

# Misallocation, Internal Trade, and the Role of Transportation Infrastructure\*

Jose Asturias

School of Foreign Service in Qatar, Georgetown University

Manuel García-Santana

Université Libre de Bruxelles (ECARES)

Roberto Ramos

Bank of Spain

January 2014

## Abstract

We investigate the role of transportation infrastructure in explaining resource misallocation and income in India. We extend the endogenous variable markups model by [Atkeson and Burstein \(2008\)](#) into a multi-region setting in which asymmetric states trade with each other. High transportation costs that result from poor infrastructure quality generate misallocation of resources by increasing dispersion in market power across firms. Using a rich micro-level dataset constructed from manufacturing and geospatial data, we find preliminary evidence that is consistent with our theory. Using the construction of the Golden Quadrilateral as a natural experiment, we find that prices declined by 20% in districts crossed by this road. We calibrate the model and simulate an improvement in Indian road quality. We find the aggregate gains for improved road quality. We also decompose these gains into Ricardian and pro-competitive components.

*JEL classifications:* F1, O4.

*Keywords:* Misallocation, Total Factor Productivity, Internal Trade, Infrastructure.

---

\*We thank the CEPR for financial support of this project and the Private Enterprise Development in Low-Income Countries (PEDL) program. The authors would like to thank Paola Conconi, Arnaud Costinot, Timothy Kehoe, Francis Kramarz, Josep Pijoan-Mas, James Schmitz, and seminar participants held in ECARES, Midwest Macro Fall 2013 (University of Minnesota), and Spanish Economics Association Meeting 2013 (Santander). Asturias: [jga35@georgetown.edu](mailto:jga35@georgetown.edu); García-Santana: [msantana@ulb.ac.be](mailto:msantana@ulb.ac.be); Ramos: [roberto.ramos@bde.es](mailto:roberto.ramos@bde.es).

# 1 Introduction

It is well documented that low income countries use resources less efficiently than high income countries. In fact, differences in aggregate Total Factor Productivity (TFP) account for around 50-60% of income inequality across countries. Driven by this observation, many recent works have studied quantitatively how much of these differences in TFP across countries can be explained by how resources are allocated across heterogeneous production units. In a very influential work, [Hsieh and Klenow \(2009\)](#) use a standard [Melitz \(2003\)](#) type of model and plant level data to show that misallocation of resources across production units in India and China explain around 50% and 30% of the TFP gap between these countries and the US.

In this paper, we study to what extent the “observed” misallocation is explained by a common feature of low income countries: poor transportation infrastructure. In an economy where transportation costs are high, local firms enjoy market power that allows them to charge high mark-ups in the local market. This creates a wedge between the mark-up of firms operating at home and the mark-up of firms operating in other locations. When there is an improvement in infrastructure transportation costs decline, reducing the “cost advantage” of local firms and equalizing mark-ups across all producers.

We develop this idea in a multi-region trade model with endogenous markups and calibrate it to the Indian economy. In our model, in any given sector there is a finite number of firms competing à la Cournot. In this context, demand elasticity and hence market power of a given producer depends on its sectoral share. More productive firms have higher sectoral shares and charge higher markups. Poor transportation infrastructure provide a cost advantage to local firms: the “effective” productivity of firms producing outside the location gets reduced due to transportation costs.

Our model is similar to that recently used by [Edmond, Midrigan, and Xu \(2012\)](#) and [Atkeson and Burstein \(2008\)](#). We extend the model to include multiple non-symmetric countries. In this model, the gains from trade crucially depend on (i) the amount of initial misallocation and (ii) the pattern of comparative advantage across the states trading among each other. The model needs to generate misallocation in the first place, so that trade can reduce it. However, even if there is substantial misallocation in the first place, reducing transportation costs affect allocation of resources only if firms operating abroad are potential effective competitors of local firms. In a context in which the pattern of comparative advantage is very strong, firms producing in every single location are very productive in the production of a particular good relative to firms producing everywhere else. Then, when transportation costs decline, the market distribution of market shares and hence mark-ups does not change much, implying small changes in the allocation of resources.

We find preliminary evidence in the data that is consistent with our theory. We use the construction of the Golden Quadrilateral (GQ), a large project improving the road infrastructure in India, as a natural experiment. The project aimed to upgrade the highway network connecting the four major cities in India. We apply a differences-in-differences strategy to study the effects of the

GQ on prices. Endogeneity of infrastructure is of obvious concern, so we adopt a strategy commonly used in the empirical literature. In this strategy, we think that the goal of the construction of the GQ to connect the four major cities of India, which are the nodal points. If a non-nodal district happens to lie along that path, the change in road quality is exogenous to that district. We find that the prices in non-nodal districts crossed by the GQ were 24 percentage points lower than those district further away.

We calibrate the model to the Indian economy using plant-product-level data and geo-spatial data. First, we exploit structural equations of the model to infer transportation costs among Indian states and the elasticity of substitution across sectors. Both transportation costs and the elasticity of substitution across sectors play a key role in determining the amount of misallocation in the model. As mentioned above, higher transportation costs generate a higher variation in mark-ups in the economy, and hence more misallocation.

In order to identify transportation costs, we use the implication of the model that differences in prices for a monopolist firm across destinations reflect differences in transportation costs. To implement this strategy, we first identify all firms in the data that have very high levels of the marketshare of production in India. Using pricing data from only these identified monopolists, we regress the prices they charge across different locations against a measure of effective distance. Measures of effective distance include geographical distance and the quality of transportation infrastructure. Using the coefficients of the regression we construct a matrix of bilateral transportation costs between Indian states.

As a next step, we identify the elasticity of substitution across sectors. For goods produced by monopolists, we are able to construct trade flows across Indian locations, since we observe where the good is being used as an intermediate input. For these goods the model implies a gravity equation that relates bilateral flows to transportation costs, a strategy similar to that of [Eaton and Kortum \(2002\)](#). We use interal trade flows and the estimated transportation costs to measure how trade flows decline with increases in transportation costs.

We calibrate the rest of parameters of the model to match relevant statistics of the Indian manufacturing sector. Conditional on our estimated values for the elasticity of substitution and transportation costs, the amount of misallocation in the model depends on the elasticity of substitution within sectors and the the shape of the distribution of productivities. For a given value of the elasticity of substitution across sectors, the higher the elasticity within sectors the higher the variation dispersion in mark-ups in the economy, since there will be larger range of market power levels in the economy. The productivity distribution across firms plays a key role as well. For a given value of the elasticities within sector and across sector, the higher the variation in firms productivity, the higher the variation in markups and hence the higher the amount of misallocation.

We use the calibrated version to get some preliminary results from the model. Again, the goal is to quantify the TFP and income gains of a reduction in transport costs. We simulate

an improvement in the quality of roads so that every state in India has the best infrastructure that currently exists in country. We find that the average state gains 1-20% depending on the specification used. We also find that small states gain the most. We also decompose these gains into Ricardian and pro-competitive components. In general, we find that pro-competitive gains are small but this is sensitive to the specification used.

## 2 Related literature

Our paper is related to several strands of literature. First, it contributes to the recent literature emphasizing misallocation of resources across firms as an important source of income differences across countries. Restuccia and Rogerson (2008) and Guner, Ventura, and Yi (2008) were the first to show that policies distorting the allocation of resources across heterogeneous firms can generate high TFP and income losses. Hsieh and Klenow (2009) showed that, when looking at plant level data through the lens of a simple model of heterogeneous firms, misallocation in India and China can account for a big share of the TFP differences with respect to the US. Motivated by these works, there has been many papers focusing on different government policies and distortions as the source of misallocation, and quantifying the implied TFP losses.<sup>1</sup>

For instance, Buera, Kaboski, and Shin (2011), Midrigan and Xu (2014), Moll (Forthcoming), Caselli and Gennaioli (2013), Erosa and Allub (2013), and Lopez-Martin (2013) focus on financial frictions. Gourio and Roys (2013); Garicano, Lelarge, and Reenen (2012) and García-Santana and Pijoan-Mas (2012) study the marginal effect of size-dependent policies in France and India respectively. Our paper is also related to Peters (2013), who calibrates a model of imperfect competition with heterogeneous firms to Indonesian data to investigate the impact of misallocation on growth.

Our paper also builds on a large set of work that studies the gains from international trade in a context of imperfect competition (e.g. Markusen (1981); de Blas and Russ (2010); Holmes, Hsu, and Lee (2012); Devereux and Lee (2001); Epifani and Gancia (2011)). Two recent papers that are closely related to ours are Edmond, Midrigan, and Xu (2012) and Donaldson (Forthcoming). The former calibrate a similar model to the one we use to Taiwanese plant-level data to quantify gains from trade. The latter applies Eaton and Kortum (2002) type of model to investigate the impact of transportation infrastructure in the colonial India.

Finally, our diff-in-diff strategy is very similar to the one used in recent works whose aim is to investigate the impact of transportation infrastructure. In particular, the construction of the GQ project has attracted the attention of several authors who have applied this type of empirical strategy. Datta (2012) uses the Enterprises Surveys of the World Bank to investigate the impact of the GQ on firms' average stock of input inventories. Ghani, Goswami, and Kerr (2013) uses several rounds of the Annual Survey of Industries to document an increase in entry rates and

---

<sup>1</sup>See Restuccia and Rogerson (2013) for a nice survey of the literature.

average plants productivity in districts located near the GQ. Alder (2014) uses luminosity data to study the impact of the GQ on Indian regions.

### 3 Datasets

In this section, we describe the datasets we use in the paper. We first describe the plant level data, ASI and NSS, and we next present a description of the geographic and infrastructure variables we use.

#### 3.1 Plant level data representative of the Indian manufacturing sector

We use a representative dataset of the Indian manufacturing sector which is constructed out of two different plant-level datasets. We next carefully describe both, and explain how we aggregate them in order to have a sample representative of the whole Indian manufacturing sector.

##### 3.1.1 Annual Survey of Industries (ASI)

The ASI has been collected annually by Indian Government's Central Statistical Organization since 1959, except for the year 1972. The ASI covers all the Indian territory except for the states Arunachal Pradesh, Mizoram, and Sikkim and Union Territory of Lakshadweep. ASI is the main source of manufacturing statistics in India.<sup>2</sup> The unit of enumeration in ASI is a plant, targeting all plants employing 10 or more workers using power and those employing 20 or more than 10 workers without using power.<sup>3</sup> ASI has two parts: the ASI census and the ASI sample. Plants with 100 or more workers are categorized as the census sector, which means that all of them are surveyed. In order to account for the rest of registered plants, all registered plants with less than 100 employees are randomly sampled. The sample frame is carefully designed: all plants are stratified at the sector-industry 4 digit level of NIC and at least 1/5th of the plants in each strata are selected for the sample.<sup>4</sup> The information provided by the establishments is very rich, covering several operational characteristics: sales, employment, capital, remuneration to workers, expenditure in intermediate goods, etc.

---

<sup>2</sup>This dataset has become commonly used in the development literature. See for instance Aghion, Burgess, Redding, and Zilibotti (2005), Chari (2011), Hsieh and Klenow (2009) and Bollard, Klenow, and Sharma (2013).

<sup>3</sup>According to India's Factories Act of 1948, establishments with more than 10 workers (20 if without power) are required to be registered. It means that ASI can be seen as a dataset covering the formal sector in India.

<sup>4</sup>The data reported by the plants that is used to construct ASI is actually carefully monitored by the National Sample Survey Organization, which belongs to the Ministry of Statistics and Programme Implementation: (i) when plants report their records, they are initially verified by the field staff; (ii) verified information by the field staff is then manually scrutinized by senior level staff; (iii) the data is sent to the data centre where it is verified again before enter it in the computers; (iv) once the data is entered, the members of the IT team look anomalies and check consistency with previous surveys.

### 3.1.2 National Sample Survey (NSS)

Taking advantage of the high cover of the economic census that the Indian Government was carrying out since 1977, an integrated survey on informal manufacturing enterprises was first created in 1994, as part of the 51th round of National Sample Survey. NSS covers all informal establishments in the Indian manufacturing sector, where “informal” refers to all manufacturing enterprises not covered by ASI. The NSS covers all Indian districts except for Ladakh and Kargil districts in the state of Jammu and Kashmir. One of the main features of the sample frame is the use of a list frame approach, which consists on constructing a list with the biggest plants (around 8,000) under the coverage of the survey. This is done to make sure of including in the sample a sufficient large number of “relatively” big plants and hence to have a representative sample in terms of plant size.<sup>5</sup> For the rest of the informal manufacturing plants in the universe (other than the ones including on the list), an area sample frame is carried out. For rural zones, frame areas are defined by the list of villages as per census 2001, whereas for urban zones frame areas are defined by urban frame sample blocks. Once the frame areas are defined, a stratified multi-stage sampling is carried out, and sampling weights are provided in order to have population estimates of the whole informal manufacturing sector. As in the case of the ASI, the information provided by the establishments is very rich, covering both variables related to production, use of labor and capital, and expenses in intermediates.

### 3.1.3 Aggregating ASI and NSS

Most of the information extracted from ASI can be easily aggregated with the one from NSS. The reason is that the questionnaires filled by both formal and informal manufacturing establishments are remarkably similar. Since for every single plant that appears in ASI and NSS population weights are provided, we can compute aggregate measures of number of plants operating, number of employees, production at the industry-product level, use of intermediate goods, and prices of both outputs and inputs at the district, state, and country level. Table I shows some descriptive statistics computed for both ASI and NSS for 2000-01 and 2005-06. After merging ASI and NSS we have around 190,000 observations for the year 2000-01 and 140,000 observations for the fiscal year 2005-06. Once these observations are properly weighted, we have more around 17 million manufacturing plants in our data, which employ around 45 million of workers. Of this total level of employment, around 83% and 80% is accounted for by informal plants in 2000-01 and 2005-06 respectively. However, these informal plants only account for around 20% of total value added, which shows, unconditional on industry of operation, the huge differences in productivity between formal-informal plants in India.

---

<sup>5</sup>This list was prepared based on information collected on the census of Small Scale Industries in 2003 conducted by the Development Commissioner of Small Scale Industries.

TABLE I  
DESCRIPTIVE STATISTICS:  
ASI & NSS PLANTS

	Mean	Percentiles			Mean	Percentiles		
	(Std. Dev)	25	50	75	(Std. Dev)	25	50	75
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
<b>Panel A: ASI 2000-01 (Obs = 41,096; plants = 171,743)</b>					<b>ASI 2005-06 (Obs = 57,304; plants = 179,918)</b>			
Number of Employees	46.51 (382.07)	10	18	42	50.65 (347.60)	10	20	48
Gross Value Added per Worker (thousands of rupees)	191.87 (686.32)	27.19	63.20	128.48	286.75 (1104.25)	11.82	57.10	136.16
Number of Products per plant	1.53 (1.11)	1	1	2	1.53 (1.12)	1	1	2
<b>Panel B: NSS 2000-01 (Obs = 152,494; plants = 17,024,108)</b>					<b>NSS 2005-06 (Obs = 82330; plants = 16,953,555)</b>			
Number of Employees	2.17 (2.55)	1	2	2	2.11 (5.59)	1	2	2
Gross Value Added per Worker (thousands of rupees)	16.23 (17.09)	4.18	8.56	17.52	23.12 (47.64)	4.80	9.54	20.74
Number of Products per plant	1.04 (0.26)	1	1	1	1.05 (0.28)	1	1	1

Table I shows descriptive statistics of Indian plants for the fiscal year 2000-2001 and 2005-06 according to NSS and ASI. Panel A shows statistics of plants in the Annual Survey of Industries (ASI). Panel B shows statistics of the National Sample Survey (NSS).

### 3.1.4 Prices and consumption of intermediates

One of the characteristics that makes ASI and NSS rich datasets is the detailed information about production and intermediates use at the product level. For each plant in our data we observe all the goods produced and all the goods used as intermediate inputs.<sup>6</sup> The description of the “goods” is given by the Annual Survey of Industries Commodities Classification (ASICC). This classification was first created to have a clear description of the goods produced by ASI plants and was later taken as standard classification to define goods in the Indian manufacturing sector. Goods are very narrowly defined in ASICC. For instance, for food goods ASICC distinguishes between black tea, leaf, raw, black tea, leaf, blended, black tea, leaf, unblended, black tea, dust, blended, black tea, dust, unblended, etc. For instance, for processed minerals, ASICC distinguishes among around 12 different types of coal: coal for carbonisation, slack, soft, coke dust, etc.<sup>7</sup> But this is not all. For both the list of outputs and inputs we observe both the revenues/expenditures and physical quantities are reported. This means that we can actually compute the output prices charged by plants and input prices paid by plants.<sup>8</sup> To compute the price of outputs, we divide the ex-factory sales (factory gate sales) value and divide it by physical units. For the price of inputs, we divide the expenditure in a particular good by physical units.

## 4 Some facts

In this section we provide some evidence supporting the importance of market power in the manufacturing sector in India as well as its association with transportation infrastructure. In the first part of this section, we show that firms with higher sectoral shares seem to enjoy more market power and charge higher markups. We next show that, conditional on observables like factors costs and size, firms located in areas with a lower transportation infrastructure tend to charge higher prices. In the second part of this section, we study the relationship between transportation infrastructure and measures of sectoral concentration at the state level. We emphasize that we do not claim any causal relationship in the associations provided in this section. These facts will be very useful later on though, as we will use them to test the cross-state implications of our model.

---

<sup>6</sup>All plants report intermediate inputs imported from outside India separately from those not imported. This is crucial for our analysis, since we abstract from international trade in this paper.

<sup>7</sup>ASICC contains around 5,400 different classified products, around 80% and 83% of which are produced by plants in our sample in 2000-01 and 2005-06 respectively. 75% and 85% are used as intermediates for plants in our sample in 2000-01 and 2005-06 respectively.

<sup>8</sup>Although these datasets are becoming widely used, there has been not much attention to the price information. A notable exception is [Kothari \(2013\)](#).



## 4.1 The cross-state size distribution of plants

As we mentioned in the introduction, our hypothesis is that poor transportation infrastructure provides market power to local producers, allowing them to charge higher markups. In this context, plant size of these producers becomes inefficiently low, since their marginal productivity of labor is very high. Additionally, wages will be inefficiently low, implying an inefficiently low plant size of unproductive plants. We next study the cross-state association between connectivity and the share of employment accounted for by small plants. We use GIS analytics and NSS-ASI for the year 2005-06 to construct measures of connectivity and plants size distribution.

Figure I shows the cross-section relationship between the percentage of towns with a national highway within 1 mile and the unconditional share of employment allocated to plants with less than 5 workers. We first observe a wide cross-state variation on how resources are allocated. In states like Manipur or Mizoram almost all resources in the manufacturing sector are allocated to plants with less than 5 workers, which account for 89% and 86% of total employment. In contrast, in other states like Gujarat or Maharashtra around 75% of employment is allocated plants with more than 5 workers. We also observe a clear negative relationship between connectivity and the importance of small plants across states. This correlation has to be taken with caution though, since states with good infrastructure are different from states with poor infrastructure in many other dimensions. One of the main differences between well and not well connected states is the sectoral composition. States with better transportation infrastructure probably end up specializing in bigger scale industries. It is also be the case that better connected states are usually bigger, which influences the plants size distribution as well.

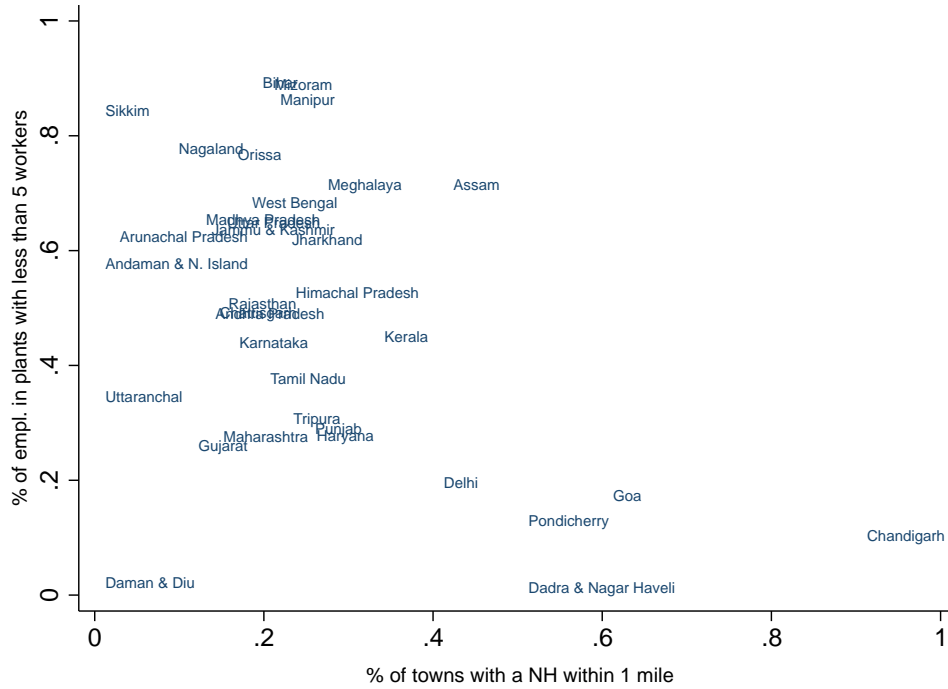
In order to provide a cleaner relationship between infrastructure and plants size distribution we run the following regression:

$$\text{Sh.EMP.}(<5 \text{ work.})_{ij} = \beta_0 + \beta_1 \text{Infrast}_i + \beta_2 \text{LogEMP}_i + \text{Industry}_j + u_{ij} \quad (1)$$

where  $\text{Sh.EMP.}(<5 \text{ work.})$  is the share of employment in plants with less than 5 workers in industry  $j$  in state  $i$ ,  $\text{Infrast}_i$  is our measure of infrastructure quality for a state  $i$ ,  $\text{LogEMP}_i$  is the log of total employment in state  $i$ , and  $\text{Industry}_j$  are industry dummies computed at the ASICC level. Table II shows the results. We find that, even controlling for the size of the state and the industry composition there is a strong negative association between the state infrastructure measured by the percentage of towns close to a national highway and the amount of resources allocated to small plants. By looking at column 3 for instance, we observe that within industry across states with similar size, in states in which the percentage of towns with a national highway within 10 mile is 50%, the share of employment accounted for by plants with less than 5 employees is around 3.5 percentage points lower. This association is much stronger when measuring infrastructure by looking at towns very near national highways. In column 4 we see that in states in which the percentage of towns with a national highway within only 1 mile is 50%, the share of employment

FIGURE I

RELATIONSHIP BETWEEN STATE CONNECTIVITY AND PLANTS SIZE DISTRIBUTION



accounted for by plants with less than 5 employees is around 7.5 percentage points lower.

TABLE II  
CROSS-STATE PLANTS SIZE DISTRIBUTION

	(1)	(2)	(3)	(4)
Dep. Variable: Share of employment accounted for by plants with less than 5 workers				
% of Towns with NH within 10 miles	-0.0376*** (0.00835)		-0.0618*** (0.00922)	
% of Towns with NH within 1 mile		-0.0762*** (0.0135)		-0.150*** (0.0167)
log Total Employment in the state			-0.0131*** (0.00178)	-0.0166*** (0.00194)
Product FE	YES	YES	YES	YES
Observations	24,254	24,254	24,254	24,254
R-squared	0.484	0.485	0.487	0.03

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Column (1) of table II shows state-product level regressions of the share of employment accounted for by plants with less than 5 workers against the percentage of towns in the state with a national highway within 10 miles and product fixed effects. Column (2) shows the same regression but including the percentage of towns in the district with a national highway within 1 mile instead. Columns (3) and (4) are the same regressions as columns (1) and (2) but including log employment of the state. Product fixed effects correspond to 5-digit Indian products codes (ASICC). The share of employment accounted for by plants with less than 5 workers have been computed using NSS-ASI for the year 2005-06. Robust standard errors are in parenthesis: \*: 10%; \*\*: 5%; \*\*\*: 1%.

## 4.2 Industry concentration among local producers

We next present some evidence relating states-industry concentration to transportation infrastructure. Under our hypothesis, the higher the connectivity of the state, the higher the competition from abroad faced by local firms. Imagine we depart from a situation in which a state is in autarky. For any given industry, the most productive firms will have high market shares and charge high markups, which will imply a high level of measured industry’s sales concentration. Now let us assume that the state opens to trade with other states. Then, industry’s concentration at home will decrease, since now all firms from abroad will start competing with the domestic ones, forcing the domestic firms that used to enjoy market power to decrease markups. Unfortunately, we can not test this prediction with our data directly. As mentioned in the introduction, we observe what intermediates are consumed by every single plant but the state of origin of the intermediate is not reported. It means we can not compute, for a given industry, the actual state concentration index, as we ignore what fraction of the consumption of goods comes from abroad. What we can test is whether more “opened” states have a higher industry concentration indexes within local firms. It is natural to think that when firms from all the states can compete in an industry, it will be the most productive ones from every state which account for most of the industry sales. We next look at the relationship between states’ level of transportation infrastructure and the Herfindahl index computed using only local plants. To this end, we run the following regression:

$$HH_{ij} = \alpha_0 + \alpha_1 \text{Infrast}_i + \alpha_2 \text{LogEmp}_i + \text{Industry}_j + \epsilon_{ij} \quad (2)$$

where:

$$HH_{ij} = \sum_{l=1}^{N_{ij}} \left( \frac{\text{Sales}_l}{\sum_{l=1}^{N_{ij}} \text{Sales}_l} \right)^2$$

is the Herfindahl index within local plants in industry  $j$  in country  $i$ ;  $N_{ij}$  is the number of firms in state  $i$  operating in industry  $j$  and  $\text{Sales}_l$  are the sales of a firm  $l$  in industry  $j$  in state  $i$ .  $\text{Infrast}_i$  is our measure of infrastructure quality for a state  $i$ ,  $\text{Log Emp}_i$  is the log of total employment in state  $i$ , and  $\text{Industry}_j$  are industry dummies computed at the ASICC level. We show the results of this regression in Table III. We find that industries in states with higher levels of transportation infrastructure have higher levels of concentration. For example, we find a coefficient of around 0.31 from the percentage of towns with a national highway within 1 mile in column 2. This implies, that industries located in states in which this percentage of town is for instance, 50%, are associated to a Herfindahl index within local plants 15 percentage points higher. This association is weaker when controlling for the size of the state. We see in column 4 that industries located in states in which the percentage of towns with a national highway within 1 mile is 50% are associated to a Herfindahl index within local plants around 5.5 percentage points higher.

TABLE III  
CROSS-STATE INDUSTRY CONCENTRATION

	(1)	(2)	(3)	(4)
<u>Dep. Variable: Herfindahl index</u>				
% of Towns with NH within 10 miles	0.107*** (0.0107)		0.018* (0.0111)	
% of Towns with NH within 1 mile		0.309*** (0.0171)		0.111*** (0.0192)
log Total Employment in the state			-0.048*** (0.00216)	-0.044*** (0.00226)
Product FE	YES	YES	YES	YES
Observations	24,260	24,260	24,260	24,260
R-squared	0.484	0.03	0.03	0.03

Robust standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Column (1) of table III shows state-product level regressions of the Herfindahl index against the percentage of towns in the state with a national highway within 10 miles and product fixed effects. Column (2) shows the same regression but including the percentage of towns in the district with a national highway within 1 mile instead. Columns (3) and (4) are the same regressions as columns (1) and (2) but including log employment of the state. Herfindahl indexes have been computed using sales. Product fixed effects correspond to 5-digit Indian products codes (ASICC). The Herfindahl index has been computed using sales at the product-plant level using NSS-ASI for the year 2005-06. Robust standard errors are in parenthesis: \*: 10%; \*\*: 5%; \*\*\*: 1%.

## 5 Transportation costs and prices: evidence from the Golden Quadrilateral

Although the preliminary evidence showed above is promising, we emphasize that these correlations have to be taken with caution. Reverse causality and omitted variables are key issues when analysing the relationship between infrastructure and economic development: a better transportation infrastructure may endogenously arise in response to a better economic performance. Moreover, broad policy strategies may drive both infrastructure spending and economic development. In order to deal with this identification problem, several authors have exploited the fact that the goal of infrastructure projects is usually to connect historical cities or large economic centers, as in, for instance, [Atack, Bateman, Haines, and Margo \(2010\)](#) and [Banerjee, Duflo, and Qian \(2012\)](#). The causal effect of transportation infrastructure is identified by applying a differences-in-differences approach: compare non-nodal areas close to the transportation network connecting large cities with those further away before and after the presence of the transportation network.

We follow this methodology to study the impact of a large-scale highway construction and improvement project in India, the Golden Quadrilateral (GQ), on price levels in Indian districts. The Golden Quadrilateral is a highway network connecting the four major cities in India: Delhi, Mumbai, Chennai, and Calcutta, with a total length of 5,846 km. The project was launched in

2001, as the first phase of the Indian National Highways Development Project (NHDP), the largest highway project ever undertaken by this country. The network was finished in 2011, but in 2006 95 per cent of the project was completed. Recent empirical works have investigated the impact of the Golden Quadrilateral project on economic outcomes. Using the 2002 and 2005 rounds of the World Bank Enterprise Surveys, [Datta \(2012\)](#) finds that firms in cities affected by the GQ reduced significantly their average stock of input inventories and reported decreased transportation obstacles to production. [Ghani, Goswami, and Kerr \(2013\)](#) use several rounds of the Annual Survey of Industries to document an increase in entry rates and plant productivity in districts located 0-10 km from the GQ. Compared to these papers, our focus is to test whether prices of intermediaries were reduced in those districts that became connected to the GQ relative to the districts further away from the network.

## 5.1 Data preparation

We use the 2000-2001 and 2005-2006 rounds of both the Annual Survey of Industries and the National Sample Survey in order to look at the distribution of prices of intermediaries before and after the construction of the GQ project.<sup>9</sup> For each round and district, we compute the price of each product as a weighted average of the prices paid by the plants using that product as intermediate in that district. Each price is calculated as the value of consumption of the input over the quantity consumed. We observe the price of 1,480 products that are consumed in the same district in both 2001 and 2006. The total number of districts is 506. Several districts in 2006 are carved out from districts in 2001. We use as benchmark the districts of the 2000-2001 round, merging those splits. Additionally, using GIS analytics we construct a dummy variable taking value 1 if the district falls within the GQ network.

## 5.2 Specification

We apply the usual differences-in-differences specification to study the impact of the GQ project on the prices paid in different districts. In particular, we run the following regression:

$$\Delta \log P_{jd} = \sum_j \alpha_j + \beta_1 \text{GQ}_d + \sum_s \delta_s + \epsilon_{jd} \quad (3)$$

where  $\Delta \log P_{jd}$  is the log change in the price of input  $j$  in district  $d$  between 2001 and 2006 and  $\text{GQ}_d$  is a dummy variable taking value 1 if district  $d$  is crossed by the Golden Quadrilateral, and zero otherwise. We include input fixed effects and state fixed effects in order to account for input-specific price trends and aggregate shocks affecting prices at the state level. Standard errors are

---

<sup>9</sup>Although 95 per cent of the GQ was finished by the end of 2006, it is natural to think that it requires some time to observe the total effect on economic outcomes. We are currently in the process of acquiring ASI and NSS for 2010-2011 in order to extend the time span of the analysis.

clustered at the district level in order to account for possible serial correlation of price shocks within districts.

**TABLE IV**  
**PRICES AND THE GOLDEN QUADRILATERAL: DIFFERENCES-IN-DIFFERENCES**

	(1)	(2)	(3)	(4)
<u>Dep. Variable:</u> Log change in input prices between 2001 and 2006				
Golden Quadrilateral	-0.3700*** (0.1146)	-0.2883** (0.1279)	-0.2791** (0.1236)	-0.2364* (0.1324)
Input fixed effects	YES	YES	YES	YES
State fixed effects	NO	YES	NO	YES
Nodal districts	YES	YES	NO	NO
Observations	8,906	8,906	7,682	7,682
R-squared	0.44	0.45	0.43	0.44

Table IV shows the estimation of equation (3). The dependent variable is the log change in input prices between 2001 and 2006 in the district. The variable of interest is the connectivity of the district, defined as whether the districts falls withing the Golden Quadrilateral network. Specifications across columns differ by the sample considered and the inclusion of state fixed effects. Columns (1) includes all districts. Column (2) adds state fixed effects. Column (3) excludes nodal districts and state fixed effects. Column (4) excludes nodal districts and controls for state fixed effects. Input fixed effects are included in all specifications. Robust standard errors are in parenthesis, clusterd at the district level. Significance levels: \*: 10%; \*\*: 5%; \*\*\*: 1%.

### 5.3 Results

The estimates of equation (3) can be found in Table IV. Column (1) shows that the change in input prices in those districts crossed by the GQ were 37 percentage points lower relative to those district further away. That is, prices fell more in those districts in which connectivity was improved. The coefficient is highly statistically significant and economically relevant. In column (2) we introduce state fixed effects in order to control for aggregate shocks affecting both input prices and connectivity at the state level. The point estimates drops to 29 percentage points, remaining statistically significant at a 95 per cent confidence level. Column (3) excludes the nodal districts of the GQ network. The GQ was aimed to improve the land connection between these districts. Therefore, their inclusion in the network might be the result of a different economic performance by those districts, or by other omitted factors. This would generate biased estimates, possibly upward. Hence, following Datta (2012) and Ghani, Goswami, and Kerr (2013) we analyze the robustness of the results by restricting the sample to non-nodal areas. Following these papers, we exclude the four nodal districts (Mumbai, Calcutta, Chennai, and Delhi) as well as contiguous suburbs (Thane for Mumbai and Ghaziabad, Gurgaon, Gautam Buddh Nagar and Faribadad for Delhi). Columns (3) and (4) show that the results prevail in this restricted sample. Input prices

in non-nodal districts crossed by the GQ fell relative to areas not crossed by the highway network. The elasticities are around 20 per cent lower than those found in the full sample, but the statistical significance is preserved. Overall, we find robust evidence that the improvement in connectivity lead to a decrease in input prices.

## 6 Model

In this section we present our static general equilibrium model of internal trade. In each of the  $N$  states there is a measure 1 of industries. Within each industry, there are a firms that compete in an oligopolistic manner. Labor is immobile across states.

### 6.1 Consumers

In each state  $i$  there is a representative household with utility function

$$C_i = \left( \int_0^1 C_i(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} \quad (4)$$

where  $C_i(j)$  is the composite good for industry  $j$  and  $\theta > 1$  is the elasticity of substitution across composite goods of different industries. The industry-level composite good is defined as

$$C_i(j) = \left( \sum_{o=1}^N \sum_{k=1}^{K_o} c_i^o(j, k)^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}} \quad (5)$$

where  $c_i^o(j, k)$  is the good consumed by state  $i$  and provided by firm  $k$  in industry  $j$  from state  $o$ ,  $N$  is the number of states,  $K_o$  is the number of firms that operate in each industry in state  $o$ , and  $\gamma > 1$  is the elasticity of substitution of goods produced by different firms in the same industry. We consider the cases that  $\gamma > \theta$ , which means that goods are more substitutable *within* industries than *between* industries.

The budget constraint of the representative household in state  $i$  is given by:

$$\int_0^1 \left( \sum_{o=1}^N \sum_{k=1}^{K_o} p_i^o(j, k) c_i^o(j, k) \right) dj = W_i L_i + \Pi_i \quad (6)$$

where  $W_i$  is the equilibrium wage,  $L_i$  is labor endowment, and  $\Pi_i$  is income derived from profits of firms located in  $i$ .

### 6.2 Firms

In each industry in state  $o$  there are  $K_o$  firms. Firms draw their productivities from a pareto distribution  $G(a) = 1 - a^{-\kappa}$ , where  $a \geq 1$ . A firm with productivity level  $a$  has a labor requirement of  $1/a$  to produce one unit of good. Because firms do not pay any fixed cost to operate in a market, firms will operate in all  $N$  states.

To find the pricing rule of the firm, we first find the demand faced by that firm. Equations 4, 5, and 6 generate demand

$$c_i^o(j) = \left( \frac{P_i}{P_i(j)} \right)^\theta \left( \frac{P_i(j)}{p_i^o(j, k)} \right)^\gamma Y_i, \quad (7)$$

where

$$P_i = \left( \int_0^1 P_i(j)^{1-\theta} dj \right)^{\frac{1}{1-\theta}}, \quad (8)$$

$$P_i(j) = \left( \sum_{o=1}^N \sum_{k=1}^{K_o} p_i^o(j, k)^{1-\gamma} \right)^{\frac{1}{1-\gamma}}. \quad (9)$$

Firms within industries firms compete a la Cournot. Firm  $j$  takes as given the demand characterized in equation 7 and quantity supplied by competitor firms in the industry and solves the following problem

$$\pi_d^o(j, k) = \max_{c_d^o(j, k)} p_d^o(j, k) c_d^o(j, k) - \frac{W_o \tau_d^o}{a_o(j, k)} c_d^o(j, k), \quad (10)$$

where  $a_{okj}$  is the productivity of firm  $j$  in industry  $k$  in country  $o$ ,  $\tau_d^o$  is the iceberg transportation cost to send a unit of good to  $d$ . The solution of this problem is

$$p_d^o(j, k) = \frac{\epsilon_d^o(j, k)}{\epsilon_d^o(j, k) - 1} \frac{W_o}{a_o(j, k)} \tau_d^o, \quad (11)$$

where

$$\epsilon_d^o(k, j) = \left( \omega_d^o(j, k) \frac{1}{\theta} + (1 - \omega_d^o(j, k)) \frac{1}{\gamma} \right)^{-1}, \quad (12)$$

and  $\omega_d^o(k, j)$  is the market share of firm  $j$  in industry  $k$  in market  $d$

$$\omega_d^o = \frac{p_d^o(j, k) c_d^o(j, k)}{\sum_{o=1}^N \sum_{k=1}^K p_d^o(j, k) c_d^o(j, k)}. \quad (13)$$

Aggregate profits of firms in country  $i$  are characterized by

$$\Pi_i = \int_0^1 \left( \sum_{i=1}^N \sum_{k=1}^K \pi_i^o(j, k) \right) dj. \quad (14)$$

### 6.3 Balanced trade and labor clearing condition

All countries  $i$  must have balanced trade

$$\int_0^1 \left( \sum_{o=1, o \neq i}^N \sum_{k=1}^K p_i^o(j, k) c_i^o(j, k) \right) dj = \int_0^1 \left( \sum_{d=1, d \neq i}^N \sum_{k=1}^K p_d^i(j, k) c_d^i(j, k) \right) dj. \quad (15)$$

The labor clearing condition for country  $i$  is

$$\int_0^1 \left( \sum_{d=1}^N \sum_{k=1}^K \frac{c_d^i(j, k)}{a_i(j, k)} \tau_d^o \right) dj = L_i. \quad (16)$$



## 6.4 Definition of equilibrium

*Equilibrium.* For all countries  $i$  and  $i'$ , industries  $j$ , and firms  $k$ , an equilibrium is a set of allocations of consumption goods  $\{c_{i'}^i(j, k), C_i(j)\}$ , firm prices  $\{p_{i'}^i(j, k)\}$ , industry prices  $\{P_i(j)\}$ , and aggregate variables  $\{W_i, P_i, \Pi_i\}$  such that

1. Given firm prices, industry prices, aggregate variables,  $\{c_i^o(j)\}$  is given by 7 and solves the consumer's problem in 4, 5, and 6.
2. Given aggregate variables,  $p_d^o(j, k)$  is given by 11, 12, and 13, and solves the problem of the firm in 10.
3. Aggregate profits satisfy 14, aggregate prices satisfy 8, and industry prices satisfy 9.
4. Trade flows satisfy 15.
5. Labor markets satisfy 16.

## 6.5 Misallocation in the model

We now want to study the misallocation that arises in the model due to poor transportation infrastructure in India. Because labor is only one factor of production, papers such as Holmes, Hsu, and Lee (2012) have shown that misallocation arises only due to *dispersion* in markup across producers of the same country. Consumers will consume too little of the goods with high marks and too much of the goods with low markups. It is important to note that the level of markups do not have any bearing on misallocation. It can be shown that as long as all producers from the same country have the same markups across markets, then the equilibrium allocation is the first-best.

The misallocation that arises in the model can be linked to firm-size distribution. More productive firms charge higher markups since they can capture a larger portion of the industry's marketshare. This means that firms with high productivity draws are smaller than they would be in the first best. In other words, the economy would be better off by reallocating labor from firms with low productivity draws (low markup firms) to firms with high productivity draws (high markup firms). Misallocation in the model implies that large firms are not large enough and that small firms are too large.

In previous works, the cross-firm variation of marginal product of labor has been interpreted as firms facing idiosyncratic distortions as a consequence of government policies.<sup>10</sup> Through the lens of our model, this misallocation is interpreted as firms enjoying different market powers due to the lack of perfect competition.<sup>11</sup> This is because the variation in markups is proportional to

---

<sup>10</sup>Restuccia and Rogerson (2008); Hsieh and Klenow (2009); Guner, Ventura, and Yi (2008) are some examples.

<sup>11</sup>We are not the first in interpreting the variation of marginal productivity of labor in the data in this way. See for instance Peters (2013), who studies the link between misallocation and growth.

the variation in marginal product of labor. The constant returns to scale technology means that the marginal product of labor is  $w \frac{\epsilon(j,k)}{\epsilon(j,k)-1}$ , where  $w$  is the wage at the origin. Thus, firms charging high markups are precisely the ones with high levels of marginal product of labor.

**First best level of TFP** We now want to derive an expression for the first best level of TFP in the economy, which is achieved when there is no misallocation. First best TFP is

$$\tilde{A}_i = \frac{\tilde{Y}_i^*}{L_i},$$

where  $\tilde{Y}_i^*$  is first best aggregate output. We know that this allocation can be implemented as a competitive equilibrium if all markets are perfectly competitive. Thus, we get that

$$\tilde{A}_i = \frac{L_i \tilde{W}_i}{L_i} \tilde{P}_i^{-1} = \tilde{W}_i \tilde{P}_i^{-1}, \quad (17)$$

where  $\tilde{W}_i$  and  $\tilde{P}_i$  are equilibrium wages and aggregate prices under perfectly competitive goods markets. This implies that the first best level of TFP in a given country would be equal to real income under perfect competition.

Under perfectly competitive goods market, all firms price at marginal cost. If we substitute these prices into equation 17 along with equations 8 and 9, we get that the first best level of TFP is

$$\tilde{A}_i = \tilde{W}_i \left( \int_0^1 \tilde{A}_i(j)^{\theta-1} dj \right)^{\frac{1}{\theta-1}},$$

where

$$\tilde{A}_i(j) = \left( \sum_{o=1}^N (\tilde{W}_o \tau_i^o)^{1-\gamma} \sum_{k=1}^{K_o} (a_o(j,k))^{\gamma-1} \right)^{\frac{1}{\gamma-1}}.$$

## 7 Inferring parameter values

The goal of this section is to use our data to infer parameter value for the model. We also show that removing misallocation in the model would yield substantial gains in Indian manufacturing.

The first step of the process is to use pricing data to infer transportation costs between states in India. We estimate these transportation costs based on distance and infrastructure quality between the origin and destination states. We show that the quality of infrastructure can matter just as much as distance in determining transportation costs. The next step is to derive a gravity equation for trade flows from the model and use it to estimate the across-industry elasticity of substitution,  $\theta$ . We will explain how we calibrate the remaining parameters in section 7.4.

## 7.1 Firms with near monopolies

The first step is to use pricing data to find transportation costs between states of India. This is a challenge because the prices charged by firms depend both on transportation and the level of competition in the destination market.<sup>12</sup>

To solve this problem, we exploit an implication of the model: for monopolists, variation on prices across destinations is due to variation in transportation costs. Equation 13 shows that when the sectoral share of a firm equals one, the demand elasticity that the firm faces becomes  $\theta$ . Hence, the price charged by the firm, which operates in location  $o$ , in destination  $d$  becomes:

$$p_d^o(j, k) = \frac{\theta}{\theta - 1} \frac{W_o}{a_o(j, k)} \tau_d^o \quad (18)$$

By dividing the expressions of the price charged by a monopolist in different destinations

$$\frac{p_d^o(j, k)}{p_{d'}^o(j, k)} = \frac{\tau_d^o}{\tau_{d'}^o}. \quad (19)$$

Differences in prices for monopolist producers across destinations reveal the ratio of transportation costs across those destinations.

The first step is to identify plants that we call near monopolists, a plant that produces more than 95% of the value of output nationwide at the ASICC 5 digit level (the highest level of disaggregation in the dataset).<sup>13</sup>

We identify 140 products that are manufactured by near monopolists. Table V shows the distribution of these products across industries. It is comforting that the largest category is “Manufacture of chemicals and chemical products,” which contains around 40% of the identified products. This is consistent with the production structure of the chemical industry. The production of a certain chemical is often concentrated in one plant and that production is shipped to many locations. This is a result of the economies of scale in production. One important subset of such chemicals are commodity chemicals, which consists of over 60% of the value of chemical production in India.

## 7.2 Inferring transportation costs in India

Once we have identified near monopolists in our data, we want to estimate transportation costs between the states of India. We estimate equation 18 as follows

$$\log P_{jod} = \text{Origin}_o + \beta_1 \log \text{Distance}_{od} + \beta_2 \text{Inf}_d + \text{Product}_j + u_{jod} \quad (20)$$

where  $P_{jod}$  is the average price paid in district  $d$  for product  $j$  produced in district  $o$ ,  $\text{Origin}_o$  is origin fixed effect which aims to control for prices paid at origin  $o$ ,  $\text{Distance}_{od}$  is the distance in

<sup>12</sup>A firm in the model maps to a plant in the data. In developing countries 85% of firms have only one plant.

<sup>13</sup>We also exclude goods that are not used in at least 5 other districts.

TABLE V  
INDUSTRIES WITH NEAR MONOPOLISTS

Sector code (NIC)	Sector name	Number of plants
15	Manufacture of food products and beverages	10
16	Manufacture of tobacco products	1
17	Manufacture of textiles	13
19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear	3
21	Manufacture of paper and paper products	4
22	Publishing, printing and reproduction of recorded media	1
23	Manufacture of coke, refined petroleum products and nuclear fuel	6
24	Manufacture of chemicals and chemical products	54
25	Manufacture of rubber and plastics products	4
26	Manufacture of other non-metallic mineral products	10
27	Manufacture of basic metals	13
28	Manufacture of fabricated metal products, except machinery and equipment	1
30	Manufacture of office, accounting and computing machinery	2
31	Manufacture of electrical machinery and apparatus n.e.c.	7
32	Manufacture of radio, television and communication equipment and apparatus	5
33	Manufacture of medical, precision and optical instruments, watches and clocks	2
34	Manufacture of motor vehicles, trailers and semi-trailers	1
35	Manufacture of other transport equipment	1
36	Manufacture of furniture; manufacturing n.e.c.	2

Table V shows the distribution of near monopolists across different sectors defined by the National Industry Classification at the 2-digit level (NIC2). We define a plant as near monopolist when it accounts for by more than 95% of total sales of a product in India.

miles between district  $o$  and district  $d$ ,  $\text{Inf}_d$  is the percentage of towns in district  $d$  with a national highway within 5 miles,  $\text{Product}_j$  is product dummy that aims to control for different levels of transportation costs across products, and  $u_{jod}$  is the error term.<sup>14</sup>

Table VI presents the results from running an OLS regression of equation 20. Column (1) shows the results of the regression when including only distance, whereas column (2) also includes transportation infrastructure. We find that increases in distance is associated with increasing transportation costs and that the coefficient is not highly sensitive to including infrastructure quality. Column (2) indicates that a 10% increase in distance will be associated with an increase of 1.3% in transportation costs. It is comforting that this coefficient is consistent with studies. We also find that infrastructure matters can have significant impacts on transportation costs. Column (2) indicates that a district with the best possible infrastructure quality will have transportation costs that are 33% lower than a district with the worst infrastructure quality.

---

<sup>14</sup>Although we observe prices paid by firms that consume the intermediate good at origin, we take as our benchmark the specification with origin dummies because the number of observations becomes twice bigger. The estimated coefficients are the same magnitude under the two specifications though.

TABLE VI  
ESTIMATING TRANSPORTATION COSTS

	(1)	(2)
Dep. Variable: Log average Price at district of destination		
log Distance	0.128** (0.0542)	0.127** (0.0571)
% towns with NH within 5 miles		-0.326** (0.159)
District of origin FE	YES	YES
Product FE	YES	YES
Observations	1,508	1,411
R-squared	0.789	0.790
Pseudo R-squared	0.768	0.767

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Column (1) of table VI shows district of origin-district of destination-product regressions of log average price paid at destination against log miles, district of origin fixed effects and, product fixed effects. Column (2) shows the same regression but including the percentage of towns in the district with a national highway within 5 miles. Product fixed effects correspond to 5-digit Indian products codes (ASICC). Robust standard errors are in parenthesis: \*: 10%; \*\*: 5%; \*\*\*: 1%.

**How important is infrastructure quality?** We want to get a sense of the relative importance of distance and infrastructure quality in determining transportation costs. We take Delhi, which is located in the northern part of the country, as our point of origin. We then rank all other districts in India based on distance from Delhi, and find the district that is in the 10th percentile and the 90th percentile in terms of distance. We will call these districts as being close and far from Delhi respectively. The close district is 31 miles away while the far district is 1397 miles away.

Next, we rank all the districts in terms of infrastructure quality and find the district with the infrastructure quality in the 10th and 90th percentile. These measures indicate infrastructure quality that is bad and good. The district with bad infrastructure quality has no towns within 25 miles of a national highway, while a district with good infrastructure has all towns within 25 miles of a national highway.

Now we want to illustrate the importance of infrastructure quality relative to distance. We compare the transportation costs for a close district with bad infrastructure to that of a far district with good infrastructure for a fixed product. From equation 20, we can derive the following expression

$$\frac{P_{jod}}{P_{jod'}} = \left( \frac{\text{Distance}_{od}}{\text{Distance}_{od'}} \right)^{\beta_1} \left( \frac{e^{\text{Inf}_d}}{e^{\text{Inf}_{d'}}} \right)^{\beta_2}$$

Substituting in the coefficients from the regression we get that transportation costs to the close district are 8% higher than the far district.

We have recently acquired detailed geospatial data that will allow us to take into account the road quality between two districts. The final result from that analysis is that we will determine the relative cost to travel on roads of different quality in India (for example, it is twice as costly to

travel on a one-lane road than a two-lane road). These differences in the relative cost of traveling on different types of roads and the road structure between districts will determine transportation costs.

### 7.3 Estimating the across-industry elasticity of substitution $\theta$

In order to estimate the across-industry elasticity of substitution, we first derive the gravity equation for trade flows implied by the model for a monopolist firm. Combining 7 and 18 we get

$$c_d^o(j, k)p_d^o(j, k) = \left( \frac{W_o \tau_d^o}{a_o(j, k)} \frac{\theta - 1}{\theta} \right)^{1-\theta} P_d^\theta Y_d. \quad (21)$$

Take logs of both sides of the equation:

$$\begin{aligned} \log(c_d^o(j, k)p_d^o(j, k)) &= (1 - \theta) \log(W_o) + (1 - \theta) \log(\tau_d^o) + (\theta - 1) \log(a_o(j, k)) \\ &+ \log(P_d^\theta Y_d) + (1 - \theta) \log \frac{\theta - 1}{\theta}. \end{aligned} \quad (22)$$

The model predicts that transportation costs between locations reduce trade flows, which will be how we identify this parameter. The intuition is that the strength of the relationship between transportation costs and bilateral flows depends on the value of  $\theta$ . The intuition behind the identification strategy is that if small differences in transportation costs among destinations are associated with big differences in trade flows, then the value of  $\theta$  must be high (and vice-versa). It is also important to note that this relationship only holds when firms are monopolists.

To estimate equation 22 we run the following regression:

$$\log(c_{jod}p_{jod}) = \text{Origin}_o + \text{Dest}_d + (1 - \theta) \log \hat{\tau}_d^o + \text{Product}_j + \epsilon_{jod} \quad (23)$$

where  $c_{jod}p_{jod}$  is the value of sales of product  $j$  consumed in district  $d$  and produced by a monopolist in district  $o$ ,  $\hat{\tau}_d^o$  is the predicted iceberg transportation cost between district  $o$  and  $d$  from before, and  $\epsilon_{jod}$  is the error term.

Table VII presents the results from running an OLS regression of equation 23. We find that higher transportation costs between origin and destination are associated with lower amounts of trade between them. The empirical specification indicates that transportation costs increasing by 10% is associated with 17% lower trade flows. This relationship implies that the value of  $\theta$  is 2.84.

### 7.4 Determining the level of transportation costs

By estimating the regression in equation 20, we can determine how transportation costs are related to changes in distance in infrastructure for a given good and origin. However, this does not indicate the level of transportation costs. In order to determine the level of transportation costs, for each origin district we find the destination district with the lowest transportation cost. We set the iceberg

TABLE VII  
GRAVITY EQUATIONS FOR MONOPOLISTS

	(1)	(2)
Dep. Variable: Log Value of Sales at destination		
$\hat{\tau}_d^o$	-1.731*** (0.500)	-1.838** (0.759)
log Size <sub>d</sub>	0.227*** (0.059)	
District of origin FE	YES	YES
District of dest. FE	NO	YES
Product FE	YES	YES
Observations	1,580	1,580
R-squared	0.295	0.509

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Column (1) of table VII shows district bilateral flows regressions of log Value of Sales at destination against transportation costs between origin and destination, size of district of destination measured by the log of total employment, district of origin fixed effects, and product fixed effects. Column (2) shows the same regression but district of origin fixed effects to control for the size of the district of destination. Product fixed effects correspond to 5-digit Indian products codes (ASICC). Robust standard errors are in parenthesis: \*: 10%; \*\*: 5%; \*\*\*: 1%. When computing the standard errors of the coefficient associated to  $\hat{\tau}_d^o$  correcting for the variance of the estimator in the regression 20, the coefficient remains significant at 5%.

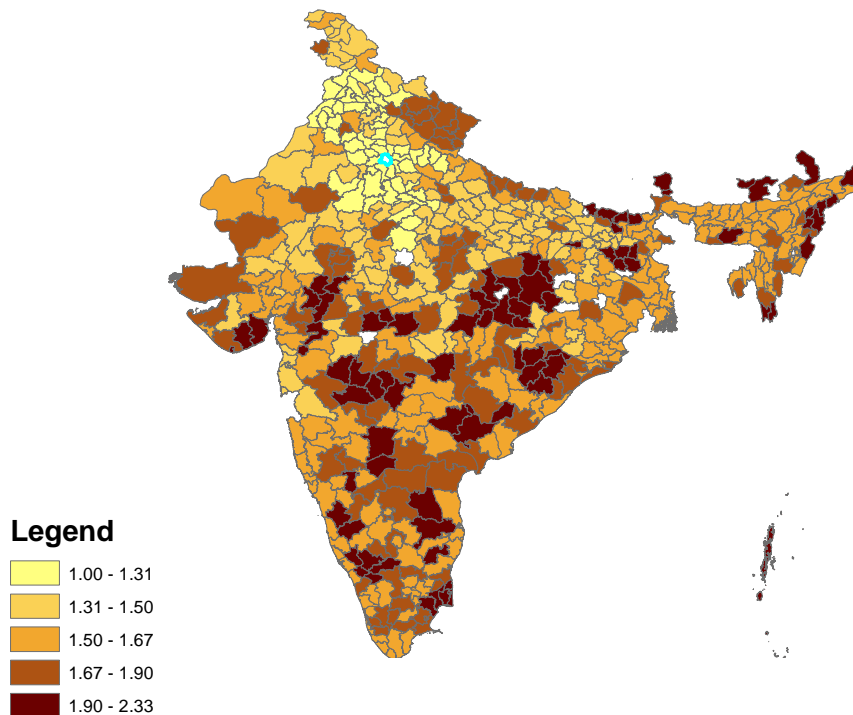
transportation cost to 1.01 for that district. We then use the regression results from equation 20 to determine transportation costs for the remaining districts.

**What do transportation costs look like?** As a starting point we will take the districts located in the region of Delhi. Figure ?? shows the histogram of the transportation costs to all the districts in India. The average district has iceberg transportation cost of 1.61 and the standard deviation is 0.25. Next, we want to look at the geographic relationship of these transportation costs. Figure II shows a map of the transportation costs to all possible destination districts. First, we see that distance does play a role in determining the transportation costs: transportation costs tend to be lower in closer districts. However, we see many instances in which transportation costs are less in districts that are very far away. Second, we see that there is clustering of areas with poor infrastructure, which contributes to high transportation costs.

**Aggregating transportation costs to the state-level** The model that we simulate will be based on the Indian states trading with each other. Thus, it is necessary to aggregate the district-to-district transportation costs to state-to-state transportation costs. In order to do so, for any two states we take the average bilateral transportation costs between all the districts in those states. This gives us an average measure of transportation costs between the two states.

Given these new set of transportation costs, we repeat the exercise from above in which we map the transportation costs from Delhi to all of the states in India. Figure III shows a map of these

**FIGURE II**  
 MAP OF TRANSPORTATION COSTS FOR DISTRICTS LOCATED IN REGION OF DELHI  
 DISTRICT-LEVEL  
 UNDER TRANSPORTATION INFRASTRUCTURE IN 2001



transportation costs.

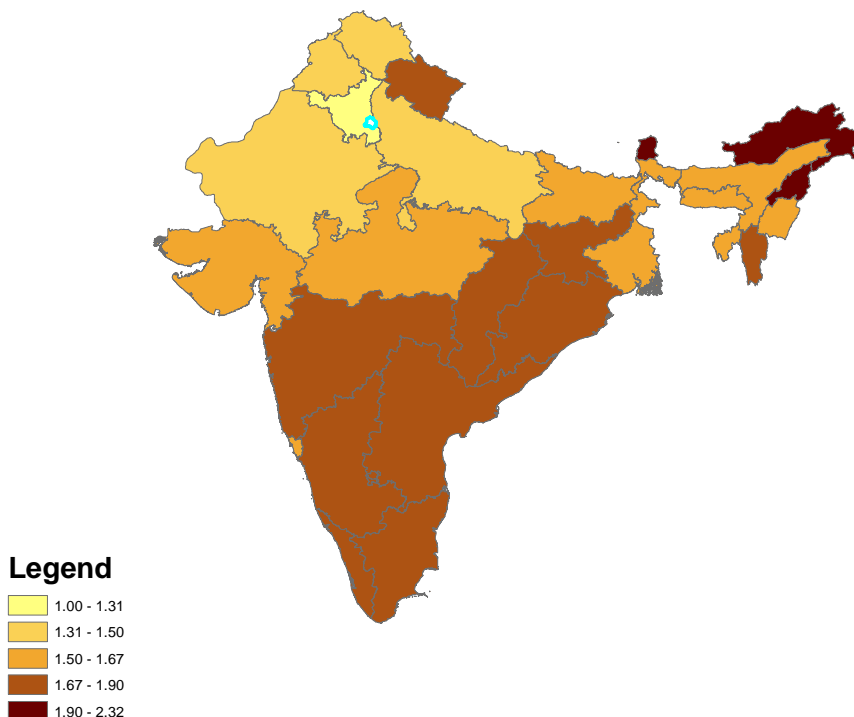
### 7.5 Calibrating the remaining parameters

Once we have values for transportation costs between states and the across-sector elasticity of substitution, we need to give values to the labor endowment of states  $L_i$ , the number of firms operating in each industry across states  $K_i$ , the parameters governing the firms productivity distribution, and the elasticity of substitution within sector  $\gamma$ . At this preliminary stage we assume that the productivity distribution follows a Pareto with shape parameter equal to 2. We will calibrate the remaining parameters so that in equilibrium the model matches relevant statistics in the data.

**Labor endowment and number of firms across states** For labor endowments of each state,  $L_i$ , we first normalize the labor endowment of the state with the lowest manufacturing value added to 1. We then set the labor endowments of the remaining states to match the ratio of manufacturing value-added (relative to that state that we normalized to) that we observe in the data. The number of firms in each sector in state  $i$ ,  $K_i$ , is set to match the cross-industry median number of producers



FIGURE III  
MAP OF TRANSPORTATION COSTS FOR DISTRICTS LOCATED IN REGION OF DELHI  
STATE-LEVEL  
UNDER TRANSPORTATION INFRASTRUCTURE IN 2001



in each state. The model perfectly matches these two dimensions of the data for all the states. Table IX shows the specific targets for every state.

**The productivity distribution of firms** As a preliminary attempt, we use a Pareto distribution with shape parameter equal to 2. We assume that all firms operating in India make draws from the same distribution, independently of the state where the firm operates.

**The pattern of comparative advantage across states** Edmond, Midrigan, and Xu (2012) show that the “cross-trade partners” pattern of correlation in productivity draws is crucial to determine the size of pro-competitive gains from trade, since it determines how strong the pattern of comparative advantage is, and hence to what extent local firms with market power face new competition when the economy opens to trade.<sup>15</sup> Following Edmond, Midrigan, and Xu (2012) we will calibrate two different economies. In the first economy draws across states will be completely

<sup>15</sup>In a situation in which productivity draws across states are independent, the pattern of comparative advantage is very strong, and the pro-competitive gains from trade are zero or even negative. In the model developed by Eaton, Kortum, Bernard, and Jensen (2003) the implied pro-competitive gains are exactly zero. In a more recent paper Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012) show that pro-competitive gains can be even negative.

TABLE VIII  
PARAMETER VALUES

Param.	Definition	Value
(A) PARAMETERS ESTIMATED WITH STRUCTURAL EQUATIONS		
$\tau_d^o$	Iceberg transportation costs between states	varies by state pair
$\theta$	Elasticity of substitution across sectors	2.84
(B) PARAMETERS CALIBRATED IN EQUILIBRIUM		
$L_i$	Labor endowment of the states	varies by state
$K_i$	Number of firms operating in the median sector	varies by state
$\gamma$	Elasticity of substitution within sector	34.03
(B) PREDETERMINED PARAMETERS		
$\alpha$	Shape parameter Pareto	2.00

Notes: Table VIII refers to a calibration in which productivity draws across states are independent.

uncorrelated, while in the second economy the productivity draws will be perfectly correlated across states.

**The elasticity of substitution within sector** Once we have values for the across-sector elasticity of substitution, the bilateral iceberg costs, and the parameter governing the productivity distribution, we need to pick a value for the elasticity of substitution within sector  $\gamma$ . Note that conditional on transportation costs and the productivity distribution, the amount of misallocation in the model will depend on the gap between the two elasticities  $\gamma$  and  $\theta$ . We already have a value for  $\theta$  from our gravity equation estimation. We then set  $\gamma$  to match the ratio of the (value added) weighted average labor share relative to the unweighted average labor share. The identification of  $\gamma$  comes from the fact that, in the model, the higher the gap between the elasticity across sectors and the elasticity within sector, the higher the difference between the labor share of firms with high sectoral shares and firms with low sectoral shares. Then, for a given value of  $\theta$ , high values of  $\gamma$  will make the model to generate high differences between the unweighted and the weighted labor shares. The difference between these two averages is actually high in the data. According to our merged NSS-ASI data for 2000-01, the unweighted average labor share is equal to 0.69, while the value-added weighted average labor share is 0.54. This implies a ratio of the weighted to the unweighted average labor share equal to 0.78. This gap between the unweighted and weighted average comes from a strong cross-firm negative relationship between labor shares and sectoral shares. For instance, when we running a simple regression of firm labor share against sectoral share and industry dummies, OLS coefficients imply that plants with a sectoral share close to zero

TABLE IX  
CROSS-STATE TARGETS FOR CALIBRATION

State	Median Number of Producers	Relative Man. Value Added
Andhra Pradesh	5	692.3
Arunachal Pradesh	10	1.0
Assam	16	117.8
Bihar	21	91.1
Chattisgarh	2	263.2
Delhi	4	163.6
Goa	1	119.4
Gujarat	5	1987.9
Haryana	4	591.1
Himachal Pradesh	2	290.3
Jammu&Kashmir	3	89.1
Jharkhand	4	469.6
Karnataka	5	850.2
Kerala	11	257.9
Madhya Pradesh	4	361.5
Maharashtra	6	3101.2
Manipur	53	5.3
Meghalaya	26	18.6
Mizoram	13	3.7
Nagaland	13	3.7
Orissa	5	322.7
Punjab	5	344.9
Rajasthan	5	417.0
Sikkim	15	1.5
Tamil Nadu	5	1279.4
Tripura	12	10.1
Uttar Pradesh	6	955.5
Uttaranchal	2	151.8
West Bengal	5	611.3

Table IX Column (1) shows the state median number of producers across industries, where industries are defined according to ASICCC product classification. Column (2) shows the state relative value added of the manufacturing sector.

are associated to a labor share of around 0.43. In contrast, plants with sectoral shares of 0.50 are associated to a labor share of around 0.21.<sup>16</sup> Our calibrated parameter  $\gamma$  is equal to 34 (see table VIII). This number is very high compared to the one estimated by Edmond, Midrigan, and Xu (2012). The reason is that, conditional on the productivity distribution we assume, and our estimated transportation costs and across sector elasticity of substitution  $\theta$  the model needs a very high value of  $\gamma$  to match to strong negative cross-section relationship between labor share and sectoral share we observe in the Indian economy.

## 7.6 How important is misallocation?

Armed with our calibration we can actually compare the allocations that emerge in equilibrium and compare them with a situation in which there is not misallocation at all (the first best allocation).

<sup>16</sup>Edmond, Midrigan, and Xu (2012) find a lower ratio of the unweighted with respect to the weighted average for the case of Taiwan: 0.70. However, they find very similar point estimates for the relationship between labor and sectoral shares.

In our model, there is not distinction between TFP and income per capita, since labor is the only factor of production and there is not endogenous entry. We first find that the initial levels of misallocation are similar under the two different patterns on comparative advantage. In the case of perfect correlated draws, the cross-states unweighted average TFP loss across states is 5.77 per cent. In the case of uncorrelated draws it is a little bit lower, 5.64 per cent. interestingly, there is substantial variation across states. For instance, in the case of uncorrelated draws, misallocation in states like Tamil Nadu or Gujarat implies around 15% of TFP losses, while in in states like Kerala and Tripura the amount of misallocation is lower, generating only around 3% TFP losses. The model generates higher amounts of misallocation in smaller states, which implies that the aggregate level of misallocation in India in the case of perfect correlated draws is 5.23, slightly lower than the cross-state unweighted average. Overall, the levels of misallocation that the model generates under our preliminary calibration are very low compare to, for instance, the ones found by Hsieh and Klenow (2009).

## 8 Simulating a decline in transportation costs

Now we describe our quantitative results when simulating the economy with lower transportation costs. We run the following exercise: we calculate a new bilateral matrix of transportation costs assuming that all districts have perfect transportation infrastructure. This is, we compute the predicted transportation costs using equation 20 and imposing that the percentage of towns with a highway within 25 miles equal to one. This corresponds to the districts in the top 15% in terms of infrastructure quality.

Figure IV and V show this new counterfactual distribution of transportation costs from Delhi at the district and state level respectively. Under perfect transportation infrastructure, the pattern of transportation costs become only dependent on distance, which is reflected in the concentric circles each with increasing transportation costs.

When declining transportation costs, two different types of income gain arise.<sup>17</sup> Under lower transportation costs the trading goods becomes cheaper and hence the first best level of TFP increases.<sup>18</sup> Following the recent literature we call these gains *Ricardian gains*. However, decreasing transportation costs has also an effect on the allocation of resources. By increasing competition between firms, the distribution of markups in the economy becomes less dispersed and hence the level of misallocation decreases. We call these gains *pro-competitive gains*.

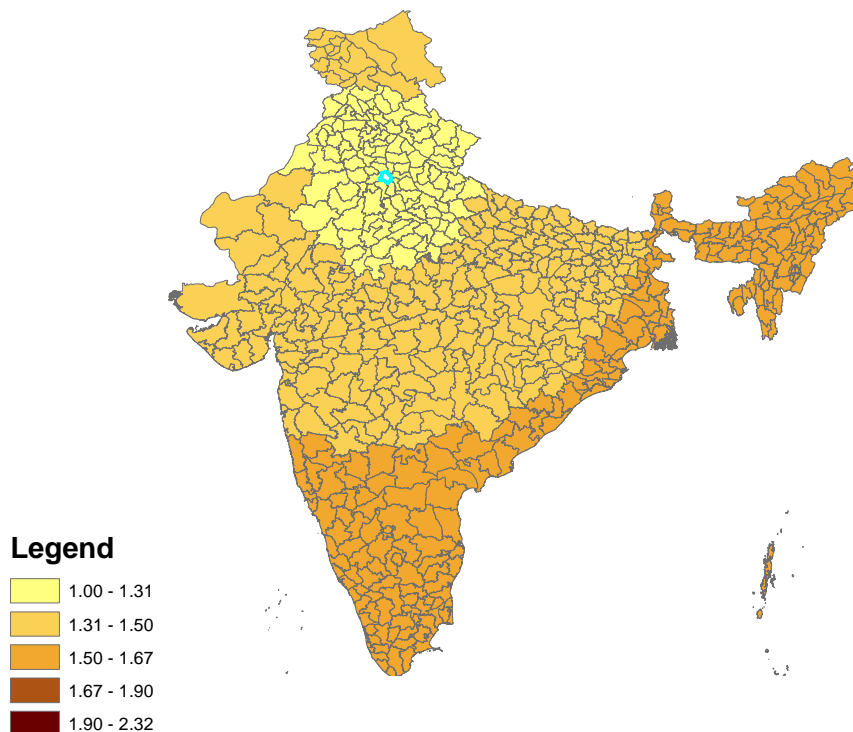
We carry out the counterfactual exercise for two benchmark economies in which we use uncorrelated and perfectly correlated draws across states. Panel (A) of Table X shows the results in the case of uncorrelated draws while panel (B) shows the results for the case of perfectly correlated draws. For these two cases, we report the results for the states of Sikkim, Uttar Pradesh, and

---

<sup>17</sup>Note that in our model income per capita and TFP are the same objects, since there is not capital or endogenous entry in the economy.

<sup>18</sup>See equation 17.

**FIGURE IV**  
 MAP OF TRANSPORTATION COSTS FOR DISTRICTS LOCATED IN REGION OF DELHI  
 DISTRICT-LEVEL  
 UNDER PERFECT TRANSPORTATION INFRASTRUCTURE

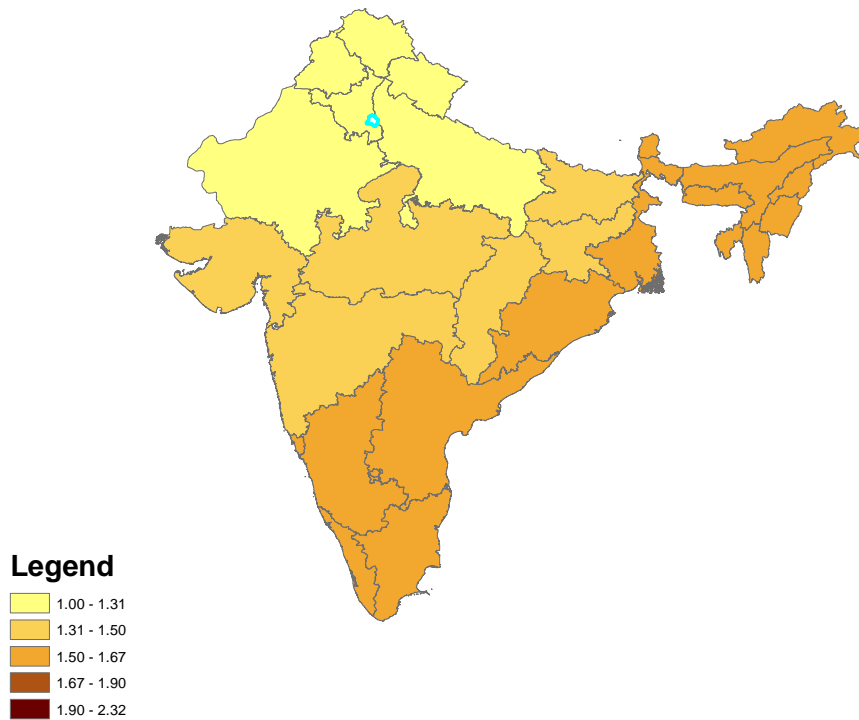


Goa. These are the states with the highest, median, and lowest gains respectively for the case in uncorrelated draws.

The size of the effects differ substantially across states. For instance, in Sikkim, income per capita increases by around 32% and 7% in the two different scenarios. These number are much smaller for the case of Goa, where income per capita only increases by 1.40% and 0.50% for the cases of uncorrelated and perfectly correlated draws.

As already emphasized in the literature, we find that the importance of *pro-competitive* gains on the overall gains crucially depends on the pattern of correlation of productivity shocks. When the pattern of comparative advantage is very strong (uncorrelated draws) most of the income gains are explained by *Ricardian* effects. For example, in the case of Uttar Pradesh, the almost 100% of the income gains are due to *Ricardian* effects. The opposite is true when decreasing transportation costs in the case of perfectly correlated draws, in which almost 100% of the increase in income is due to *pro-competitive* gains. This is driven by the different pattern on the change on the standard deviation of markups. For instance, in the state of Goa, the dispersion on markups increases a 0.54% after decreasing transportation costs in the case of uncorrelated draws, which is associated to negative *pro-competitive* effects.

FIGURE V  
MAP OF TRANSPORTATION COSTS FOR DISTRICTS LOCATED IN REGION OF DELHI  
STATE-LEVEL  
UNDER PERFECT TRANSPORTATION INFRASTRUCTURE



## 9 Conclusions

The goal of this paper has been to determine the extent to which misallocation in India is driven by high internal transportation costs. We assembled a rich micro-level data on Indian manufacturing as well as measures of road quality based on geospatial data. We find evidence that is consistent with misallocation being presenting the data as well as evidence that transportation costs play a role. We calibrate a model in which all the states of India trade with each other and in which firms face oligopolistic competition. Finally, we quantify the gains from improving infrastructure and decompose these gains into the Ricardian and pro-competitive channels under different scenarios of productivity distributions.

TABLE X  
QUANTITATIVE RESULTS

	(A) Uncorrelated draws			(B) Perfect correlated draws		
	Perfect Infrastructure			Perfect Infrastructure		
	$E_0$	$E_{cf}$	$\Delta$ (%)	$E_0$	$E_{cf}$	$\Delta$ (%)
<b>(I) SIKKIM (HIGHEST GAINS)</b>						
Income per capita	12.24	16.96	32.53	4.75	5.11	7.32
(Ricardian)			32.81			0.02
(Pro-competitive)			-0.31			7.30
Price Index	1.30	93.42	-28.40	3.66	2.92	-0.20
s.d markups	0.022	0.022	0.001	0.016	0.007	-52.704
<b>(II) UTTAR PRADESH (MEDIAN GAINS)</b>						
Income per capita	10.88	11.75	7.72	2.71	2.72	0.04
(Ricardian)			7.71			0.00
(Pro-competitive)			0.01			0.04
Price Index	0.93	0.85	-8.32	2.82	2.70	-4.15
s.d markups	0.020	0.021	-0.142	0.018	0.015	-16.226
<b>(III) GOA (LOWEST GAINS)</b>						
Income per capita	1.40	1.44	2.50	0.50	0.51	0.05
(Ricardian)			3.41		260.09	0.05
(Pro-competitive)			-0.91		11.96	0.00
Price Index	0.90	0.88	-1.52	2.72	2.75	1.07
s.d markups	0.021	0.057	0.542	0.015	0.013	-11.596

Notes: Panels (A) and (B) refer to the case in which we assume perfect correlated productivity and uncorrelated productivity draws across Indian states respectively.  $E_0$  refers to the benchmark calibrated economies;  $E_{cf}$  refers to the counter-factual economies;  $\Delta$  (%) refers to the relative change between  $E_0$  and  $E_{cf}$ . (I) shows the results for the state of Sikkim, which is the state with the highest gains in the case of uncorrelated draws. (II) shows the results for the state of Uttar Pradesh, which is the state with the median gains in the case of uncorrelated draws. (III) shows the results for the state of Goa, which is the state with the lowest gains in the case of uncorrelated draws.

## References

- AGHION, P., R. BURGESS, S. REDDING, AND F. ZILIBOTTI (2005): “Entry liberalization and inequality in industrial performance,” *Journal of the European Economic Association*, 3(2/3), 291–302.
- ALDER, S. (2014): “Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development,” *Working Paper*.
- ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2012): “The Elusive Pro-Competitive Effects of Trade,” *Working Paper*.
- ATAK, J., F. BATEMAN, M. HAINES, AND R. A. MARGO (2010): “Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850-1860,” *Social Science History*, 34(2), 171–197.
- ATKESON, A., AND A. BURSTEIN (2008): “Pricing-to-market, trade costs, and international relative prices,” *The American Economic Review*, 98(5), 1998–2031.
- BANERJEE, A. V., E. DUFLO, AND N. QIAN (2012): “On the Road: Access to Transportation Infrastructure and Economic Growth in China,” pp. 1–50.
- BOLLARD, A., P. KLENOW, AND G. SHARMA (2013): “India’s Mysterious Manufacturing Miracle,” *Review of Economic Dynamics*, 16, 59–85.
- BUERA, F., J. KABOSKI, AND Y. SHIN (2011): “Finance and Development: A Tail of Two Sectors,” *American Economic Review*, 101, 1964–2002.
- CASELLI, F., AND N. GENNAIOLI (2013): “Dynastic Management,” *Economic Enquiry*, 51(1), 971–996.
- CHARI, A. V. (2011): “Identifying the Aggregate Productivity Effects of Entry and Size Restrictions: An Empirical Analysis of License Reform in India,” *American Economic Journal: Economic Policy*, 3(2), 66–96.
- DATTA, S. (2012): “The impact of improved highways on Indian firms,” *Journal of Development Economics*, 99(1), 46–57.
- DE BLAS, B., AND K. RUSS (2010): “Understanding Markups in the Open Economy under Bertrand Competition,” Discussion paper, National Bureau of Economic Research.
- DEVEREUX, M. B., AND K. M. LEE (2001): “Dynamic gains from international trade with imperfect competition and market power,” *Review of Development Economics*, 5(2), 239–255.



- DONALDSON, D. (Forthcoming): “Railroads and the Raj: the Economic Impact of Transportation Infrastructure,” *American Economic Review*.
- EATON, J., AND S. KORTUM (2002): “Technology, Geography, and Trade,” *Econometrica*, 70(5), 1741–1779.
- EATON, J., S. KORTUM, A. B. BERNARD, AND J. B. JENSEN (2003): “Plants and Productivity in International Trade,” *The American Economic Review*, 4, 1268–1290.
- EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2012): “Competition, markups, and the gains from international trade,” Discussion paper, National Bureau of Economic Research.
- EPIFANI, P., AND G. GANCIA (2011): “Trade, markup heterogeneity and misallocations,” *Journal of International Economics*, 83(1), 1–13.
- EROSA, A., AND L. ALLUB (2013): “Financial Frictions, Occupational Choice, and Economic Inequality,” *Working Paper*.
- GARCÍA-SANTANA, M., AND J. PIJOAN-MAS (2012): “The Reservation Laws in India and the Misallocation of Production Factors,” *Working Paper*.
- GARICANO, L., C. LELARGE, AND J. V. REENEN (2012): “Firm Size Distortions and the Productivity Distribution: Evidence from France,” *Working Paper*.
- GHANI, E., A. G. GOSWAMI, AND W. KERR (2013): “Highway to success in India: the impact of the golden quadrilateral project for the location and performance of manufacturing,” *World Bank Policy Research Working Paper*, (6320).
- GOURIO, F., AND N. ROYS (2013): “Size-Dependent Regulations, Firm Size Distribution, and Reallocation,” *Working Paper*.
- GUNER, N., G. VENTURA, AND X. YI (2008): “Macroeconomic Implications of Size-Dependent Policies,” *Review of Economic Dynamics*, 11(4), 721–744.
- HOLMES, T. J., W.-T. HSU, AND S. LEE (2012): “Allocative Efficiency, Markups, and the Welfare Gains from Trade,” Discussion paper, University of Minnesota.
- HSIEH, C.-T., AND P. KLENOW (2009): “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124(4), 1403–1448.
- KOTHARI, S. (2013): “The Size Distribution of Manufacturing Plants and Development,” *Working Paper*.
- LOPEZ-MARTIN, B. (2013): “From Firm Productivity Dynamics to Aggregate Efficiency,” *Working Paper*.

- MARKUSEN, J. R. (1981): “Trade and the gains from trade with imperfect competition,” *Journal of International Economics*, 11(4), 531–551.
- MELITZ, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), 1695–1725.
- MIDRIGAN, V., AND D. XU (2014): “Finance and Misallocation: Evidence from Plant Level Data,” *American Economic Review*, 2(104), 422–58.
- MOLL, B. (Forthcoming): “Productivity Losses from Capital Frictions: Can Self-Financing Undo Capital Misallocation?,” *American Economic Review*.
- PETERS, M. (2013): “Heterogeneous Mark-Ups, Growth, and Endogenous Misallocation,” Discussion paper, Working Paper.
- RESTUCCIA, D., AND R. ROGERSON (2008): “Policy Distorsions and Aggregate Productivity with Heterogeneous Establishments,” *Review of Economic Dynamics*, 11(4), 707–720.
- (2013): “Misallocation and productivity,” *Review of Economic Dynamics*, 16(1), 1–10.

