated precisely. Travel in grades ranging from 2 to 10% is estimated to cost more than 10 times the average. The estimate of cost at this grade is not precise, but it is statistically significant.

We next evaluate factors that could potentially affect freight cost but that are not segments of route distance: these are district $\times$ time or district-pair variables. For these factors we estimate the analogue of the baseline specification but including the log of the freight rate, distance, tonnage and the variables of interest. In Col. 2 we include rain shocks (the log difference between precipitation in a given district and month and its previous 10-year average) at origin and destination. We find a small but statistically significant premium for driving into a rain shock: a one-standard deviation increase in rain at destination causes a 0.4% increase in the freight rate. The point estimate for rain shocks at the origin is a third the size and not significant.

Language. In Col. 3 of Table 9 we include the log of the probability that a speaker from the origin and the destination district share any spoken language, as in Melitz and Toubal (2014). We find a significant and substantial effect on rates: a 10% increase in the probability of a shared language decreases the freight rate by 3.1%. This effect may be driven by information frictions, as a common language may help a driver find a return load at destination, or by cultural preferences. Although our finding is consistent with more than one cultural channel, it is reasonable to rule out factors related to proximity (such as a driver home preference, as in Yang, 2022) given the extensive set of spatial controls.

Competition within the truck freight sector. Market structure in the trucking industry is fragmented, with both large, national operators and many small and medium firms with varying degrees of formalization. Firms can supply services across state lines, and the digital platform from which we obtain transaction data is one tool used by carriers to supply services outside their home market.

Defining a market and measuring market structure in this context is challenging. Market size, information frictions or trade flow imbalances may result in varying degrees of firm entry and competition. We do not have a measure of competition that varies over time and space at the resolution of our rate transaction data. We use data from a list of transport operators, which is compiled by the Indian Bank Association (IBA) to qualify firms that meet certain financial and accounting standards, to observe the degree to which some route markets are served by a set of larger and more financially mature firms. The IBA maintains a "Recommended Transport Operators" list to facilitate the practice of its member banks to provide discount lending against trucking receipts. Inclusion in the list increases the probability that a trucking firm will be able to access discount lending. To be included in the list the transport firm needs to own at least seven heavy commercial vehicles, meet and maintain the IBA's accounting and solvency standards, and regularly submit documentation to the IBA that includes financial statements. The list includes data on the state pairs served by each of 673 qualified firms as of June 2023.

We find that entry by IBA qualified firms is related to lower freight rates, with rates 1.0% lower per firm operating in a route. Figure 9 displays coefficient estimates for 12 interactions of market entry, showing non-parametrically that negative relationship between IBA firm entry and freight rates holds in a roughly monotone pattern. IBA "monopoly" routes have 10% higher rates than routes with four providers, and the pattern continues with additional IBA entry. An important caveat to this result is that more IBA firms choose to enter state pairs that can be expected to carry more trade (for example, between the Golden Quadrilateral endpoints in Delhi, Maharashtra, Tamil Nadu and West Bengal). Although our results control for infrastructure, for market conditions at each market endpoint and for tonnage as a first-order measure of scale effects, they do not rule out broader economies of scale in trucking operations (e.g., thicker bilateral markets with a lower probability of an empty return trip) which might contribute to lower rates through operational efficiency rather than additional competition.

Competition from rail. Rail transported 27% of freight cargo in 2018-2019, compared to 64% of cargo transported by road (Indian Railways, 2022). Rail and road are complements in the supply of a multimodal logistics network, but on any given trade lane are in competition for market share. A city pair where rail is cost-effective may have lower or more price-sensitive demand for truck freight, with effects on production cost and market power in the trucking sector.

We estimate the effects of rail competition on truck rates by leveraging variation in the strength of rail's competitiveness. First, we use data from the Indian Railways (IR) Freight Operations Information System website to determine which stations are qualified to load or unload containers or general cargo. For each trip endpoint in our freight transaction data we calculate straight-line distances to the nearest loading and unloading stations for each of these two types of cargo. Second, we query the IR Rates Branch System (RBS, at rbs.indianrail.gov.in/ShortPath) for the distance between these stations along the rail network. Lastly, we calculate the ratio of rail to road distance for each freight transaction (i.e., rail circuity), and augment our baseline regression with measures of both circuity and the local distance to access a rail station at origin and destination. We do not employ data on actual rail freight rates, which depend on a classification of more than 600 types of cargo and are subject to regulation and multiple government schemes. However, all rates contain a distance portion that is determined by IR from the same RBS data that we employ.

We find that truck freight is cheaper on routes where the alternative provided by rail is relatively more direct. Specifically, we add quintiles of rail circuity to our baseline specification. In an unreported regression we find that shipments in the second to fifth quintile of circuity are Rs. 56, 648, 806 and 870 more expensive than the the most direct routes, the last two estimates are significant at the 5% level. The economic magnitude of this average effect is small: Rs. 870 is 2.6% of the average rate in the most circuitous quintile. Local distances to access rail stations are not significant in this specification. The finding of a small average effect masks substantial heterogeneity by distance. Most rail freight includes an intermodal overhead cost, which can only be amortized over a long distance.[17] Consistent with this fact, for shipments under 300 km the modal share of freight for trucks is 2.8 times that of rail, whereas this ratio is 2.1 for shipments beyond 300 km (Indian Railways, 2022). We therefore interact our measure of rail circuity with route distance, separately binning our sample into terciles of distance and quintiles of rail circuity. Figure 10 displays the coefficient estimates from these interactions. Rail's circuity is unrelated to freight rates for the shortest two thirds of freight trips (up to 1,082 km in our sample). We find a stark effect of rail circuity on freight rates for trips in the top third of distance, with rates increasing monotonically as rail becomes a weaker source of competition across two different measures of competitiveness. The rate difference between shipments with the strongest to the weakest competition from rail in terms of rail network circuity is Rs. 9,761, which is 12.3% of the average freight rate for the longest third of the shipment sample.[18] A second dimension of rail competitiveness is the proximity of cargo endpoints to rail stations. We measure rail access as the sum of distance at both endpoints, and we bin this access distance in the intervals of less than 10 km, 10 to 25 km or more than 25 km (corresponding to 38, 35, and 27% of the sample). We again find a monotone relationship with access for trips in the top third of distance. As in previous regressions, district origin and destination fixed effects absorb transportation market conditions that hold at either trip endpoint rather than the od-pair, and a linear distance term controls for the average effect of distance, including any residual variation in distance within each of the interaction cells. In unreported regressions we also find that: i) the effect of rail proximity is stronger for general cargo, and weaker for containers (only 4.9% of rail cargo by metric tons was composed of containers in 2019, vs. 14.7% in our sample of truck freight), ii) rate monotonicity also holds for container trips of intermediate distance, iii) additionally conditioning on terciles of tonnage, the monotone effect of rail circuity is strongest for the heaviest cargo in the top distance tercile, and iv) the monotone effects of rail circuity on rates hold exactly for cargo in the categories of Chemical and powder, Construction materials, Cylinders, Electronic goods, Machine and auto parts, Package and consumer boxes, and Scrap, hold approximately for Food and agriculture, and do not hold for Alcoholic beverages, Batteries, or Petroleum and paint products.

We view our results on the relationship between rail competitiveness and truck freight as consistent with two, possibly concurrent, market mechanisms. On the demand side, note that supply in the trucking sector is fragmented and that the barriers to entry for any specific route are low (particularly for supply though a digital platform, as in our transaction data). However, for any degree of market power we would expect a competitive rail alternative to reduce truck freight rates though price competition, as it increases the price elasticity of residual demand faced by truck drivers. On the supply side, we expect that competitive rail carries a greater share

[17] An exception to this are dedicated industrial stations (*industrial sidings*), such as between coal mines and power stations.

[18] A specification in logs is also monotone: we find a 0.100 log difference in rates between the quintiles with strongest and weakest rail competition.

of freight demand on a given city pair (as suggested by the relationship between distance and mode; Indian Railways, 2022). To the extent that driver labor supply is upward sloping within particular city-pairs (note that a shared language at origin and destination is also a determinant of cost) then rail could have a general equilibrium effect on truck driver wages even if the sector is perfectly competitive.

City-specific channels. Cities on a freight route may have idiosyncratic cost or amenity values to truck drivers, beyond the cost of urban congestion found in section 2.3 above. Consistent with that cost, Figure 11 reports on an alternative specification to Eq. 4 in which city-specific premia are estimated for each of the 180 cities in Akbar *et al.* (2023). Consistent with Table 7, cities with higher ACDS mobility have lower estimated urban premia. Additional city-specific costs might include daytime restrictions on truck movement (as imposed in Delhi) or other costs of trucking. Figure 11c plots urban premia against survey data on the average bribe collected from truckers by traffic or highway police (SaveLife Foundation, 2020). Although the survey only provides data for 10 cities, there is a suggestive correlation between urban premia and expected bribes.[19]

4 Conclusions

This paper has documented how freight rates are related to market conditions, road classes, urban mobility and other cost factors such as gradient, travel time, fuel and toll costs. We do not claim that the statistical relationships documented in this work are causal. For example, private sector entry into the provision of roads or freight services may be more likely in markets with high traffic or inelastic freight demand, potentially making road availability a function of the prevailing freight rates.

Our analysis does provide the first description of how freight user cost varies across space in India, statistically decomposing the contribution of road and route attributes from market conditions at origin and destination. We describe how the supply of road infrastructure, toll prices and congestion along freight routes are related to logistics costs – a major priority of economic development policy in India.

[19]Guwahati is an exception and outlier to this correlation. Guwahati is the largest city in Northeast India, and may provide amenity value to truck drivers as the only major city along the corridor from Northeast India to the rest of the country. SaveLife Foundation (2020) reports that police bribe requests in Guwahati double those over all other sampled cities, and that road blockades and extortion by local groups are frequent there during Durga Pooja festivities.

References

Akbar, P., Couture, V., Duranton, G. and Storeygard, A. (2023). Mobility and Congestion in Urban India. *American Economic Review*, **113** (4), 1083–1111.

Alam, M., Herrera Dappe, M., Melecky, M. and Goldblatt, R. (2022). Wider economic benefits of transport corridors: Evidence from international development organizations. *Journal of Development Economics*, **158**, 102900.

Allen, T. and Arkolakis, C. (2022). The Welfare Effects of Transportation Infrastructure Improvements. *The Review of Economic Studies*, **89** (6), 2911–2957.

Asher, S. and Novosad, P. (2020). Rural Roads and Local Economic Development. *American Economic Review*, **110** (3), 797–823.

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, publisher: Oxford Academic.

Brancaccio, G., Kalouptsidi, M., Papageorgiou, T. and Rosaia, N. (2022). Search Frictions and Efficiency in Decentralized Transport Markets.

Dash, D. K. (2023). Road ministry pitches for more highway project on PPP citing quality issue in government funded works. *The Times of India*.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System. *The Review of Economic Studies*, **81** (3), 1046–1070.

Fabre, A. and Straub, S. (2023). The Impact of Public–Private Partnerships (PPPs) in Infrastructure, Health, and Education. *Journal of Economic Literature*, **61** (2), 655–715.

Fan, J., Lu, Y. and Luo, W. (2021). Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters. *The Review of Economics and Statistics*, pp. 1–46.

Ghani, E., Goswami, A. G. and Kerr, W. R. (2016). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *The Economic Journal*, **126** (591), 317–357.

Indian Railways (2022). National Rail Plan for India - 2030.

Lall, S., Sinha-Roy, S. and Shilpi, F. (2022). *Trucking costs and the margins of internal trade: Evidence from a Trucking Portal in India*. Policy Research Working Papers, The World Bank.

Mangrum, D. and Molnar, A. (2020). *The Marginal Congestion of a Taxi in New York City*. Working Paper.

Melitz, J. and Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, **93** (2), 351–363.

Patel, U. R. and Bhattacharya, S. (2010). Infrastructure in India: The economics of transition from public to private provision. *Journal of Comparative Economics*, **38** (1), 52–70.

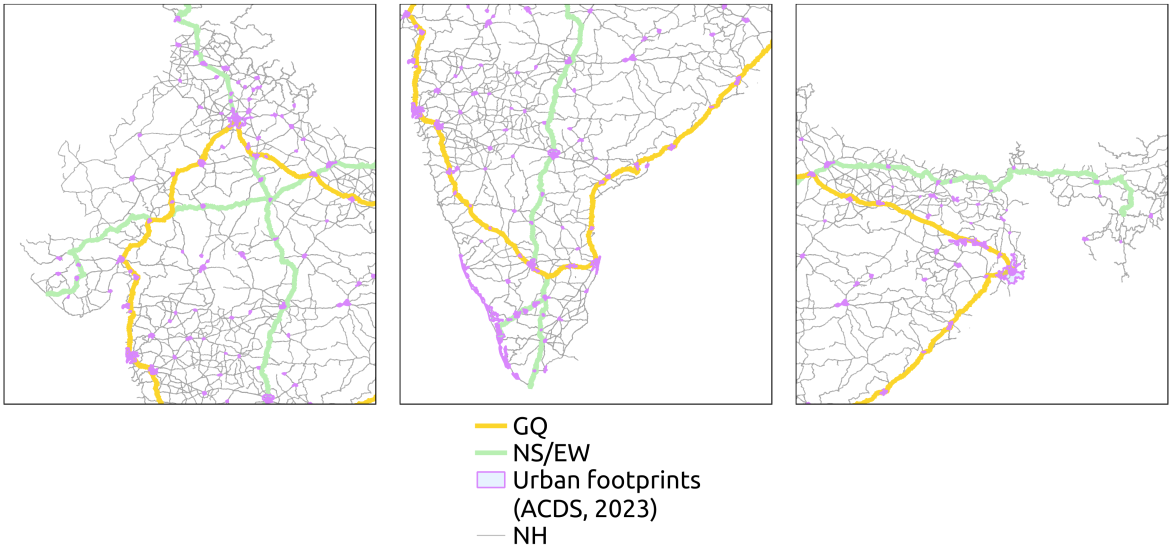
Rajan, T. A., Gopinath, G. and Behera, M. (2014). PPPs and Project Overruns: Evidence from Road Projects in India. *Journal of Construction Engineering and Management*, **140** (5).

The Economic Times (2024). Length of national highways in India increased 60 pc in last 10 years: Road Transport and Highways secretary Anurag Jain. *The Economic Times*.

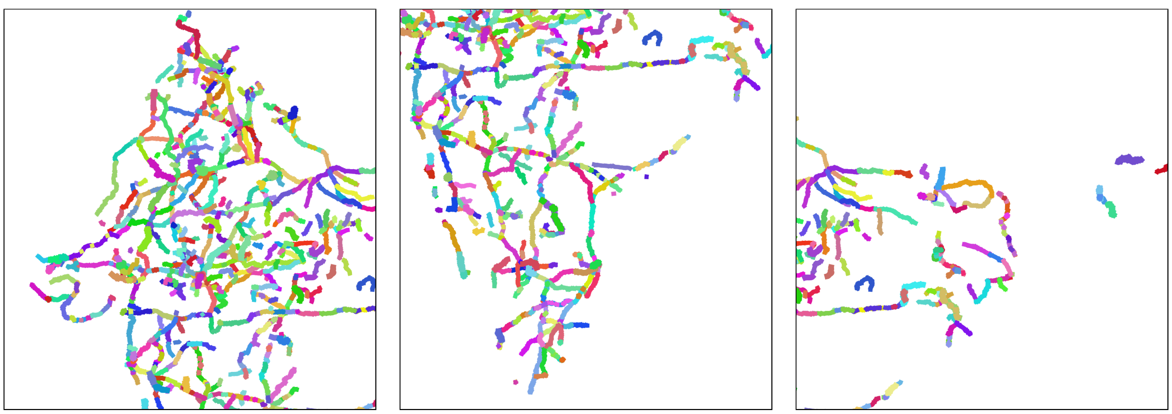
Transport Corporation of India (TCI) and IIM Calcutta (2015). *Operational efficiency of freight transportation by road in India: 2014-2015*. Tech. rep.

Yang, R. (2022). (Don't) Take Me Home: Home Preference and the Effect of Self-Driving Trucks on Interstate Trade.

Tables and figures



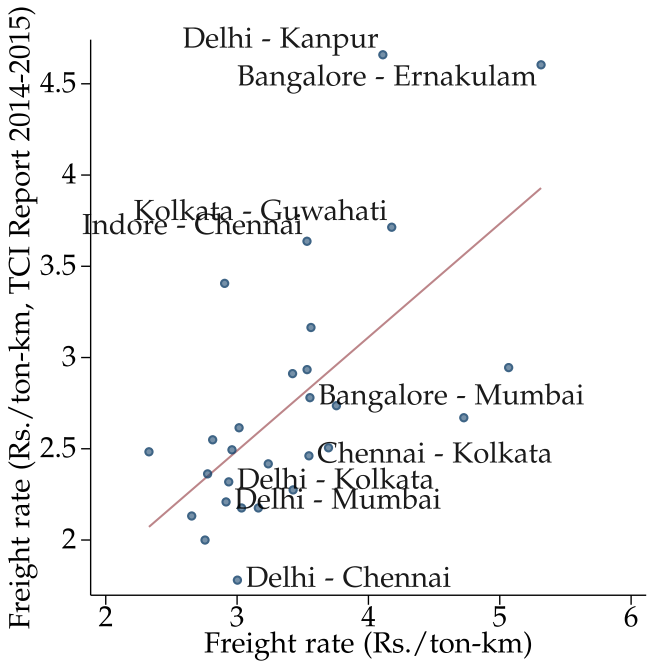
(a) Road classes and urban footprints



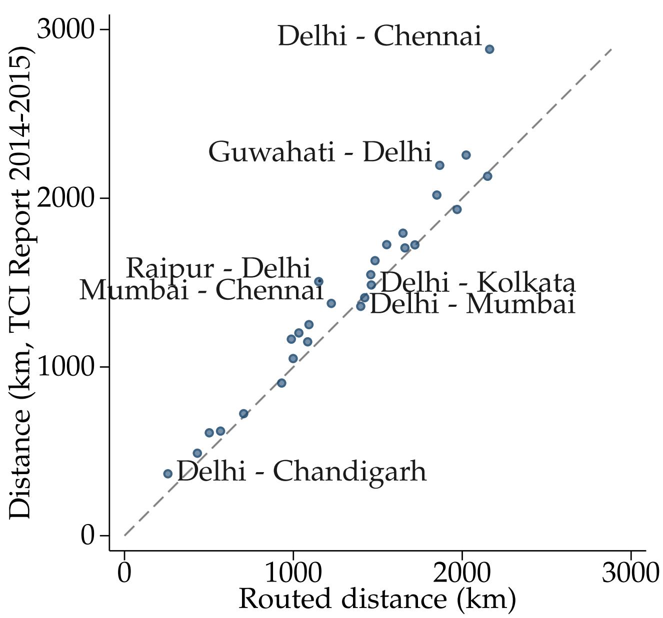
(b) Inventory of Public-Private-Partnership (PPP) roads

Figure 1: Spatial layers

PPP road inventory shows 667 projects divided into 912 project sub-components.



(a) Freight rate per ton-km



(b) Route distance

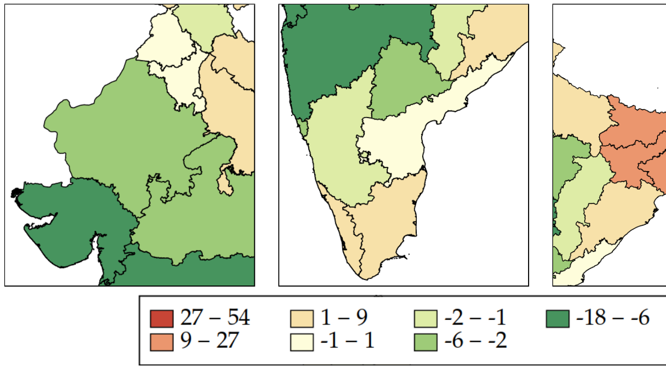
Figure 2: Cross-sample comparison to 28 market survey included in TCI Report, 2014-2015

Comparison of digital freight dataset to survey data from Transport Corporation of India (TCI) and IIM Calcutta (2015) for 30 trips in the Delhi-Bangalore and Delhi-Mumbai lanes (15 in each direction) and 4 trips (2 in each direction) in each of 26 other markets. Digital platform statistics include all transactions between a city-pair's urban footprints (average 478 observations per route). Platform rates are first averaged per direction and then averaged again across both directions of each city pair for comparability with the statistics reported in TCI (2015).

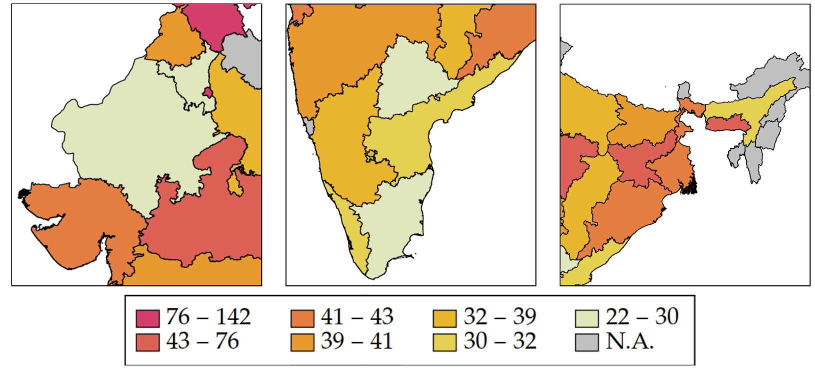
Table 1: Freight transaction data. Summary statistics

	Mean	S.D.	Min	Max
Freight transaction				
Rate (Rs.)	43125.3	37132.8	10000	850000
Rs. per ton	2643.3	2029.1	314.3	36000
Rs. per ton-km	3.78	4.50	0.26	737.9
Route km	890.7	684.3	1.13	3824.3
Route km (web)	909.4	705.8	1.50	3918.7
Tonnage	18.2	8.32	1	35
Within state	0.22	0.41	0	1
Within district	0.0090	0.095	0	1
No. states visited	3.11	1.82	1	12
Intra-urban	0.00075	0.027	0	1
Urban origin	0.52	0.50	0	1
Urban dest.	0.64	0.48	0	1
Petrol price along route (Rs./l)	76.6	4.33	62.9	90.2
Diesel price along route (Rs./l)	70.1	3.91	56.5	82.2
Hours per 100 km (web)	1.91	0.26	0.60	8.25
Origin district				
Precipitation (mm)	71.1	125.6	0	2111.3
Precipitation, 10-year avg. (mm)	74.8	117.4	0	1462.5
Destination district				
Precipitation (mm)	77.5	136.3	0	2111.3
Precipitation, 10-year avg. (mm)	85.3	135.5	0	1767.3
District pair				
IBA: No. of qualified operators	2.95	3.49	1	34
IBA: Multi-market contact (joint prob.)	0.25	0.24	0	1
Prob. same L1 language	34.7	36.5	0.0056	99.8
Observations	479187			

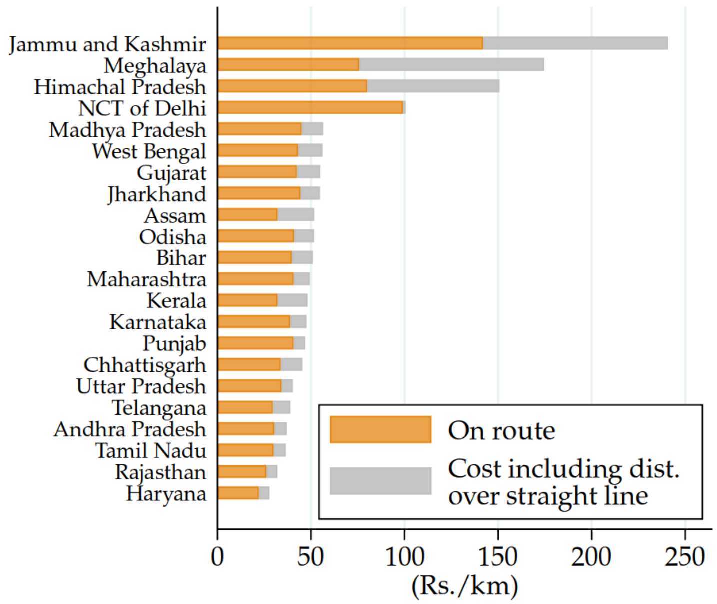
Each observation is a freight transaction, which are observed over 70,300 address-level origin destination pairs (average 6.88 transactions per od-pair). Network routing algorithm fails on 21 pairs (28 for web tool), resulting in 483,624 observations in the sample (483,596 for web measurement of route distances and travel time).



(a) State relocation F.E. (1,000 Rs.)



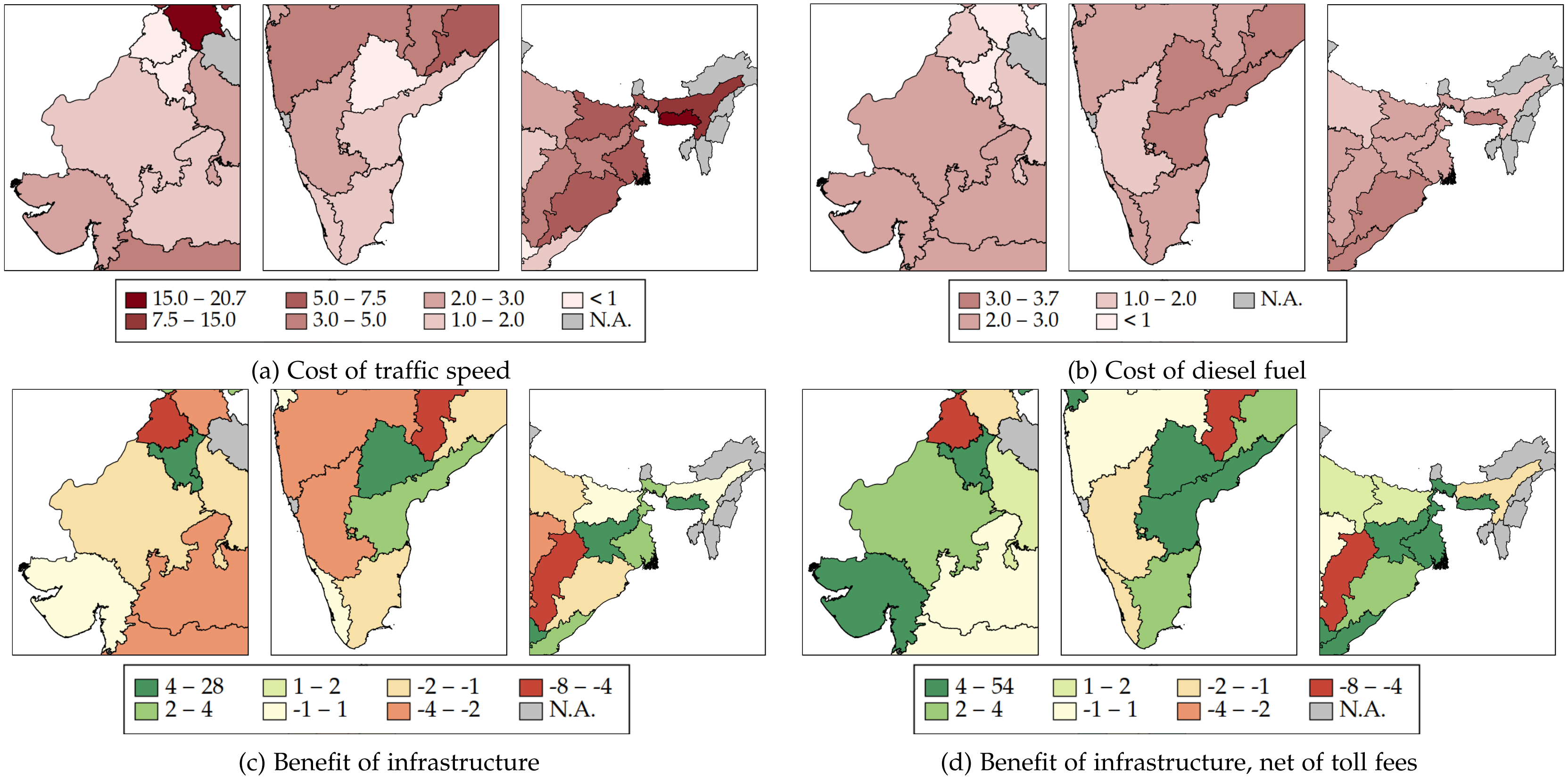
(b) Avg. state cost (Rs./km)



(c) Road availability

Figure 3: Attributes of the roads used in the freight sample. State averages

Figure (a) displays the net effect of relocation to each state on freight cost ($\hat{\eta}_s$ in equation 1), in thousands of rupees. Figure (b) displays the average state cost in rupees per km. Chart (c) displays an extrapolation of the state cost per km (from Figure b) to the incremental distance required to cross each state. Incremental distances are estimated as defined on page 7.



(a) Cost of traffic speed

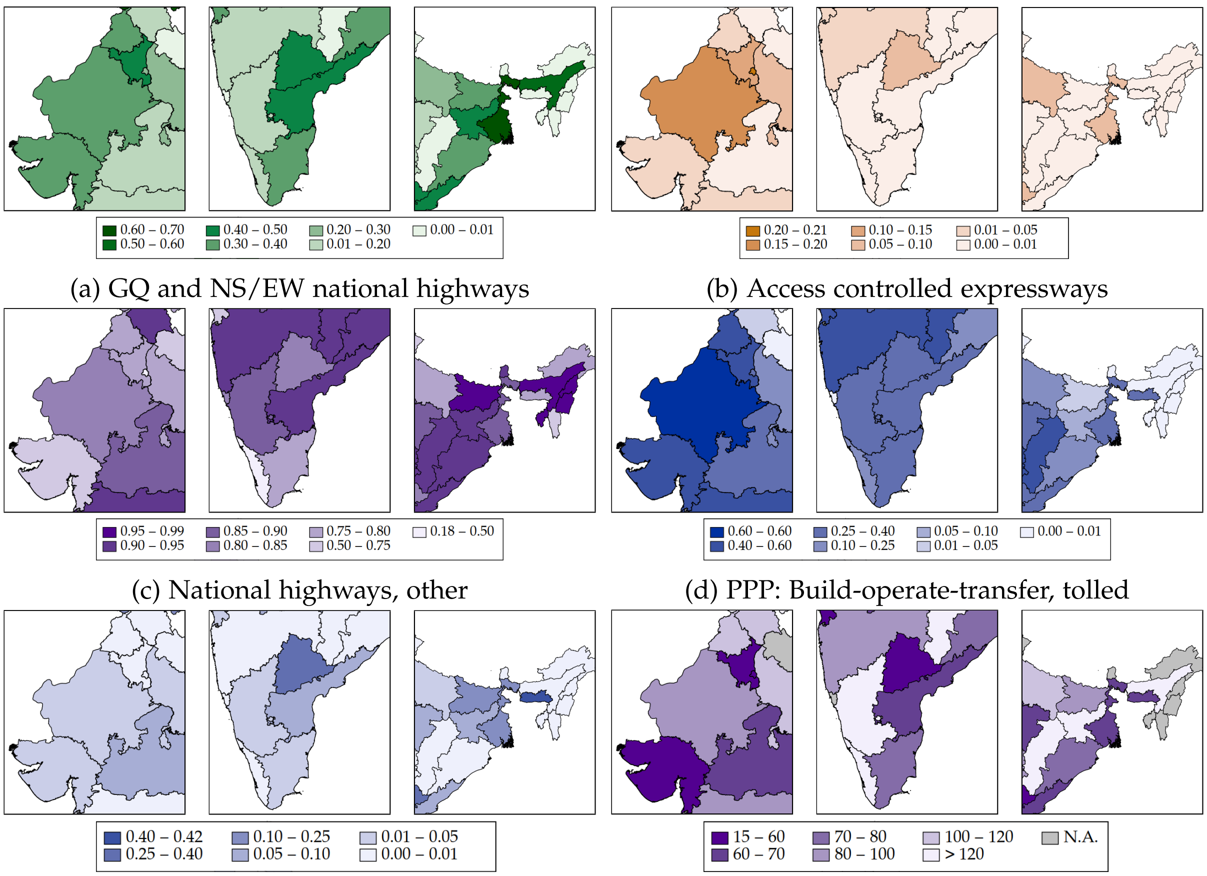
(b) Cost of diesel fuel

(c) Benefit of infrastructure

(d) Benefit of infrastructure, net of toll fees

Figure 4: State-level variation in sensitivity to cost and benefit factors

Figures (a)-(d) display the difference in Rs./km between the baseline estimated average state cost (Eq. 1) and the state average after controlling for each of three factors that may account for freight cost (Eq. 2). The rate outcome in Figure (d) is net of toll fees, predicted for each route as explained in Appendix A.2.



(a) GQ and NS/EW national highways

(b) Access controlled expressways

(c) National highways, other

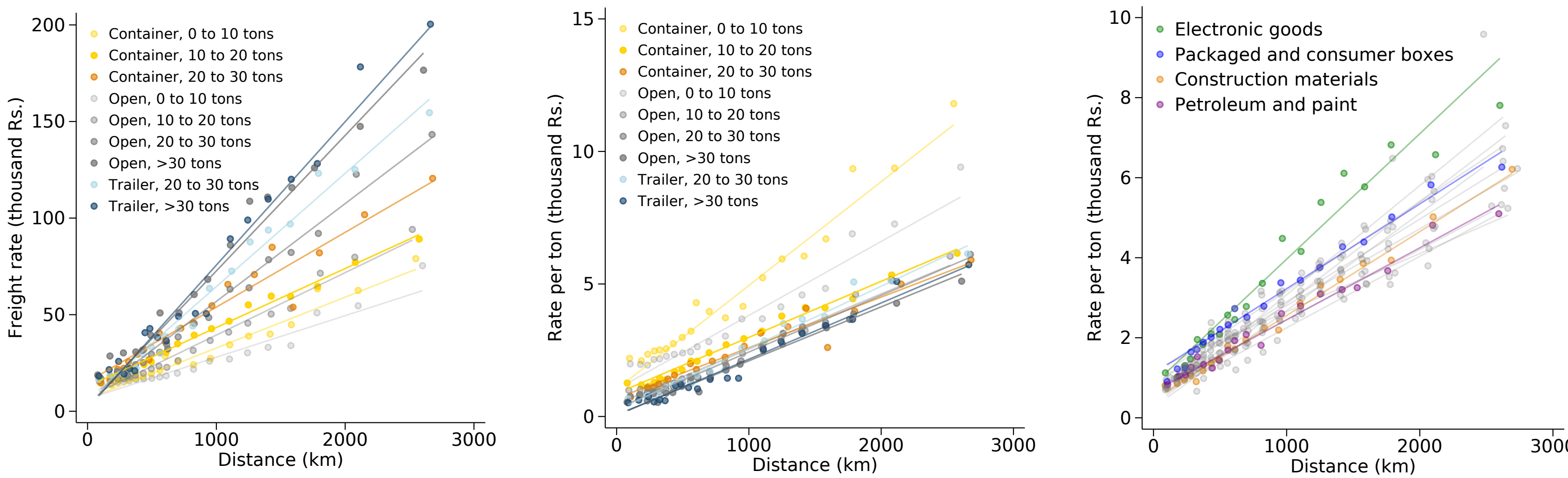
(d) PPP: Build-operate-transfer, tolled

(e) PPP: Build-operate-transfer, annuity

(f) Avg. km between toll plazas

Figure 5: Attributes of the roads used in the freight sample. State averages

Figures (a)-(e) display the fraction of all sample freight km within a state that route over each class of road. Road segments may qualify for several road classes, so shares may add up to more than one. Figure (f) displays the average number of km between open-loop toll plazas, as encountered over the sample of freight route movements within each state.



(a) Freight rates by vehicle type and tonnage tier

(b) Per-ton rates by vehicle type and tonnage tier

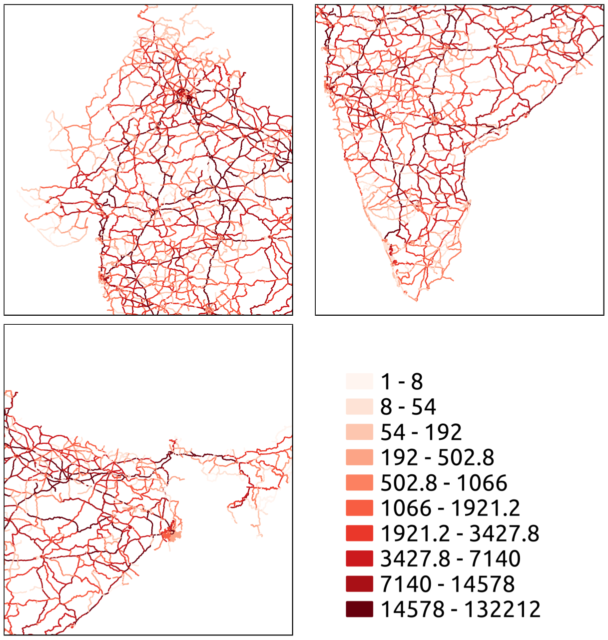
(c) Per-ton rates by cargo type, selected examples highlighted

Figure 6: Binned scatter-plots of freight rates by cargo, vehicle and tonnage categories

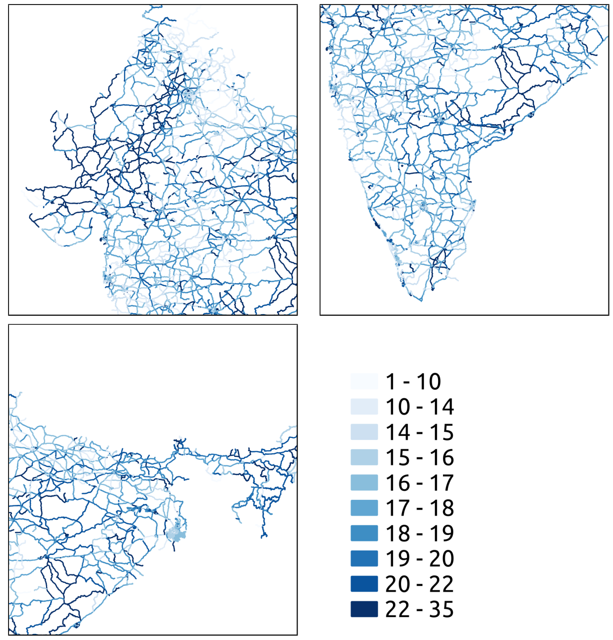
Table 2: Shipment-specific determinants of freight cost

	Dependent variable: freight rate (Rs.)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Route distance (km)	43.2^{a}	35.8^{a}	35.4^{a}	35.4^{a}	33.0^{a}	35.5^{a}		31.4^{a}	17.6^{a}
	(1.05)	(0.30)	(0.29)	(0.29)	(0.53)	(0.94)		(0.48)	(2.37)
Route distance sq.						-0.0^{a}			
						(0.00)			
Tonnage	1404.2^{a}	1490.0^{a}	1324.9^{a}	-607.2^{a}	-544.4^{b}	-524.0^{b}	485.5^{a}	-533.2^{b}	-575.1^{a}
	(62.70)	(29.53)	(80.33)	(217.54)	(213.75)	(213.09)	(151.25)	(213.16)	(214.65)
Tonnage sq.				35.6^{a}	34.3^{a}	33.9^{a}	-6.0^{c}	34.2^{a}	35.0^{a}
				(4.42)	(4.33)	(4.32)	(3.18)	(4.32)	(4.34)
Distance (km) X									
0 to 10 tons							19.1^{a}		
							(0.54)		
10 to 20 tons							27.5^{a}		
							(0.56)		
20 to 30 tons							39.1^{a}		
							(0.56)		
>30 tons							56.3^{a}		
							(0.85)		
Within state					-2380.9^{a}	-1748.2^{a}	-881.3^{b}	-2260.1^{a}	-2143.8^{a}
					(436.38)	(476.53)	(356.83)	(436.21)	(431.78)
Within district					-464.6	-77.7	3968.3^{b}	-808.5	-511.2
					(2387.60)	(2396.15)	(1616.41)	(2326.40)	(2320.93)
No. states visited					763.4^{a}	931.2^{a}	835.3^{a}	1107.5^{a}	842.8^{a}
					(207.42)	(211.43)	(171.34)	(196.95)	(202.80)
Urban origin					1807.0^{a}	1792.6^{a}	988.9^{a}	1786.2^{a}	1862.7^{a}
					(341.99)	(341.08)	(269.56)	(346.08)	(343.19)
Urban dest.					165.6	122.8	-415.9	133.9	260.2
					(411.02)	(410.25)	(335.15)	(410.07)	(407.15)
Platform owned fleet			-3571.4^{a}	-3567.2^{a}	-3556.7^{a}	-3540.0^{a}	-2935.2^{a}	-3609.3^{a}	-3787.3^{a}
			(469.00)	(466.87)	(463.45)	(462.38)	(417.21)	(469.90)	(468.25)
Travel time (web, hours)									863.0^{a}
									(131.40)
Cargo & vehicle type FE	N	N	Y	Y	Y	Y	Y	Y	Y
O, D & MY FE	N	Y	Y	Y	Y	Y	Y	Y	Y
Observations	479160	479160	479160	479160	479160	479160	479160	479132	479132
R^2	0.74	0.89	0.90	0.90	0.90	0.90	0.93	0.90	0.90
R^2-within	0.74	0.69	0.72	0.72	0.73	0.73	0.79	0.73	0.73

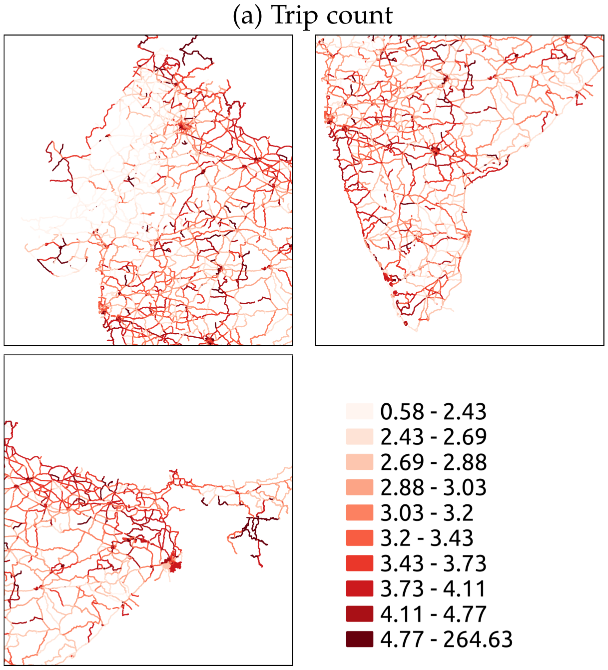
Note: An observation is a freight transaction. Standard errors are clustered at the level of the district origin and destination pair (28,141 clusters) and statistical significance is indicated by letters: *a* at 1%, *b* at 5%, *c* at 10%. Sample in Cols. 8 and 9 drops 27 observations that the Google API fails to route.



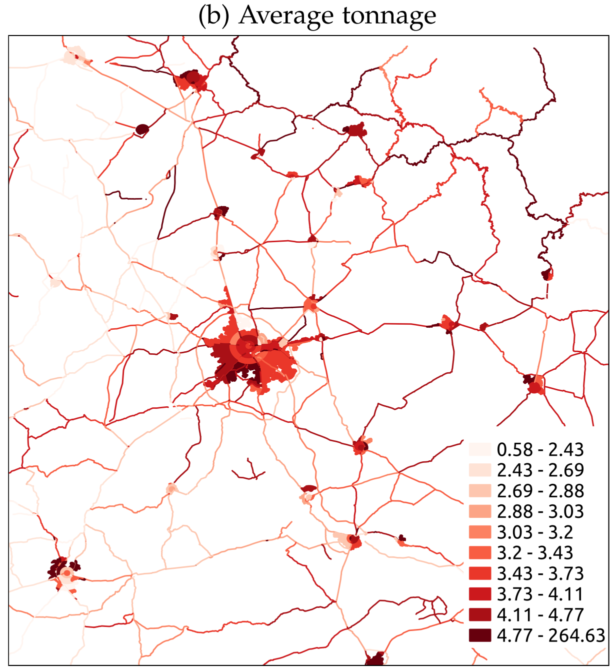
(a) Trip count



(b) Average tonnage



(c) Average rate per ton-km (Rs.)



(d) Average rate per ton-km (Rs.), Delhi area

Figure 7: Statistics of freight transaction data over the road and urban layers

Every segment shows (a) the count or (b,c) average over every routed transaction in the sample that intersects the segment.

Table 3: Shipment-specific determinants of freight cost: traffic speed

	Speed interval (km/h)							
	[0,10]	(10,20]	(20,30]	(30,40]	(40,50]	(50,60]	(60,70]	70+
Cost per km, point estimate	4185.6	144.8	69.2	50.9	36.0	28.4	30.7	27.2
and standard error	(934.4)	(55.2)	(14.4)	(4.1)	(1.5)	(.83)	(.80)	(.78)
Surplus cost (Rs./km)	4158.8	116.0	40.4	22.1	7.2			
Add. minutes per km	10.8	2.8	1.2	0.5	0.1	*(assumed at free flow)*		
Implied time cost (Rs./h)	23,093	2,486.4	2,021.7	2,582.2	3,255.0			

Note: Estimated freight costs (Rs./km) conditional on speed interval (km/h) are obtained from the specification in Table 2, Col. 5 where distance is additionally interacted with eight bins for traffic speed on the same origin-destination pair, but where path routing and varying speed along the route are obtained at the "maneuver"-level from average predictions on Google Maps in October 2022. Note that speed intervals are an attribute of road segments and not the speed of an actual freight transaction. Standard errors are clustered at the level of the district origin and destination pair (28,128 clusters) and all coefficients are significant at the 1% level. We assume that free flow speed for trucks is 50 km/h, which is the speed at which the estimates of cost per speed interval stabilize at their lowest value. We average the estimates above 50 km/h at 28.8 Rs. per km and use this speed and cost to obtain a surplus time, surplus monetary cost, and implied time cost. Calculations are done at the mid-point of each speed interval.

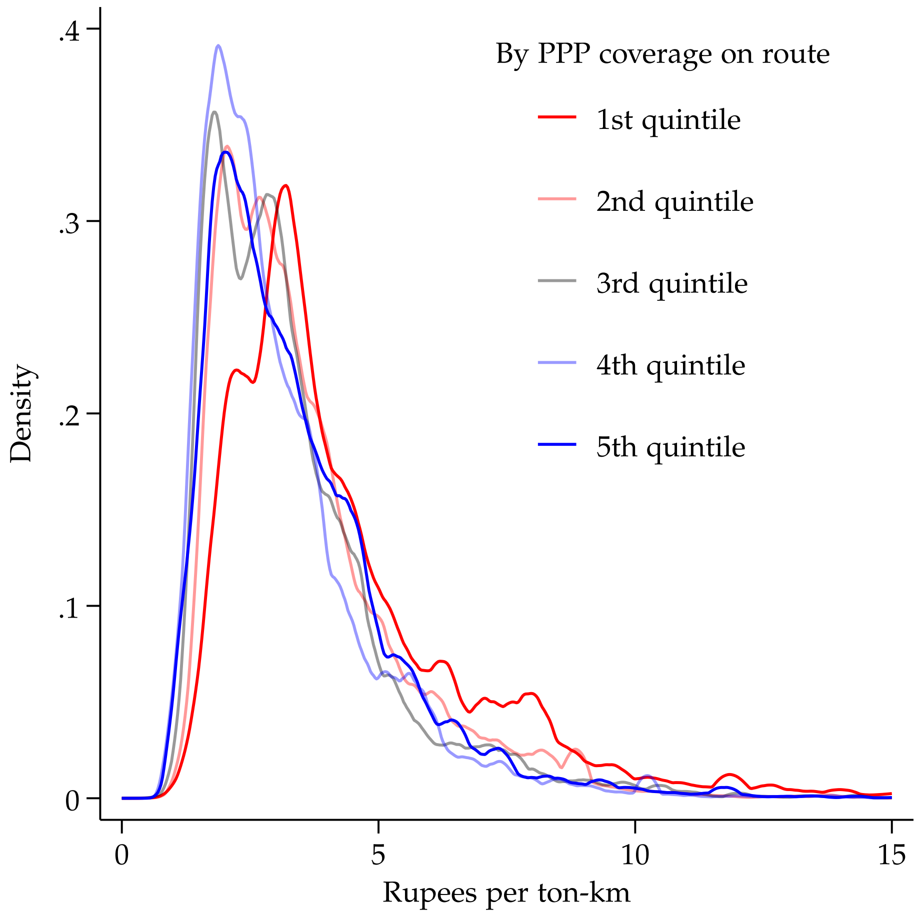


Figure 8: Freight cost by quintiles of Public-Private Parnership (PPP) road coverage along the route.

Statistics from route-km spatially joined to PPP road layers. Rates on freight transactions with cost lower than Rs. 15 per ton-km, which account for 99.1% of all transactions.

Table 4: Road-class determinants of freight cost

	Dependent variable: freight rate (Rs.)				
	(1)	(2)	(3)	(4)	(5) (Distance X 9 cats. spec)
Route distance (km)	33.760[a]	33.562[a]	32.894[a]	33.133[a]	
	(0.636)	(1.587)	(1.577)	(1.583)	
Km on a PPP road	-2.182[a]	1.719[b]			
	(0.736)	(0.840)			
BOT (Annuity)			-5.909[b]	-5.905[b]	-4.457[b]
			(2.589)	(2.588)	(2.252)
BOT (Tolled)			1.626[c]	1.533[c]	-0.136
			(0.916)	(0.915)	(0.787)
Other modes			8.536[a]	8.747[a]	5.230[a]
			(2.270)	(2.273)	(1.939)
Km on A.C. Expressway		-6.460[a]	-6.278[a]	-5.813[b]	-7.223[a]
		(2.195)	(2.329)	(2.327)	(1.810)
Km within city		6.928	7.260	7.441	-5.519
		(4.451)	(4.752)	(4.752)	(3.793)
Km on GQ/NSEW		-5.848[a]	-5.987[a]	-6.025[a]	-4.287[a]
		(0.622)	(0.891)	(0.890)	(0.812)
Km on other NH		-0.964	0.098	0.060	2.230[c]
		(1.648)	(1.625)	(1.625)	(1.256)
Hours per 100 km				1087.128	
				(685.065)	
Spatial controls	Y	Y	Y	Y	Y
Cargo & vehicle type FE	Y	Y	Y	Y	Y
O, D & MY FE	Y	Y	Y	Y	Y
Observations	479160	479160	479160	479132	479160
R^2	0.90	0.90	0.90	0.90	0.93
R^2-within	0.73	0.73	0.73	0.73	0.80

Note: An observation is a freight transaction. Clustered at the level of the district origin and destination pair (28,141 clusters).

Table 5: Road-class determinants of freight cost, net of toll fees

	Dep. var.: freight rate (Rs.) net of predicted toll fees				
	(1)	(2)	(3)	(4)	(5) (Distance X 9 cats. spec)
Route distance (km)	32.922[a]	32.382[a]	31.595[a]	31.935[a]	
	(0.673)	(1.551)	(1.547)	(1.552)	
Km on a PPP road	-8.338[a]	-2.676[a]			
	(0.754)	(0.842)			
BOT (Annuity)			-10.901[a]	-10.895[a]	-9.483[a]
			(2.572)	(2.570)	(2.280)
BOT (Tolled)			-2.623[a]	-2.756[a]	-4.313[a]
			(0.916)	(0.914)	(0.798)
Other modes			3.616	3.920[c]	0.427
			(2.242)	(2.246)	(1.949)
Km on A.C. Expressway		-5.619[a]	-5.751[b]	-5.080[b]	-6.617[a]
		(2.152)	(2.291)	(2.286)	(1.804)
Km within city		-2.059	-2.354	-2.093	-14.533[a]
		(4.461)	(4.770)	(4.768)	(3.912)
Km on GQ/NSEW		-8.517[a]	-8.388[a]	-8.442[a]	-6.750[a]
		(0.613)	(0.891)	(0.890)	(0.820)
Km on other NH		-0.476	0.655	0.606	2.743[b]
		(1.619)	(1.598)	(1.599)	(1.270)
Hours per 100 km				1568.636[b]	
				(670.087)	
Spatial controls	Y	Y	Y	Y	Y
Cargo & vehicle type FE	Y	Y	Y	Y	Y
O, D & MY FE	Y	Y	Y	Y	Y
Observations	479160	479160	479160	479132	479160
R^2	0.89	0.89	0.89	0.89	0.92
R^2-within	0.69	0.70	0.70	0.70	0.77

Note: An observation is a freight transaction. Clustered at the level of the district origin and destination pair (28,141 clusters).

Table 6: Tolling: average cost to freight, and estimated net benefit

Panel A: freight rate regressed on toll counts by quintiles of the toll fee

Fee quintile	RF	Average cost	Net benefit: Point estimate	Net benefit: 95% CI
Q1	-45.5	176.1	221.5	[70.5 , 372.5]
Q2	74.3	272.8	198.5	[-1.4 , 398.4]
Q3	130.9	319.0	188.2	[26.7 , 349.7]
Q4	-616.6	378.8	995.4	[868.1 , 1122.8]
Q5	-69.2	544.4	613.6	[443.7 , 783.5]
A.C. Expw.	-234.4	531.3	765.7	[424.3 , 1107.1]

Panel B: freight rate regressed on toll counts by quintiles of project length (rows) and terciles of the per-km toll fee (columns)

	Price per km: toll fee in Rs. per project length km								
	Bottom third			Mid. third			Top third		
Proj. length	RF	Average cost	Net benefit	RF	Average cost	Net benefit	RF	Average cost	Net benefit
Q1	-293.3	151.8	[-118.4 , 1008.7]	-299.3	172.7	[-44.4 , 988.4]	-477.6	270.8	[435.6 , 1061.2]
Q2	-1155.9	221.1	[732.4 , 2021.8]	637.7	273.8	[-788.7 , 60.9]	-942.7	424.2	[993.7 , 1740.0]
Q3	-357.0	245.6	[303.1 , 902.0]	-372.2	337.1	[354.3 , 1064.4]	2.6	454.6	[-120.3 , 1024.1]
Q4	-282.0	285.4	[246.0 , 888.8]	26.1	387.5	[-24.4 , 747.2]	530.2	553.1	[-454.5 , 500.2]
Q5	-802.8	390.5	[873.8 , 1512.8]	-1088.5	543.7	[1215.1 , 2049.2]	1686.4	620.0	[-1724.3 , -408.5]
NA				267.1	315.3	[-78.0 , 174.3]			
A.C. Expw.				-372.4	531.3	[562.6 , 1244.8]			

Note: Our toll data covers 1,332 toll plazas. Of these, 713 are in the NHAI Toll Information System which includes data on toll fees and project length. We extend the sample to include 118 plaza clusters for access-controlled (A.C.) expressways, and an additional 501 open-loop toll plazas. Fees on A.C. expressways depend on the entry and exit points on the expressway; we obtain fee data from posted rates and project-specific documents. For other open-loop plazas (which are typically on state highways) we predict fees based on road attributes (e.g. PPP category and PPP project length, where available) as well as electronic toll counts and revenue from Electronic Toll Collection Reports of the Indian Highways Management Company Ltd. (IHMCL). Specifications in these tables report on the relationship between freight rates and the number of toll plazas crossed on the truck route. In Panel A, separate coefficients are estimated for toll counts in each quintile of the distribution of toll plaza fees (ranked based on fees for a truck up to 3 axles). In Panel B, tolls are separately counted for quintiles of project length (where data is available) and again by terciles of the toll fee per km. Tolls for which project length is not available are included separately and reported on the sixth row of the table, and A.C. expressway counts are also separately included. Remaining controls and fixed effects as in the baseline specification (Table 2, Col. 5), with the exception that we include a separate distance terms for each of four truck toll fee categories. Standard errors are clustered by origin-destination district pair (28,141 clusters).

Table 7: Urban determinants of freight cost

	Dependent variable: freight rate (Rs.)				
	(1)	(2)	(3)	(4)	(5)
	Population	Mobility	Congestion	Uncongested	Average
Urban metric:	Density	Index	Index	Mobility Ind.	Speed
Route distance (km)	34.430[a]	35.119[a]	34.942[a]	33.830[a]	33.177[a]
	(1.597)	(1.660)	(1.628)	(1.563)	(1.602)
Km under urban metric quintile:					
1st	-27.074	19.203[a]	-13.516	28.490[a]	35.154[b]
	(19.148)	(6.308)	(13.382)	(10.707)	(15.189)
2nd	-27.941[a]	17.371[c]	-20.590[c]	3.069	39.994[a]
	(9.246)	(9.988)	(10.875)	(7.734)	(13.139)
3rd	18.470	-4.910	-29.153[a]	6.385	-0.357
	(13.022)	(10.995)	(9.591)	(9.245)	(7.813)
4th	43.479[a]	-4.279	54.552[a]	-8.951	2.021
	(8.073)	(12.900)	(12.043)	(12.132)	(9.097)
5th	13.694	-43.205[a]	19.606[a]	-12.971	-5.022
	(13.722)	(11.184)	(6.735)	(10.697)	(6.238)
Km on a PPP road	1.489[c]	2.192[b]	0.594	2.292[a]	2.074[b]
	(0.834)	(0.874)	(0.876)	(0.879)	(0.898)
Km on A.C. Expressway	-8.372[a]	-5.956[a]	-5.677[b]	-4.997[b]	-5.381[b]
	(2.232)	(2.186)	(2.346)	(2.368)	(2.309)
Km on GQ/NSEW	-5.185[a]	-6.066[a]	-5.283[a]	-5.812[a]	-5.613[a]
	(0.635)	(0.621)	(0.802)	(0.632)	(0.637)
Km on other NH	-1.998	-2.213	-1.862	-1.437	-0.991
	(1.714)	(1.723)	(1.665)	(1.626)	(1.682)
Spatial controls	Y	Y	Y	Y	Y
Cargo & vehicle type FE	Y	Y	Y	Y	Y
O, D & MY FE	Y	Y	Y	Y	Y
Observations	479160	479160	479160	479160	479160
R^2	0.90	0.90	0.90	0.90	0.90
R^2-within	0.73	0.73	0.73	0.73	0.73

Note: An observation is a freight transaction. Clustered at the level of the district origin and destination pair (28,141 clusters).

Table 8: Determinants of travel time

	Dependent variable: trip duration (hours, web)					
	(1)	(2) Population	(3) Mobility	(4) Congestion	(5) Uncongested	(6) Average
Urban metric:		Density	Index	Index	Mobility Ind.	Speed
Route distance (km)	0.017[a]	0.018[a]	0.018[a]	0.017[a]	0.017[a]	0.017[a]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Km within city	0.005[a]					
	(0.000)					
Km under urban metric quintile:						
1st		0.019[a]	0.002[a]	-0.015[a]	0.016[a]	0.023[a]
		(0.002)	(0.000)	(0.001)	(0.000)	(0.001)
2nd		-0.006[a]	0.018[a]	0.013[a]	0.002[a]	-0.018[a]
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
3rd		0.016[a]	0.013[a]	0.019[a]	0.000	0.010[a]
		(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
4th		-0.004[a]	0.011[a]	0.007[a]	0.010[a]	0.010[a]
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
5th		0.003[b]	-0.012[a]	0.003[a]	-0.009[a]	0.000
		(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Km on a PPP road	-0.003[a]	-0.003[a]	-0.003[a]	-0.003[a]	-0.003[a]	-0.003[a]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Km on A.C. Expressway	-0.002[a]	-0.002[a]	-0.002[a]	-0.002[a]	-0.003[a]	-0.003[a]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Km on GQ/NSEW	-0.000[a]	-0.000[a]	-0.001[a]	-0.001[a]	-0.001[a]	-0.000[a]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Km on other NH	0.003[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.003[a]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spatial controls	N	N	N	N	N	N
O, D & MY FE	N	N	N	N	N	N
Observations	68932	68932	68932	68932	68932	68932
R^2	0.99	0.99	0.99	0.99	0.99	0.99
R^2-within	0.99	0.99	0.99	0.99	0.99	0.99

Note: An observation is a routed origin-destination pair. Clustered at the level of the district origin and destination pair (28,569 clusters).

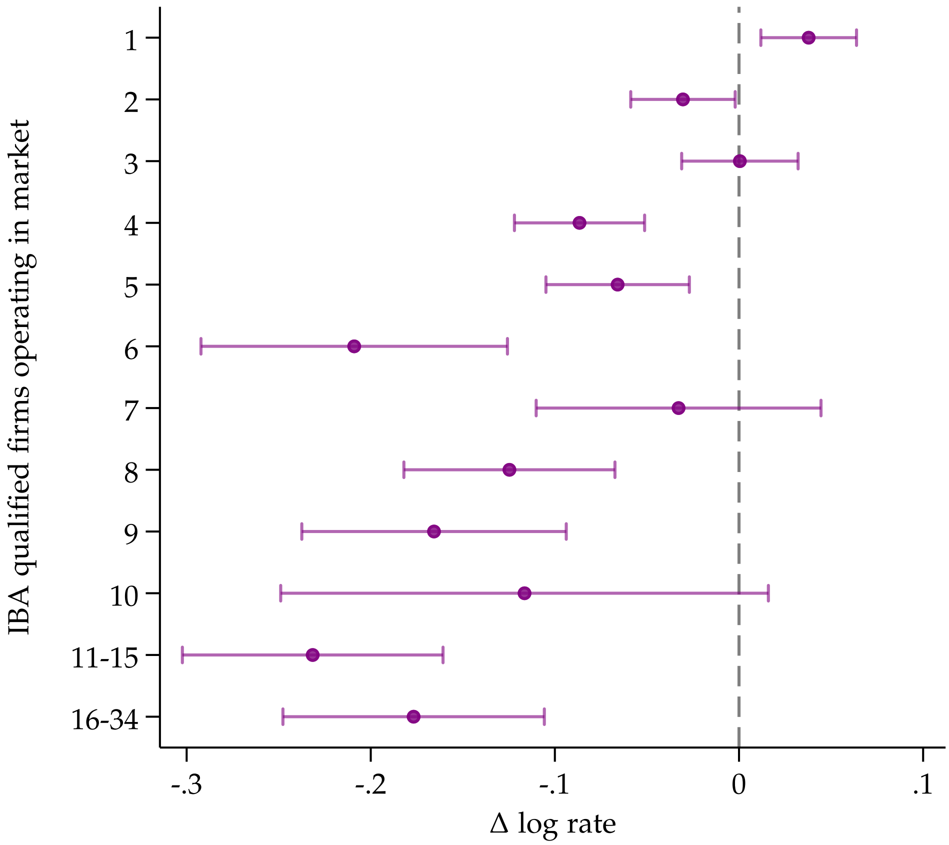
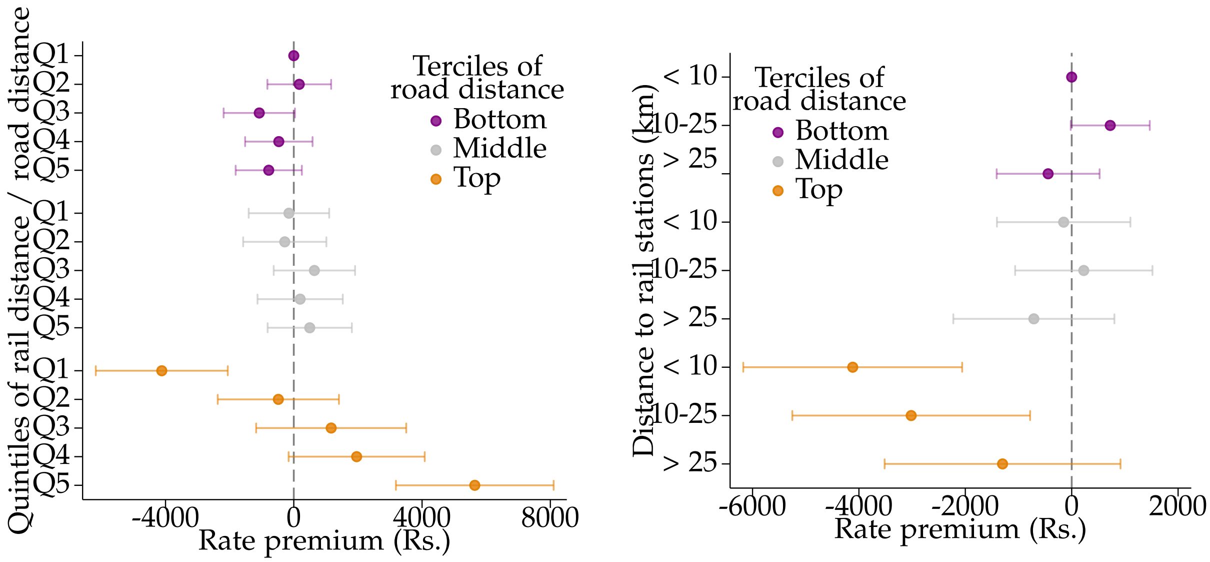


Figure 9: Freight rates by number of Indian Bank Association (IBA) qualified transport operators in the route market

Displayed estimates are the coefficients on an interaction for the number of firms listed by the Indian Bank Association (IBA) that operate on a particular route. Route markets with more than 11 operators are categorized in two interactions. The omitted category is routes with no IBA-listed operators, which account for 68.2% of all trips. IBA "monopoly" routes account for 13.1% of the sample, duopolies for 7.6% and all other categories account for 2.7% or less. Other controls and fixed effects as specified in Table 9, Col. 4. Standard errors are two-way clustered by origin and destination district (28,141 clusters); 95% confidence intervals shown.

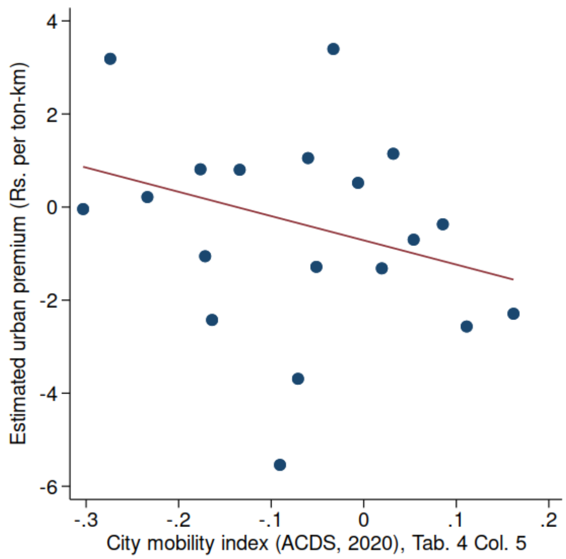


(a) Route circuity

(b) Station access

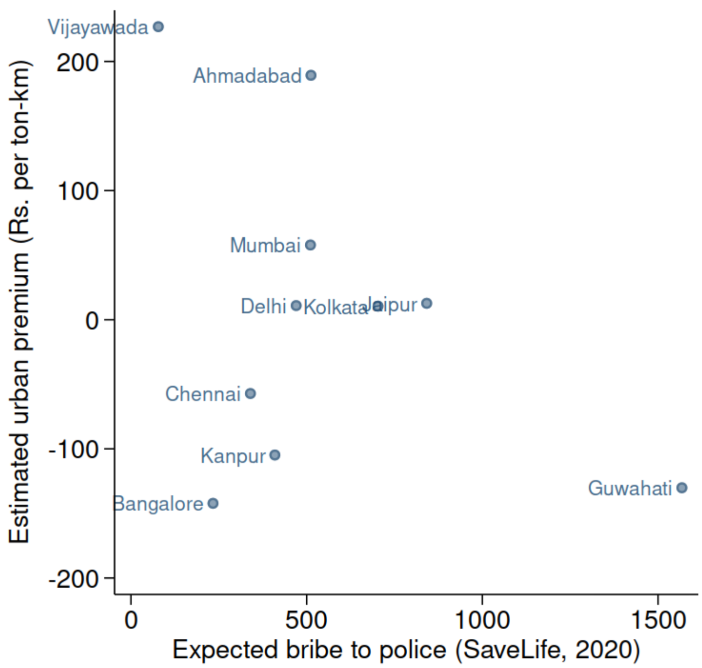
Figure 10: Freight truck rates by proximity to freight rail

Coefficients displayed in Fig. (a) are for quintiles of the ratio between the shortest distance over Indian Railways over the road distance for the same origin-destination pair, interacted with terciles of the road distance. The interaction between the lowest quantiles is the omitted category. Terciles of distance are: 1.1 to 445.1 km, 445.1 to 1081.8 km, and 1081.8 to 3824.3 km. Quintiles of the distance ratio start at 0.02 and switch at 0.95, 1.00, 1.05, 1.13, going up to 31.6. Ratios are less than 1 when the distance between nearest stations is less than the initial distance. Links to rail north of Banihal in Jammu & Kashmir, which is not connected to the national network, are included in the top quintile. Distances between trip endpoints and the nearest train station are summed at origin and destination and split into intervals defined at less than 10km (38.3% of sample), 10 to 25 km (35.2%) and more than 25 km (26.5%). Indicators for the three station access intervals are included in the specification, again interacted by distance, and their estimated coefficients are plotted in Fig. (b). We define access distance < 10 km as the baseline category, so coefficients on circuity quintile #1 in Fig. (a) are identical to coefficients for access distance < 10 km in Fig. (b). The estimated rate difference between the worst and best rail circuity for the longest third of trips is Rs. 9,761, which is 12.3% of the average freight rate for these trips. The estimated rate difference between rail station access of less than 10 km relative to more than 25 km is Rs. 2817, or 3.6% of the average freight rate. Remaining controls and fixed effects are as in the baseline specification (Table 2, Col. 5). Standard errors are clustered by origin-destination district pair (28,141 clusters), 95% confidence intervals shown.



(a) City-specific urban premia and ACDS mobility, binned scatterplot

(b) City-specific urban premia and ACDS mobility, scatter plot for statically significant premia (blue) and selected non-significant premia (red)



(c) City-specific urban-premia and expected police bribe (SaveLife Foundation, 2020)

Figure 11: Channels

Table 9: Other determinants of cost

	Dependent variable:			
	Freight rate (Rs.)	Log freight rate (Rs.)		
	(1)	(2)	(3)	(4)
Distance, grade to .5%	33.087[a]			
	(0.797)			
Distance, grade to 1%	28.994[b]			
	(13.275)			
Distance, grade to 2%	95.530			
	(88.009)			
Distance, grade to 10%	442.508[c]			
	(264.284)			
Log route distance (km)		0.458[a]	0.449[a]	0.455[a]
		(0.010)	(0.010)	(0.010)
Log tonnage		0.414[a]	0.406[a]	0.411[a]
		(0.023)	(0.024)	(0.024)
Log rain shock$_{\text{omy}}$		0.001		
		(0.001)		
Log rain shock$_{\text{dmy}}$		0.003[b]		
		(0.001)		
Log prob. common language			-0.031[a]	
			(0.005)	
No. IBA listed operators				-0.010[a]
				(0.003)
Spatial controls	Y	Y	Y	Y
Cargo & vehicle type FE	Y	Y	Y	Y
O, D & MY FE	Y	Y	Y	Y
Observations	475,014	479,046	479,160	479,160
R^2-within	0.90	0.92	0.92	0.92

Col 2. specification drops 114 trips with records of no rain, and includes dummy interactions for endpoint districts that registered no rain during the trip month. An alternative specification that drops such no rain events (9,234 cases) has identical coefficient estimates.

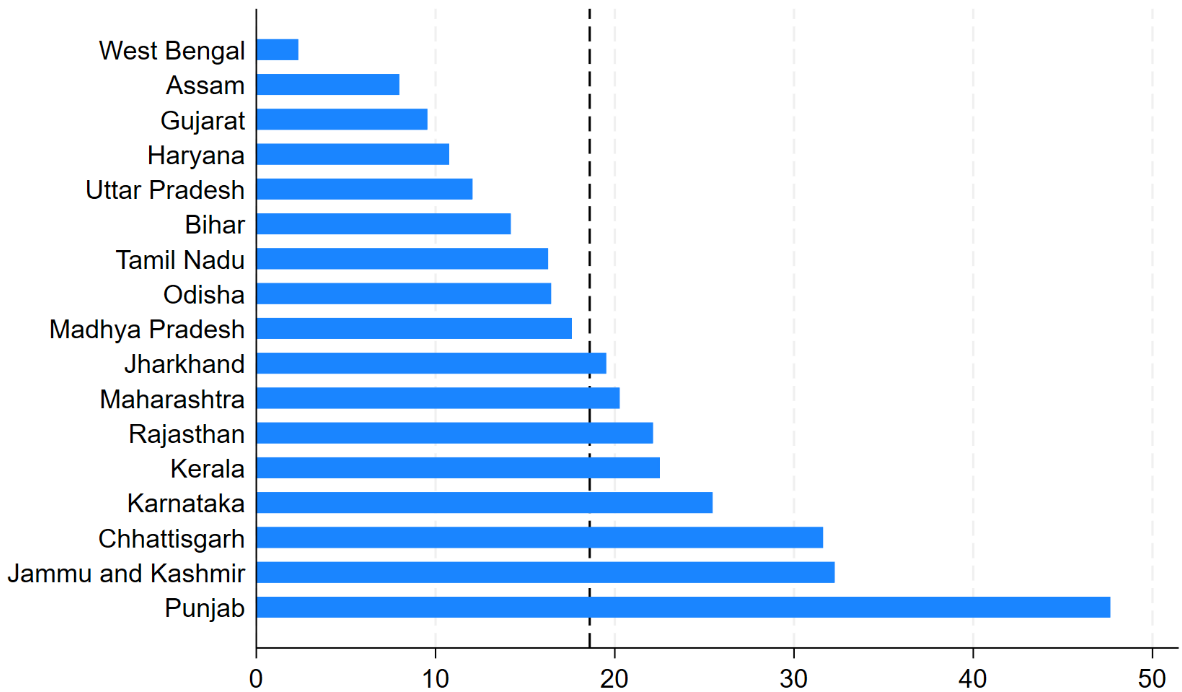


Figure 12: Percentage of state freight costs explained by state-specific spatial covariates

Dashed line at 18.6% indicates the average across states. [TODO: add details on elements included, and base cost.]

Appendix A Secondary data sources

TCI survey data, 2014-2015. We collect all rate and distance statistics provided in the Annex of Transport Corporation of India (TCI) and IIM Calcutta (2015). Rates for this report were surveyed by TCI for 30 trips in the Delhi-Bangalore and Delhi-Mumbai lanes (15 in each direction) and 4 trips (2 in each direction) in each of 26 other markets. Comparisons with digital freight rates are adjusted by the Wholesale Price Index, with monthly adjustment factors through December 2019 for the digital freight rates and an annual adjustment from 2015 to 2019 for TCI data.

Rail data. We obtain data on railway freight terminals from www.fois.indianrail.gov.in/RailSAHAY/Home.jsp, including both Indian Railways and private freight terminals. Terminal data includes location and cargo handled for loading and unloading. We obtain shortest paths on the rail network (allowing gauge transshipment) from the Indian Railways Rate Branch System, available at rbs.indianrail.gov.in/ShortPath/ShortPath.jsp.

Appendix A.1 Diesel price data

We estimate retail prices for diesel for each state s and in each month t using the following system of equations:

$$\text{Base price}_t = \text{Crude price}_t \times \overline{\text{markup}}_t \tag{5}$$

$$\begin{aligned}\text{Pre-VAT price}_t = \; & \text{Base price}_t \times (1 + \text{Dealer commission rate}_t) \\ & + \text{Dealer commission fee}_t + \text{Central excise}_t\end{aligned} \tag{6}$$

$$\begin{aligned}\text{Retail price}_{t,s} = \; & (\text{Pre-VAT price}_t \times (1 + \text{VAT rate}_{t,s}) + \text{VAT excise}_{t,s}) \\ & \times (1 + \text{VAT surcharge}_{t,s})\end{aligned} \tag{7}$$

In a first step we use Eqs. 5-7 and the monthly retail price of diesel in Delhi, Mumbai, Chennai, and Kolkata to back out a time series of monthly markups for each city, given the structure of national and state taxes and fees described by Eqs. 6 and 7, respectively. We take the geometric mean of the four cities as a time series for a national markup ($\overline{\text{markup}}_t$). Given this national markup we use state-level tax and fee data and the system given by Eqs. 5-7 to estimate the monthly retail price of diesel for each state. We compare our predictions for state-level retail prices to data on the price of diesel in state capital cities, as published by the Indian Oil Corporation Limited (IOCL) from July 2016 to June 2018. The correlation across states and periods is 0.975.

Our source data is from the Petroleum Planning & Analysis Cell (PPAC) and IOCL. PPAC reports retail prices for Delhi, Mumbai, Chennai and Kolkata from June 2017 to 2023, while IOCL reports these prices from June 2002 through September 2017 (values in both datasets are identical when they overlap). PPAC reports historical price data on crude oil, a statistic on dealer commissions for diesel, central government excises, and current state tax rates. PPAC also releases a *Ready Reckoner* report which includes updates on state taxes every every June and

November. We search news media to obtain precise timing on changes to central government excise taxes. State tax structures vary over time, and the tax structure in some states depends on the price level (typically, by setting a maximum excise tax that binds at some level of an ad-valorem rate). We adjust the rate and excise in Eq. 7 to implement state rules.

Appendix A.2 Toll plaza and fee data

We pool data from multiple sources to construct a dataset with locations and time series for toll fees for 1,328 toll plazas in India.

Toll plazas. Our baseline toll data is from the National Highway Authority of India (NHAI), which posts data on the tolls under its authority. We download current and retrieve archived versions of the NHAI's Toll Information System website, collecting data on 790 tolls for which we obtain attributes such as each project's name, location, fees, key dates, tollable project length and concessionaire.

We obtain additional administrative data from two administrative sources. We obtain data on a sample of toll locations that were planned or operational as of 2017 from the National Center Of Geo-Informatics (NCOG). We also obtain current and past versions of the National Payments Corporation of India's toll plaza master file, which indicates the coordinates and the name of the concessionaire of all toll plazas where electronic fast tags are available for sale. We match these to the NHAI data while using the location, name and characteristics of each plaza to remove potential duplications. This process adds an additional 329 toll plazas, in particular for plazas on state roads.

Lastly, we combine this data with all toll plazas from OpenStreetMap (OSM), allowing us to improve the precision of locations from the administrative data sources and add an additional 209 tolls. We define groupings for the OSM data, as OSM includes data on toll booths rather than plazas.

The NHAI website posts toll fees and their validity period. All NHAI tolls with fee that had not been updated since 2020, or tolls that were seen to not be included in the most recent archive of the NHAI website were manually researched through media articles and historical imagery on the Google Earth Desktop application to determine dates or date ranges for the opening and closing of each toll plaza.

Toll fees. Data on toll fees for the period between January 2017 and December 2022 was collected or predicted for each toll, using data from two sources: current and historical posted fees from NHAI, and per-plaza toll revenues and counts from the National Electronic Toll Collection (NETC) system.

The method used to determine fees for each plaza depended on data availability and the class of tolling system. In an "open" toll system the driver pays whenever they cross a toll plaza, regardless of the distance traveled on the road. This type of toll is the most common in national and state highways and makes up more than 90% of the tolls in our data. In contrast, closed-system tolls, also known as point-to-point tolls, require drivers to enter and exit a road at different toll plazas, with the fee depending on the distance traveled. This type of toll is more commonly used in access-controlled expressways. There are 27 access controlled expressways in the data, and tolls for these were researched individually using current and archived posted rates, official project documents and news sources.

All tolls in the NHAI database are open-system tolls. Of these, 520 tolls have recently updated fee data, including annual updates in both April 2021 and April 2022. Data from these tolls was used to estimate fees for tolls with incomplete data. While the timing of fee updates may vary, all tolls are updated annually using a formula based on the Wholesale Price Index (WPI). Tolls have multiple types of fees, both for each category of vehicle and the frequency of journey, but all charges are determined by a single fare. NHAI tolls account for about half the open-system tolls in the dataset.

For tolls without NHAI information, NETC revenue data was used to calculate toll charges for 2021, incorporating data for an additional 50 tolls. The toll fees calculated from the ETC dataset align with those in NHAI, as shown in Figure A.1.

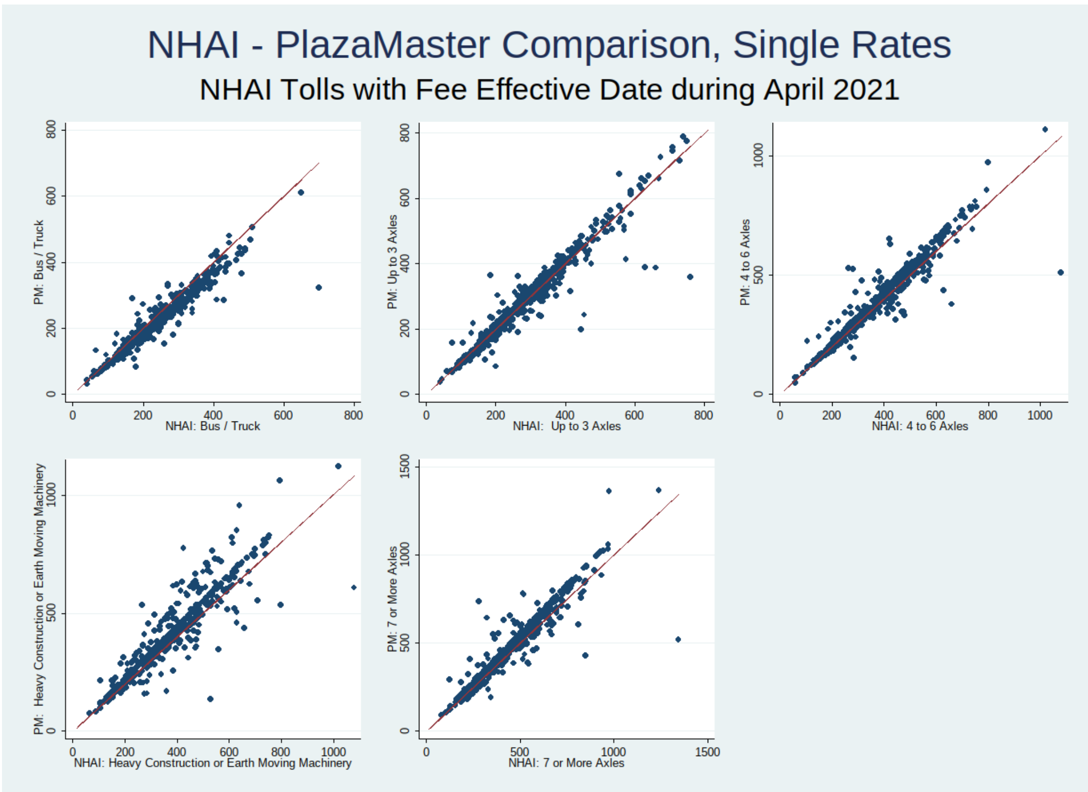


Figure A.1: Comparison between posted fees (NHAI) and per-unit revenues (NETC)

Discrepancies between these two sources may arise because we do not observe the share of revenue and counts in the NETC data that comes from single, return or monthly fees, or for special pricing for commercial vehicles registered in the plaza's district.

For the 492 tolls that are not in the NHAI or NETC data the fees in 2021 were predicted after running a regression on the sample of NHAI and NETC plazas. We regress fess on the plaza's state, road type (which may be a PPP and/or on the GQ or NS/EW highways) and, for PPP roads where this is feasible, the project length per project plaza. Given estimates for 2021 we back-cast these to the period between 2017 and 2020 following the NHAI 2008 rules for toll fee updates. The NHAI 2008 sets out a formula by which toll fees increase each April by the product of two figures: the first is 0.03% multiplied by the number of years that have passed since 2007, and the second is 40% of the increase in the Wholesale Price Index (WPI) in the last calendar year. Among the tolls in our dataset, 601 follow the 2008 rules and 92 existed prior to 2008 and follow the 1997 rules, which did not have a standard method for updating toll fees. Tolls under the 1997 rules were set to be updated in September or July using either the March year-on-year increase in the WPI or the increase in the average WPI between April and March. The March year-on-year increase method is more common, based on fee updates between 2021 and 2022 on the NHAI website. Therefore, tolls in our database were back-casted with an update every April using the 2008 formula, except when the available data explicitly stated that the toll followed the 1997 rules or had fee updates on a different month.

Appendix B Figures and tables (for online publication)

Table B.1: Shipment-specific determinants of freight cost, logs

	Dependent variable: Log of freight rate (Rs.)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Route distance (km)	0.644[a]	0.532[a]	0.531[a]	0.457[a]	-0.943[a]		0.445[a]	0.508[a]
	(0.014)	(0.013)	(0.013)	(0.010)	(0.042)		(0.009)	(0.009)
Log Route distance sq.					0.119[a]			
					(0.004)			
Log Tonnage	0.509[a]	0.414[a]	-0.197[b]	-0.230[a]	-0.215[a]	0.075	-0.235[a]	-0.246[a]
	(0.016)	(0.026)	(0.087)	(0.076)	(0.074)	(0.097)	(0.076)	(0.076)
Log Tonnage sq.			0.120[a]	0.126[a]	0.118[a]	0.046[b]	0.127[a]	0.128[a]
			(0.019)	(0.016)	(0.015)	(0.023)	(0.016)	(0.016)
Log Distance (km) X								
0 to 10 tons						0.402[a]		
						(0.010)		
10 to 20 tons						0.432[a]		
						(0.014)		
20 to 30 tons						0.478[a]		
						(0.010)		
>30 tons						0.494[a]		
						(0.010)		
Within state				0.048[a]	-0.028[a]	0.054[a]	0.047[a]	0.047[a]
				(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Within district				0.581[a]	0.112[c]	0.602[a]	0.560[a]	0.523[a]
				(0.068)	(0.058)	(0.065)	(0.066)	(0.061)
No. states visited				0.078[a]	0.013[a]	0.080[a]	0.081[a]	0.066[a]
				(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Urban origin				-0.002	-0.001	-0.004	-0.003	0.001
				(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Urban dest.				-0.009	-0.002	-0.011	-0.010	-0.009
				(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
Platform owned fleet		-0.031[a]	-0.030[a]	-0.050[a]	-0.060[a]	-0.046[a]	-0.050[a]	-0.049[a]
		(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Log Hours per 100 km								0.577[a]
								(0.041)
Cargo & vehicle type FE	N	Y	Y	Y	Y	Y	Y	Y
O, D & MY FE	N	Y	Y	Y	Y	Y	Y	Y
Observations	479187	479160	479160	479160	479160	479160	479132	479132
R^2	0.75	0.90	0.90	0.92	0.93	0.92	0.92	0.92
R^2-within	0.75	0.76	0.76	0.80	0.82	0.80	0.80	0.81

Note: An observation is a freight transaction. Clustered at the level of the district origin and destination pair (28,141 clusters). TODO: state means.

Table B.2: Shipment-specific determinants of freight cost per ton

	Dependent variable: Rate per ton (Rs.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Route distance (km)	2.305[a]	1.984[a]	1.941[a]		1.825[a]	1.979[a]
	(0.040)	(0.017)	(0.034)		(0.032)	(0.035)
Route distance sq.						
Tonnage	-73.255[a]	-47.345[a]	-46.704[a]	-373.011[a]	-46.444[a]	-46.716[a]
	(2.140)	(3.363)	(3.315)	(13.398)	(3.307)	(3.332)
Tonnage sq.				6.306[a]		
				(0.244)		
Distance (km) X						
0 to 10 tons				3.030[a]		
				(0.052)		
10 to 20 tons				1.887[a]		
				(0.034)		
20 to 30 tons				1.554[a]		
				(0.031)		
>30 tons				1.370[a]		
				(0.038)		
Within state			29.991	0.083	38.848	10.428
			(25.121)	(24.653)	(25.313)	(25.129)
Within district			389.834[a]	247.828[a]	364.961[a]	269.434[a]
			(51.916)	(66.476)	(49.461)	(50.590)
No. states visited			26.202[b]	32.258[a]	53.606[a]	17.798
			(13.017)	(11.682)	(12.708)	(13.161)
Urban origin			-25.054	1.063	-25.910	-21.853
			(15.805)	(14.732)	(16.151)	(15.813)
Urban dest.			-61.277[a]	-38.945[b]	-62.614[a]	-58.770[a]
			(18.073)	(17.612)	(18.104)	(17.897)
Platform owned fleet		-239.630[a]	-242.400[a]	-203.066[a]	-244.187[a]	-241.173[a]
		(25.084)	(24.941)	(25.234)	(25.546)	(24.882)
Hours per 100 km						232.745[a]
						(34.343)
Cargo & vehicle type FE	N	Y	Y	Y	Y	Y
O, D & MY FE	N	Y	Y	Y	Y	Y
Observations	479187	479160	479160	479160	479132	479132
R^2	0.69	0.88	0.88	0.90	0.88	0.88
R^2-within	0.69	0.69	0.69	0.76	0.69	0.70

Note: An observation is a freight transaction. Clustered at the level of the district origin and destination pair (28,141 clusters). Todo: state means.

Table B.3: Shipment-specific determinants of freight cost per ton-km

	Dependent variable: Rate per ton-km (Rs.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Route distance (km)	-0.002[a]	-0.002[a]	-0.002[a]		-0.002[a]	-0.000
	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Route distance sq.						
Tonnage	-0.093[a]	-0.102[a]	-0.081[a]	-0.486[a]	-0.082[a]	-0.082[a]
	(0.005)	(0.028)	(0.019)	(0.045)	(0.019)	(0.016)
Tonnage sq.				0.007[a]		
				(0.001)		
Distance (km) X						
0 to 10 tons				-0.003[a]		
				(0.000)		
10 to 20 tons				-0.002[a]		
				(0.000)		
20 to 30 tons				-0.002[a]		
				(0.000)		
>30 tons				-0.001[a]		
				(0.000)		
Within state			1.367[a]	1.431[a]	1.358[a]	0.611[b]
			(0.132)	(0.121)	(0.132)	(0.272)
Within district			19.029[a]	19.211[a]	19.057[a]	14.381[a]
			(3.278)	(3.290)	(3.276)	(2.207)
No. states visited			0.240[a]	0.237[a]	0.208[a]	-0.087
			(0.048)	(0.047)	(0.046)	(0.114)
Urban origin			-0.032	-0.056	-0.031	0.090
			(0.085)	(0.081)	(0.085)	(0.096)
Urban dest.			-0.358[b]	-0.377[b]	-0.358[b]	-0.262[b]
			(0.153)	(0.157)	(0.153)	(0.117)
Platform owned fleet		-0.207	-0.241[c]	-0.246[b]	-0.240[c]	-0.196[c]
		(0.174)	(0.132)	(0.111)	(0.132)	(0.115)
Hours per 100 km						8.989[a]
						(2.252)
Cargo & vehicle type FE	N	Y	Y	Y	Y	Y
O, D & MY FE	N	Y	Y	Y	Y	Y
Observations	479187	479160	479160	479160	479132	479132
R^2	0.084	0.21	0.35	0.35	0.35	0.43
R^2-within	0.084	0.091	0.25	0.26	0.25	0.34

Note: An observation is a freight transaction. Clustered at the level of the district origin and destination pair (28,141 clusters). Todo: state means.

Appendix C Spatial database of Indian highway projects

See attached document. This database will be released publicly and referenced in subsequent versions of this paper.