# Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India[*]

Jose Asturias

School of Foreign Service in Qatar, Georgetown University


Manuel García-Santana

Université Libre de Bruxelles (ECARES)


Roberto Ramos

Bank of Spain

February 10, 2015

### Abstract

What are the economic channels through which transportation infrastructure affect income? We study this question using a model of internal trade in which states trade with each other. In contrast to the previous literature, we do so in a framework that incorporates pro-competitive gains: changes in transportation costs affect the distribution of markups by changing the level of competition that firms face. We apply this model to the case of the Golden Quadrilateral (GQ), a large road infrastructure project in India. We discipline the parameters of the model using micro level manufacturing and geospatial data. We find that: i) the project generates large aggregate gains, ii) both standard and pro-competitive gains are quantitatively relevant.

# 1 Introduction

Poor transportation infrastructure is a common feature in low-income countries. For example, in 2000, it would take a truck four to five days to drive the 1,500 km distance between Delhi and Calcutta, which is five times longer than it would in the United States. International organizations and policymakers have not overlooked this fact: between 1995 and 2005, upgrades to the transportation network constituted around 12% of total World Bank lending. Out of this, 75% was allocated to upgrading roads and highways. Hence, understanding the impact of large-scale transportation infrastructure projects is a matter of great importance.

In this paper, we develop a model of internal trade that allows us to quantitatively evaluate the welfare gains which stem from improving the transportation infrastructure within a country. Our main contribution is to quantify the impact of improved transportation networks in a setting which allows distinguishing between different types of welfare gains. That is, we determine to what extent reductions in transportation costs improve productive efficiency (Ricardian gains), allocative efficiency (pro-competitive gains due to less misallocation arising from changing markups), and the terms of trade for every trade partner. We use this model to study the welfare impact of building one of the biggest highway networks in the world: the Golden Quadrilateral (GQ) in India. The GQ project upgraded and expanded the roads connecting the four major cities in the country, providing India with around 6,000 km of modern highway roads.

In our model, the states of India trade with each other. There is a continuum of sectors and each sector has firms with heterogeneous productivity competing à la Cournot. The model generates endogenous heterogeneous markups. This heterogeneity in markups results in variations in the marginal revenue product of labor (MRPL) across firms.[1] Changes in transportation costs affect misallocation in the economy since they change the market power that firms have and the distribution of markups. Our model is a version of Atkeson and Burstein (2008) and Edmond, Midrigan, and Xu (2014), extended to include multiple non-symmetric economies.

Relative to previous work on transportation infrastructure, our framework allows us to separate the standard Ricardian from the pro-competitive channel to account for the welfare gains stemming from lower transportation costs. Our decomposition of the welfare gains follows the methodology developed by Holmes, Hsu, and Lee (Forthcoming). The Ricardian component is simply the gains in real income if all firms charge their marginal cost. This component maps back to welfare in models in which firms have constant markups or operate in perfect competition. The pro-competitive component relates to the misallocation arising from the heterogeneous markups charged by firms. This misallocation arises because the consumption of goods produced by firms with high markups is inefficiently low. The last component is the terms of trade, which compare the average markup of the goods sold with the average markup of the goods purchased by the

---

[1]Recent papers have highlighted misallocation as a source of cross-country income differences. These papers include Hsieh and Klenow (2009), Restuccia and Rogerson (2008), and Guner, Ventura, and Xu (2008).

state. Ceteris paribus, states with high markups on the goods that they sell relative to the goods that they buy will enjoy a higher real income.

In order to discipline the key parameters of the model, we use a rich micro-data set of Indian manufacturing plants and geospatial data on the Indian road network at several points in time. First, we combine two separate sets of micro-level data - the Annual Survey of Industries (ASI) and the National Sample Survey (NSS) - in order to construct a very detailed description of the Indian manufacturing sector over time, covering both formal and informal firms. Crucially, from these data we derive measures of internal trade and prices paid across destinations. Second, we use GIS information on the entire Indian road network in order to compute measures of effective distance across destinations, taking the quality of the roads and the evolution of the transportation network over time into account.

We derive a set of structural equations from the model that allow us to estimate the key parameters. One implication of the model is that transportation costs can be identified by comparing the prices charged by monopolistic firms across destinations. This is the case because the prices charged by these firms only depend on transportation costs across locations, as the level of competition they face is constant. To implement this strategy, we first identify all the goods that are produced by only one plant in India. We then regress the prices paid for these goods across locations against the effective distance between the location of the monopolistic producer and the location of the plant that uses it as an intermediate input. Our measure of effective distance takes the least costly path along the Indian road network into account, incorporating differences in road quality due to the GQ. Using the coefficients of the regression, we construct a matrix of bilateral transportation costs between Indian states for both 2001 (before the GQ) and 2006 (after the GQ). Our quantitative exercise consists in comparing outcomes under these two sets of transportation costs.

Our next step is to identify the elasticity of substitution across sectors. This parameter governs the price elasticity of the demand curve of a sector, and hence determines the market power of firms that are monopolists in their sector. We use intermediate input usage data to construct trade flows for goods produced by monopolists. For these goods, the model implies a gravity equation that relates bilateral flows to transportation costs. We use internal trade flows and estimated transportation costs to measure how trade flows decline with increases in transportation costs. We set the elasticity of substitution across sectors to match the gravity equation of monopolist trade flows in the data.

We next estimate the elasticity of substitution across products within the same sector. This parameter determines the elasticity of demand faced by firms with a small market share in their sectors. In order to do that, we exploit a linear relationship between sectoral shares and labor shares implied by the model. In the model, firms with higher sectoral shares also charge higher markups, and hence have lower labor shares. The strength of this relationship depends on the

gap between the elasticity of substitution both across and within sectors. Given our estimate of the elasticity of substitution across sectors, we set an elasticity of substitution within sectors that matches the slope coefficient of an OLS cross-sectional regression of the labor shares of plants against their sectoral shares.

The values of these elasticities are crucial to quantify the size of the pro-competitive gains, since they determine the amount of misallocation that can potentially occur in the economy. The gap between the two elasticities in particular determines to what extent firms with higher sectoral shares enjoy more market power in their sector and charge higher markups. We set the rest of the parameters such that, in equilibrium, the model reproduces some important features of the Indian manufacturing sector.

Next, we quantify the effects of the construction of transportation infrastructure. To do so, we measure the impact of changing the transportation costs in the model to those estimated for 2001, before the construction of the GQ. We find that the aggregate gains for India derived from the construction of the GQ are 2.05% of real income. Since we only consider the manufacturing sector in our model, the result is in terms of manufacturing value-added. Putting the welfare gains from the model into dollar amounts yields a gain of *$3.1 billion per year*. Since the GQ cost $5.6 billion to build, our model predicts that it would take only two years for India to recover the initial cost.

We find that, on aggregate, the pro-competitive gains account for 17% of the total gains from the construction of the GQ. These pro-competitive gains are positive in all but one state, and can be as high as 26% in some of them. This means that the GQ helped reduce the misallocation arising from variations in the market power of firms.

Importantly, we also find wide heterogeneity in terms of welfare effects across states. States closest to the GQ gained the most, while those farthest had modest or even negative welfare gains. The negative effects stemming from the construction of the GQ come from the interplay of two forces. First, these states benefited from lower transportation costs. Despite their location, shipments can still travel for at least part of the route on the GQ, allowing these states to import goods at a lower price. Second, the states that are closer to the GQ started trading more intensively with each other, which implies increases in wages in these states. This translates into an increase in the cost of purchasing goods from these states. Some states which are far from the GQ lost because this higher cost of purchasing goods from other states was not compensated for by the decrease in prices due to lower transportation costs.

We also find important changes in trade patterns among Indian states. We find in particular that there was *trade diversion*: states close to the GQ diverted their trade towards states that experienced a greater decline in transportation costs. In fact, some states far from the GQ, such as the Northeast region of the country, became less open (value of exports as a fraction of state income) after the construction of the GQ due to this effect. Furthermore, we find that trade in India became more regionally concentrated. States that are on the GQ traded more intensively

with each other, as did states which are not located on the GQ. Often, trade between states on the GQ and off the GQ *declined.*

Lastly, we apply a difference-in-difference strategy to our data in order to isolate the effect of the GQ on prices and compare it with the outcome of the calibrated model. To do so, we compare the prices paid for intermediate goods by firms close to the GQ and by those that are further away before and after the construction of the highway. This strategy accounts for the potential endogeneity of infrastructure development, by focusing on price changes in non-nodal districts close to and further away from the road network. We find that, in the data, the change in prices in non-nodal districts crossed by the GQ was around 36 percentage points lower than in districts further away, implying a 1.57 times bigger decrease in prices in districts crossed by the GQ. We find a similar effect in magnitude when computing the equivalent differential effect with our calibrated model. The model predicts that the decrease in prices in states crossed by the GQ is 1.96 times larger than in average states.

The remainder of the paper is organized as follows. In Section 2, we present the related literature. In Section 3, we describe the main characteristics of the road network in India. In Section 4, we present the model. In Section 5, we describe our data, which we use to calibrate the model in Section 6. In Section 7, we present and discuss our quantitative results. In Section 8, we present quantitative results for alternative scenarios to the GQ. Finally, Section 9 concludes.

## 2  Related Literature

Our work builds on papers that quantify the gains from building transportation infrastructure using general equilibrium models of trade. The pioneering work of Donaldson (Forthcoming) studies the benefits from the construction of railroads in colonial India. Herrendorf, Schmitz, and Teixeira (2012) study the impact of transportation improvements during the 19th century United States transportation revolution on the regional distribution of population and welfare. Adamopoulos (2011) and Sotelo (2014) study the income losses due to high transportation costs for agricultural products in developing countries. Allen and Arkolakis (Forthcoming) use a novel theoretical framework to calculate the welfare effects of the construction of an interstate highway system in the United States. In a more recent paper, Donaldson and Hornbeck (2014) develop the "market access" approach in order to assess the gains from the construction of railroads in the United States. Alder (2014) uses this approach to study the impact of the hypothetical construction in India of a highway network similar to that of China. None of the papers in this literature consider how reductions in transportation costs affect markups. Our paper is the first attempt to evaluate how improvements in infrastructure impact welfare through the pro-competitive channel.[2]

Our paper also contributes to the literature that examines the pro-competitive effects of in-

---

[2]See Redding and Turner (2014) for a recent comprehensive survey of the literature on the relationship between economic outcomes and transportation costs.

ternational trade. These papers study how trade affects the markups that firms charge and the resulting impact on welfare. Markusen (1981) is an example of early work in this area. Epifani and Gancia (2011), de Blas and Russ (Forthcoming), and Holmes, Hsu, and Lee (Forthcoming) provide theoretical contributions to this literature. Devereux and Lee (2001) study how high markups induce inefficiently high levels of entry. More generally, Bernard, Eaton, Jensen, and Kortum (2003) and Melitz and Ottaviano (2008) develop widely used models of international trade with endogenous markups. We differ from these papers in that our aim is to *quantify* the pro-competitive effects of reducing transportation costs. Such quantification is useful since theory is ambiguous as to whether pro-competitive effects are quantitatively significant. In fact, theory indicates that pro-competitive effects can be negative, as stressed by Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012).

In this sense, our paper builds on Edmond, Midrigan, and Xu (2014). They quantify the pro-competitive gains channel by using a model in which Taiwan trades with the rest of the world. We extend this analysis to a non-symmetric multi-country setting and use it to evaluate the impact of infrastructure. Feenstra and Weinstein (2010) quantify the effects of changing markups on US welfare in a context of monopolistic competition.[3]

Our paper is also related to Arkolakis, Costinot, and Rodriguez-Clare (2012) and the set of commonly used trade models that they consider in their paper. In these models, all firms charge the same markup or operate under perfect competition. Our paper is different because it also considers the effects of changing markups after a reduction in trade costs.

Our work contributes to a large literature on the misallocation of resources across firms. Papers from this literature include Restuccia and Rogerson (2008), Guner, Ventura, and Xu (2008), Hsieh and Klenow (2009), and Gabler and Poschke (2013).[4] We contribute to these papers by evaluating how improvements in transportation infrastructure alleviate the misallocation.

## 3  Roads in India and the Golden Quadrilateral

India has the second largest road network in the world, spanning approximately 3.3 million kilometers. It comprises expressways, national highways (79,243 km), state highways (131,899 km), major district highways, and rural roads. Roads play an important role in facilitating trade in

---

[3]Asturias and Petty (2013) quantify pro-competitive gains in the transportation industry after a trade liberalization. Licandro and Impullitti (2013) study pro-competitive gains through a model in which firms can innovate.

[4]There are many recent papers that emphasize the misallocation of resources across firms as a source of income difference across countries. 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 Garcia-Santana and Pijoan-Mas (2014) study the marginal effect of size-dependent policies in France and India respectively. Peters (2013) calibrates a model of imperfect competition with heterogeneous firms to Indonesian data to investigate the impact of misallocation on growth. See Restuccia and Rogerson (2013) and Hopenhayn (2014) for nice surveys of the literature.

India: approximately 65 percent of freight in terms of weight and 80 percent of passenger traffic are transported on roads.[5] National highways are critical since they facilitate interstate traffic and carry about 40 percent of the total road traffic.

At the end of the 1990s, India's highway network left much to be desired. The major economic centers were not linked by expressways, and only 4% of roads had four lanes. In addition to the limited lane capacity, more than 25% of national highways were considered to be in poor surface condition.

Congestion was also an important issue, with 25% of roads categorized as congested. This was due to poor road conditions, increased demand from growing traffic, and crowded urban crossings. Frequent stops at state or municipal checkpoints for government procedures such as tax collection or permit inspection also contributed to congestion (see World-Bank (2002)).

In order to improve this situation, the Indian government launched the National Highways Development Project (NHDP) in 2001. The goal of the initiative was to improve the performance of the national highway network. The first phase of the project involved the construction of the Golden Quadrilateral (GQ), a 5,800 km highway connecting the four major metropolitan areas via four- and six-lane roads. The four metropolitan centers that were connected are Delhi, Mumbai, Chennai, and Calcutta. Apart from the increase in the number of lanes, additional features of a high-quality highway system were constructed. These features include grade separators, overbridges, bypasses, and underpasses.

The cost was initially projected to be 600 billion rupees (equivalent to $13.4 billion in 2006). As of October 2013, the total cost incurred by the Indian government was approximately half of the projected sum (250 billion rupees or $5.6 billion). In Section 7, we compare this cost with the benefits predicted by our model.

The second phase of the NHDP consists in the construction of the North-South and East-West corridor, a highway that aims to connect Srinagar in the north to Kanyakumari in the south and, Silchar in the east to Porbandar in the west. Although this second phase was approved in 2003, there have been many delays for its construction, and less than 10% of the work was completed by the end of 2006. Thus, we will not consider that project in our analysis.

**Geospatial data**   We have geospatial data for all the National Highways of India, which was supplied by ML Infomap. We complement this data using information provided by the National Highways Authority of India (NHAI) on the completion dates of various portions of the GQ. The GQ consisted of 127 stretches and we have detailed information about the start and end points.[6] Figure I shows the evolution of the GQ (in red) in 2001 and 2006. Although the GQ was finished

---

[5]The importance of railroads has declined in India over time. Although in 1950 more than 80% of freight traveled by rail, this figure has steadily been decreasing. At present, rail carries mostly bulk freight such as iron, steel, and cement. Non-bulk freight represents only around 3 percent of total rail freight in terms of ton-km.

[6]See nhai.org/completed.asp and the Annual Reports of NHAI.

in 2013, more than 90 percent of the project was completed by 2006. We will link this geospatial data to manufacturing data for 2001 and 2006.

FIGURE I

ROAD NETWORK IN INDIA AND THE GQ

A: GQ construction in 2001

B: GQ construction in 2006
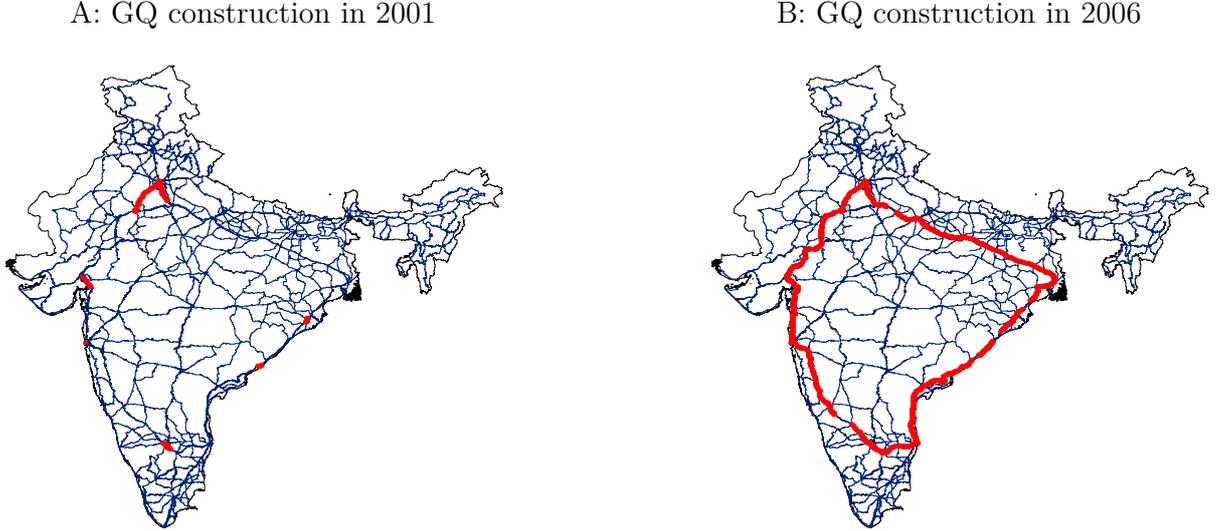


Panel A of Figure I shows a map with the road network in India at the end of 2001, including the sections of the Golden Quadrilateral that were finished by then (around 10% of the total project). Panel B shows the same map but for 2006 (around 95% of the total project).

## 4 Model

In this section, we present our static general equilibrium model of internal trade. We consider N asymmetric states trading with each other. In each state, there is a measure 1 of sectors. Within each sector, there is a finite number of firms that compete in an oligopolistic manner. Labor is immobile across states.[7]

### 4.1 Consumers

In each state $n$, there is a representative household with a utility function:

$$C_n = \left( \int_0^1 C_n(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}}, \tag{1}$$

where $C_n(j)$ is the composite good of sector $j$ and $\theta > 1$ is the elasticity of substitution across composite goods of different sectors. The sector-level composite good is defined as:

---

[7]Interstate migration flows in India are among the lowest in the world. According to the 2001 Indian Population Census, around 96% of people report to be living in the state where they were born.

$$C_n(j) = \left( \sum_{o=1}^{N} \sum_{k=1}^{K_{oj}} c_n^o(j,k)^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}}, \tag{2}$$

where $c_n^o(j,k)$ is the good consumed by state $n$ and provided by firm $k$ in sector $j$ shipped from state $o$, $N$ is the number of states, $K_{oj}$ is the number of firms that operate in sector $j$ in state $o$, and $\gamma > 1$ is the elasticity of substitution between goods produced by different firms in the same sector. We assume that $\gamma > \theta$, which means that goods are more substitutable *within* sectors than *between* sectors.

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

$$\int_0^1 \left( \sum_{o=1}^{N} \sum_{k=1}^{K_{oj}} p_n^o(j,k) c_n^o(j,k) \right) dj = W_n L_n + \Pi_n, \tag{3}$$

where $W_n$ is the equilibrium wage, $L_n$ is the labor endowment, and $\Pi_n$ is the income derived from the profits of firms located in $n$. Note also that $C_n = W_n L_n + \Pi_n$.

## 4.2 Firms

In each sector $j$ in state $o$, there is a finite number of $K_{oj}$ firms. Firms draw their productivity from a distribution with CDF $G(a)$. A firm with a productivity level $a$ has a constant labor requirement of $1/a$ to produce one unit of good. Because firms do not pay a fixed cost to operate in a market, they sell to all $N$ states.

To determine the firm's pricing rule, we first find the demand it faces. Equations (1), (2), and (3) generate the demand:

$$c_n^o(j,k) = \left( \frac{P_n}{P_n(j)} \right)^{\theta} \left( \frac{P_n(j)}{p_n^o(j,k)} \right)^{\gamma} C_n, \tag{4}$$

where

$$P_n(j) = \left( \sum_{o=1}^{N} \sum_{k=1}^{K_{oj}} p_n^o(j,k)^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \tag{5}$$

is the price index for sector $j$ in state $n$ and

$$P_n = \left( \int_0^1 P_n(j)^{1-\theta} dj \right)^{\frac{1}{1-\theta}} \tag{6}$$

is the aggregate price index in state $n$. Intuitively, the relative demand for a differentiated good within a sector depends on the price of the good relative to the price of the composite good of the sector, and also on the price of the composite good of the sector relative to the aggregate price index.

Firms within sectors compete à la Cournot. Firm $k$ takes the demand characterized by equation (4) and the quantity supplied by competitor firms in the sector as given and solves the following

problem:

$$\pi_d^o(j,k) = \max_{c_d^o(j,k)} p_d^o(j,k)c_d^o(j,k) - \frac{W_o \tau_d^o}{a_o(j,k)}c_d^o(j,k), \tag{7}$$

where $a_o(j,k)$ is the productivity of firm $k$ in sector $j$ producing in state $o$, $\tau_d^o$ is the iceberg transportation cost to ship one unit of good from $o$ to $d$. Note that, because of the constant returns to scale technology, the problem of a firm across all different destinations can be solved independently. The solution to this problem is:

$$p_d^o(j,k) = \frac{\epsilon_d^o(j,k)}{\epsilon_d^o(j,k)-1}\frac{W_o}{a_o(j,k)}\tau_d^o, \tag{8}$$

where

$$\epsilon_d^o(k,j) = \left(\omega_d^o(j,k)\frac{1}{\theta} + (1-\omega_d^o(j,k))\frac{1}{\gamma}\right)^{-1}, \tag{9}$$

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

$$\omega_d^o = \frac{p_d^o(j,k)c_d^o(j,k)}{\sum_{o=1}^N \sum_{k=1}^{K_{oj}} p_d^o(j,k)c_d^o(j,k)}. \tag{10}$$

The price that firms set in equation (8) is similar to the markup over marginal cost that is found in a setup with monopolistic competition. The difference is that the markups are endogenous here, and depend on the market structure of the sector. For example, suppose that there is only one firm in a given sector, then that firm will compete only with firms operating in other sectors and its demand elasticity will be equal to $\theta$. This means that the firm faces the sector-level elasticity of demand. At the other extreme, suppose that a firm's market share is close to zero, then the firm will compete only with firms in its own sector and its elasticity of demand will be equal to $\gamma$. Notice that a given firm will generally have different market shares and hence charge different markups across different destinations.

The aggregate profits of firms in state $n$ are characterized by:

$$\Pi_n = \int_0^1 \left(\sum_{n=1}^N \sum_{k=1}^{K_{oj}} \pi_n^o(j,k)\right) dj. \tag{11}$$

## 4.3 Balanced Trade and Labor-Clearing Condition

All states $n$ must have balanced trade:

$$\int_0^1 \left(\sum_{o=1,\ o\neq n}^N \sum_{k=1}^{K_{oj}} p_n^o(j,k)c_n^o(j,k)\right) dj = \int_0^1 \left(\sum_{d=1,\ d\neq n}^N \sum_{k=1}^{K_{nj}} p_d^n(j,k)c_d^n(j,k)\right) dj. \tag{12}$$

The labor-clearing condition for state $n$ is:

$$\int_0^1 \left(\sum_{d=1}^N \sum_{k=1}^{K_{nj}} \frac{c_d^n(j,k)}{a_n(j,k)}\tau_d^n\right) dj = L_n. \tag{13}$$

9

### 4.4 Definition of Equilibrium

*Equilibrium.* For all states $n$ and $n'$, sectors $j$, and firms $k_{nj}$, an equilibrium is a set of allocations of consumption goods $\{c_{n'}^n(j,k),\ C_n(j)\}$, firm prices $\{p_{n'}^n(j,k)\}$, sector prices $\{P_n(j)\}$, and aggregate variables $\{W_n,\ P_n,\ \Pi_n\}$ such that:

1. Given firm prices, sector prices, and aggregate variables, $\{c_{n'}^n(j,k)\}$ is given by (4), $C_n(j)$ by (2), and they solve the consumer's problem in (1), and (3).

2. Given aggregate variables, $p_{n'}^n(j,k)$ is given by (8), (9), and (10), and solves the problem of the firm in (7).

3. Aggregate profits satisfy (11), aggregate prices satisfy (6), and sector prices satisfy (5).

4. Trade flows satisfy (12).

5. Labor markets satisfy (13).

### 4.5 Misallocation in the Model

Misallocation in this setting arises due to *dispersion* in markups across producers: the marginal revenue product of labor (MPRL) of firms with high markups becomes inefficiently high, which implies that the goods produced by these firms are under-consumed relative to the goods produced by firms with low markups.[8] The model is hence relevant to think about the cross-firm misallocation emphasized by Restuccia and Rogerson (2008), Hsieh and Klenow (2009), and Guner, Ventura, and Xu (2008), among others.

These papers have interpreted this misallocation as resulting from government policies that create idiosyncratic distortions at the firm level, which affect the optimal decision of firms. In our model, dispersion in MRPL is caused by dispersion in the market power, which translates into variations in markups: firms with higher productivity draws charge higher markups because they are able to capture bigger market shares. The constant-returns-to-scale technology implies that the MRPL of a firm operating in state $o$ is $W_o{}^{\epsilon(j,k)}/_{(\epsilon(j,k)-1)}$. Thus, firms with high productivity draws (and high markups) also have a high MRPL.

This misallocation is hence similar in nature to the one studied by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), for the particular case in which the size of the idiosyncratic distortions of firms is positively correlated to their productivity. Firms with high productivity draws are smaller in size than in the case of perfect competition. Thus, India's aggregate welfare would increase by reallocating labor from firms with low productivity draws (low-markup firms) to firms with high productivity draws (high-markup firms).

---

[8]MRPL is the price of the good multiplied by the marginal product of labor. This is equivalent to the TFPR in Hsieh and Klenow (2009) since labor is the only factor of production, and the production function exhibits constant returns to scale.

# 5 Plant-Level Data on Indian Manufacturing

In this section, we describe the construction of the data set used in the paper. We link firm-level data on the Indian manufacturing sector with geospatial data in order to construct two snapshots in time (2001 and 2006) with detailed manufacturing and road quality data. The data provides the necessary information to analyze how changes in infrastructure quality affect the manufacturing sector. See Section (A) of the Online Appendix for details about the data.

We first construct a representative sample of the Indian manufacturing sector. To do so, we merge two separate sets of plant-level data: the Annual Survey of Industries (ASI) and National Sample Survey (NSS). The ASI targets plants that are in the formal sector. It is the main source of manufacturing statistics in India and has been commonly used in the development literature.[9] It concerns plants that have more than 10 workers if they have electricity and 20 if they do not. The information provided by the establishments is very rich, covering several operational characteristics: sales, employment, capital stock, wage payments, and expenditures on intermediate goods. The NSS covers all informal establishments in the Indian manufacturing sector. "Informal" refers to all manufacturing enterprises not included in the ASI. The survey is conducted every five years by the Indian Ministry of Statistics, as one of the modules in the Indian National Sample Survey.

The process of merging the data from the ASI and NSS is straightforward since very similar questions are used to collect the data. Thus, we can create a representative sample of manufacturing plants in India using the weights provided. After merging the ASI and NSS, we have around 190,000 observations for the fiscal year 2000-2001 and 140,000 observations for the fiscal year 2005-2006. Once these observations are properly weighted, for each year, we have around 17 million manufacturing plants in our data, which employ around 45 million workers.

It is important to note the huge differences in productivity between formal and informal plants in India. Informal plants account for around 80% of employment and around 20% of total value-added. Thus, it is crucial to merge these data sets to have an accurate picture of the Indian manufacturing sector.

**Prices and the consumption of intermediates**   The ASI and NSS contain detailed information about production and intermediate good usage. For each plant in our data, we observe the value and physical quantity of production and intermediate input usage broken down by product.[10] This means that we can compute the output prices charged by plants and the input prices paid by plants.[11] To compute the price of inputs, we divide the expenditure on a particular good by

---

[9]See for instance Aghion, Burgess, Redding, and Zilibotti (2005), Chari (2011), Hsieh and Klenow (2009), and Bollard, Klenow, and Sharma (2013).

[10]All plants report intermediate inputs imported from outside India separately from those which are not imported. This is important for our analysis, since we abstract from international trade in this paper.

[11]Although these data sets are starting to be widely used (see Garcia-Santana and Pijoan-Mas (2014) and Hsieh and Klenow (2009) for example), not much attention has been paid to the price information. A notable exception

physical units.

The product classification used in both the ASI and NSS is the Annual Survey of Industries Commodities Classification (ASICC). The ASICC classifies around 5,400 different products, which are very narrowly defined. For instance, the ASICC distinguishes between different types of black tea: leaf, raw, blended, unblended, dust, etc. In the processed mineral category, for example, the ASICC distinguishes between around 12 different types of coke.

# 6 Inferring Parameter Values

We calibrate our model to 2006, when the GQ was already in place. Our calibration strategy is as follows. Our model is characterized by (i) a set of bilateral iceberg costs between states $\{\tau_d^o\}_{d=1}^{d=N}$ for all $o$, (ii) the elasticity of substitution across sectors $\theta$, (iii) the elasticity of substitution within sectors $\gamma$, (iv) a number of producers for each state-sector $K_{ij}$ for $i$ and $j$, (v) a set of labor endowments $\{L_n\}_{n=1}^{n=N}$ of the states, and (vi) the parameters governing the productivity distribution of the firms.

Using structural equations from the model, we first estimate the transportation costs and the two elasticities (Sections 6.1, 6.2, 6.3, and 6.4). We next plug the number of firms per state-sector that we observe in the data into the model, and calibrate the labor endowment of the states and the productivity distribution to match the relevant statistics for the Indian manufacturing sector (Section 6.5).

## 6.1 Estimating Transportation Costs

The first step is to infer transportation costs. To do so, we use pricing data from intermediate inputs used across India. Equation (8) shows that the prices charged by firms depend both on transportation costs and market shares in the destination market.[12] In order to identify transportation costs, we exploit one implication of the model: variations in prices for monopolists (i.e. firms with market shares equal to one) are due solely to variations in transportation costs across destinations. To see this formally, equation (8), along with the fact that a monopolist firm faces a demand elasticity given by $\theta$, implies that the firm will charge:

$$p_d^o(j,k) = \frac{\theta}{\theta-1} \frac{W_o}{a_o(j,k)} \tau_d^o. \tag{14}$$

Then, the relative price charged by a monopolist across destinations is:

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

is Kothari (2013).

[12] Atkin and Donaldson (2014) also study issues regarding the fact that dispersion in prices contain information about markups and transportation costs.

which only depends on the ratio of transportation costs. Hence, the prices charged by monopolists across states reveal differences in transportation costs.

Empirically, we define a monopolist firm as a plant selling at least 95 percent of the value of each 5-digit ASICC product nationally. Using the ASI and NSS for the years 2001 and 2006, we identify 261 products that are manufactured by monopolists.[13] The largest category is "Manufacture of chemicals and chemical products," which contains around 40 percent of the products identified. This is consistent with the nature of the chemical industry, in which production is often concentrated in one plant due to economies of scale and then shipped to many locations.[14]

Once the products manufactured by monopolists are identified, the strategy is to use the variation in prices across locations where these products were used as intermediate inputs to identify transportation costs. We regress variations in prices on a measure of transportation costs that we call effective distance. This measure takes the least costly path to go from origin to destination given the road structure into account. Furthermore, the varying road quality is also incorporated into this measure. The price of an input in each district is computed as the weighted average of the prices paid by all the plants that use this input in the district.

We estimate equation (14) as follows:

$$\log p_{d,t}^o(j,k) = \beta \log \text{Effective Distance}_{d,t}^o + \sum_o \delta_o + \sum_j \alpha_j + \sum_t \eta_t + \epsilon_{d,t}^o(j,k) \tag{15}$$

where $p_d^o(j,k)$ is the average price in district $d$ paid for product $j$ produced by a monopolist located in district $o$, $\delta_o$ are a set of districts of origin fixed effects, $\alpha_j$ a set of product fixed effects, $\eta_t$ are time dummies, and $\epsilon_{d,t}^o(j,k)$ is the error term. The origin fixed effects control for local wages and the product fixed effects control for firm productivity.[15]

In order to compute the effective distance, we first convert the national highway network into a graph. The graph consists of a series of nodes that are connected by arcs. In our case, a node is the most populous city in each district and an arc is the national road that connects them. An arc is referred to as being GQ or non-GQ, depending on whether it was completed in the specific year. We then use Dijkstra's shortest-path algorithm to construct a matrix of lowest-cost distances between all the districts for the years 2001 and 2006. The transportation costs in these two years are different since this algorithm takes the fact that traveling on a better quality road (i.e. on the Golden Quadrilateral) is less costly into account. Specifically, we assume that:

---

[13]We exclude goods that are not used as intermediate inputs in at least five districts.

[14]A description of the production structure of the chemical industry in India can be found at http://smallb.in/sites/default/files/knowledge_base/reports/IndianChemicalIndustry.pdf

[15]We exploit the cross-sectional variation using the two years in our sample to estimate the relationship between prices and effective distance in equation (15). Although we calibrate the model to the year 2006, we proceed in this way in order to maximize power in our estimations.

$$\text{Effective Distance}_{n_2}^{n_1} \quad = \quad \text{Road Distance}_{n_2}^{n_1} \text{ if GQ} = 0 \tag{16}$$
$$\text{Effective Distance}_{n_2}^{n_1} \quad = \quad \alpha \text{Road Distance}_{n_2}^{n_1} \text{ if GQ} = 1,$$

where $n_1$ and $n_2$ are nodes, and $\alpha$ indicates the effective distance of the GQ relative to stretches of road that are not GQ. We use a value of $\alpha = 0.52$, which is based on average speeds calculated by the World Bank.[16] This value of $\alpha$ indicates that if a given stretch is GQ, the effective distance is roughly half of what it is if it is not GQ. The effective distance used to estimate equation (15) is the sum of the effective distance along all the arcs traveled along the shortest path.

Table I presents the results from estimating equation (15). In column (1), we show that a 10 percent increase in the effective distance is associated with a 0.86 percent increase in the price of the good.[17] In column (2), we use a more flexible specification, in order to incorporate potential non-linearities in transportation costs. Note that such a flexible specification is commonly used to estimate the parameters of trade models using gravity equations, for instance in Eaton and Kortum (2002) and Waugh (2010). We include ten deciles of log effective distance, and find that the highest deciles are associated with large increases in the price of the good. We find, for instance, that the prices paid at destinations falling in the second decile of effective distance (around 280 km) are 37% higher than the prices paid at destinations within the first decile (70 km on average). The effect is particularly strong for destinations that are very far from the location where production takes place: the prices are around 52% higher when the effective distance to the destination is in the 10th percentile of the distribution. The 10th decile includes districts located more than 1,800 kilometers away in effective distance, which is roughly the road distance from New York City to Des Moines, Iowa. Although the overall pattern is increasing, the effect seems to be non-monotonic. For example, the coefficient associated with the third decile is 8 percentage points lower than the second decile coefficient. In order to avoid having non-monotonic transportation costs to effective distance in the model, we assume that the relationship between iceberg costs and effective distance is given by a discrete cubic function $g(\text{Coeff. of Effective Distance}_d^o)$, and set the parameters that better fit the coefficients implied by the regression. See Section (B) of the Online Appendix for details.

Lastly, we assume that the iceberg cost for all destinations in the first decile is equal to one.

---

[16]The value of $\alpha$ is based on the fact that the average speed on a national highway is between 30 and 40 km/h according to World-Bank (2002). By contrast, the average speed on the GQ is estimated to be around 75 km/h. This can be computed by calculating the predicted average speed traveling from a random sample of origins-destinations over GQ roads using Google Maps.

[17]Costinot and Donaldson (2012) estimate a similar regression for the price of agricultural goods and their distance to the nearest wholesale market over time in the United States. They find coefficients for distance of a similar magnitude during the 1880-1920 period (0.09 to 0.14). Note that the effective distance is exactly equal to the road distance before the construction of the GQ.

The iceberg cost predicted for all other deciles becomes:

$$\hat{\tau}_d^o = e^{\text{g(Coeff. of Effective Distance}_d^o)}. \tag{17}$$

**Direct measures of transportation prices**  In order to have additional estimates of transportation costs across Indian locations, we have assembled an additional data set that contains information on prices charged for shipping a container by truck within India. We have in particular collected pricing data for shipping a container of size 20 ft x 8 ft x 8.5 ft for around 900,000 origin-destination Indian city pairs. Using this data set, we construct measures of bilateral iceberg costs as a function of effective distance and compare them with the ones we obtain from equation (17). The overall pattern of the two sets of iceberg costs is remarkably similar. Importantly, in both cases, transportation costs increase more than linearly at higher levels of the distribution of effective distance. See Section (C) of the Online Appendix for further details about the construction of these transportation costs.

**What do transportation costs look like?**  As a starting point, we will take the district of New Delhi (located in the National Capital Territory of Delhi) in the year 2001. Panel A of Figure II shows a map of the transportation costs to all districts from New Delhi. The legend on the map shows transportation costs divided into quartiles. The figure also shows that only a small portion of the GQ had been upgraded by this point (depicted in red). The first thing to notice is the concentric circles that surround New Delhi. This means that the further the destination, the higher the transportation costs. The concentric circles also show that straight-line distances are highly correlated to the shortest path on the highway system. The reason is that the highway system is dense, as can be seen in Figure I. The second thing to notice is the general level of transportation costs. The map shows iceberg costs of 1.43-1.50 for transporting goods from New Delhi to the southern tip of India.

Our next step is to look at transportation costs from New Delhi in the year 2006 (panel B of Figure II), after a large part of the upgrade of the GQ had been completed. The color categories for the map have not changed compared to panel A, so that the colors are comparable across maps. The lighter colors reflect a general decrease in transportation costs.

## 6.2 Estimating the Across-Sector Elasticity of Substitution ($\theta$)

The next step consists in estimating the elasticity of substitution across sectors. The identification strategy is to compare the differences in the transportation costs of the goods produced by monopolists across destinations with the trade flows across these destinations.

Formally, we derive a gravity equation implied by the model for the trade flows of monopolist firms. Combining equations (4) and (14), we derive the following condition for the trade flow values:

## TABLE I
### IMPACT OF ROAD DISTANCE AND INFRASTRUCTURE QUALITY ON PRICES

| | (1) | (2) |
|---|---|---|
| Dep. Variable: Log price at district of destination | | |
| Log Effective Distance | 0.086*** | |
| | (0.023) | |
| Log Effective Distance $2^{nd}$ decile | | 0.371*** |
| | | (0.115) |
| Log Effective Distance $3^{th}$ decile | | 0.298*** |
| | | (0.114) |
| Log Effective Distance $4^{th}$ decile | | 0.137 |
| | | (0.112) |
| Log Effective Distance $5^{th}$ decile | | 0.168 |
| | | (0.131) |
| Log Effective Distance $6^{th}$ decile | | 0.398*** |
| | | (0.121) |
| Log Effective Distance $7^{th}$ decile | | 0.355*** |
| | | (0.133) |
| Log Effective Distance $8^{th}$ decile | | 0.445*** |
| | | (0.142) |
| Log Effective Distance $9^{th}$ decile | | 0.341** |
| | | (0.141) |
| Log Effective Distance $10^{th}$ decile | | 0.516*** |
| | | (0.136) |
| District of Origin Fixed Effects | YES | YES |
| Product Fixed Effects | YES | YES |
| Year Fixed Effects | YES | YES |
| Observations | 2,235 | 2,235 |
| R-squared | 0.876 | 0.881 |

Table I shows the estimation of equation (15). The dependent variable is the log price of a product manufactured by a monopolist at destination. The variable of interest is the effective distance between the district where the product is manufactured and the district of destination. Effective distance is defined as the lowest cost path between both districts, taking into account road distance and infrastructre quality. Specifically, going across the Golden Quadrilateral reduces road distance 48 per cent, relatives to roads not in the Golden Quadrilateral. The lowest path is computed by means of road networks and applying the Dijkstra's search path alogrithm. Column (1) uses a linear specification of effective distance, whereas column (2) estimates a non-linear specification, using 10 deciles of effective distance. District of origin, product and year -2001 and 2006- fixed effects are included. Robust standard errors are in parenthesis. Significance levels: *: 10%; **: 5%; ***: 1%.

$$\log c_d^o(j, k) p_d^o(j, k) = (1 - \theta) \log W_o + (\theta - 1) \log a_o(j, k) + \log P_d^\theta Y_d \quad (18)$$

$$+ (1 - \theta) \log \tau_d^o + (1 - \theta) \log \frac{\theta - 1}{\theta}.$$

The model predicts that higher transportation costs reduce trade flows, and the strength of this relationship depends on the value of $\theta$. The intuition behind this identification strategy is that if small differences in transportation costs across 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 straightforward relationship only holds when firms are monopolists.

We estimate equation (18) as follows:

$$\log \text{Sales}_{d,t}^o(j,k) = \beta \log \hat{\tau}_{d,t}^o + \sum_o \delta_o + \sum_j \alpha_j + \sum_d \lambda_d + \sum_t \eta_t + \epsilon_{d,t}^o(j,k) \qquad (19)$$

where $\text{Sales}_{d,t}^o(j,k)$ is the value of sales of product $j$ in year $t$ consumed in district $d$ and produced by a monopolist located in district $o$, $\hat{\tau}_d^o$ is the predicted iceberg transportation cost between districts $o$ and $d$ (obtained from equation (17)), $\delta_o$ is a set of district of origin fixed effects, $\alpha_j$ is a set of product fixed effects, $\lambda_d$ is a set of district of destination fixed effects, $\eta_t$ is a set of year fixed effects, and $\epsilon_{d,t}^o(j,k)$ is the error term. The origin fixed effect controls for local wages. The product fixed effect controls for firm productivity. The destination fixed effect controls for market size and aggregate prices at the destination.

Table II presents the results of estimating equation (19). We find that higher transportation costs are associated with lower trade flows at statistically significant levels. The empirical specification indicates that transportation costs which increase by 10 percent are associated with an 8.4 percent decrease in trade flows. This relationship implies that the value of $\theta$ is 1.84.

## TABLE II
### GRAVITY EQUATIONS FOR MONOPOLISTS

|  | (1) |
|---|---|
| Dep. Variable: Log value of sales at destination | |
| $\hat{\tau}_d^o$ | -0.840** |
|  | (0.401) |
| District of Origin Fixed Effects | YES |
| District of Destination Fixed Effects | YES |
| Product Fixed Effects | YES |
| Year Fixed Effects | YES |
| Observations | 2,235 |
| R-squared | 0.538 |

Table II shows the estimates of equation (19). The dependent variable is the log value of sales at destination of products manufactured by monopolists. The variable of interest is the predicted values of equation (15), namely the predicted transport costs across districts. Origin, destination, product and year fixed effects are included. Product fixed effects correspond to 5-digit ASICC products. Robust standard errors are in parenthesis. Significance levels: *: 10%; **: 5%; ***: 1%.

## 6.3 Estimating the Within-Sector Elasticity of Substitution ($\gamma$)

We now estimate the within-sector elasticity of substitution. To do so, we derive the following condition from the model between a firm's labor share and its sectoral share for a given destination:

$$\frac{W_o l_d^o(j,k)}{\tilde{p}_d^o(j,k) c_d^o(j,k)} = 1 - \frac{1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right) \omega_d^o(j,k) \qquad (20)$$

where $\tilde{p}_d^o(j,k)$ is the factory gate price of the good. This condition implies that firms with a higher sectoral share at a destination have a lower labor share. The reason is that firms with higher sectoral shares charge higher markups, which result in lower labor shares. See Section (D) of the Online Appendix for details.

In the data, we do not observe the market share of any given firm by destination. However, a similar condition can be derived for goods that are only produced in one state. In these sectors, the market shares of firms are constant across destinations.

We find that approximately 15% of sectors operate only in one state. These sectors comprise 30,000 firms. Using data from these firms, we estimate equation (20) as follows:

$$LS_o(j,k) = \beta \omega^o(j,k) + \sum_o \delta_o + \sum_j \alpha_j + \epsilon^o(j,k) \qquad (21)$$

where $LS_o(j,k)$ and $\omega^o(j,k)$ are the labor and sectoral shares respectively in state $o$, $\delta_o$ is a set of fixed effects to control for the state where the firm operates, $\alpha_j$ is a set of product fixed effects, and $\epsilon^o(j,k)$ is the error term.

### TABLE III
#### LABOR SHARES VS SECTORAL SHARES

|  | (1) Labor Share | (2) Cap+Labor Share |
|---|---|---|
| Sectoral Share | -0.416*** | -0.493*** |
|  | (0.094) | (0.105) |
| Constant | 0.707*** | 0.807*** |
|  | (0.042) | (0.052) |
| Sector FE | YES | YES |
| Year FE | YES | YES |
| Observations | 1,181 | 1,009 |
| R-squared | 0.870 | 0.893 |

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

Column (1) of table III shows an OLS regression of firms' labor shares against sectoral shares for sectors that are operated only in one state. Column (2) shows the same regression but including also capital remuneration on the left hand side. Product fixed effects correspond to 5-digit Indian sectoral codes (ASICC). Robust standard errors are in parenthesis: *: 10%; **: 5%; ***: 1%.

We present the results in Table III. Column (1) shows the results when including only labor remuneration on the left-hand side of the equation. In column (2), we also include capital remuneration on the left-hand side of the equation. The second specification controls for across-firm

18

variations in capital intensity. We choose this second specification as our preferred one. An OLS slope coefficient of -0.49 together with an across-sector elasticity of substitution $\theta$ of 1.84 implies a value of $\gamma$ of 19.77.

## 6.4 Aggregating Transportation Costs to the State Level

In order to exploit all the variations that exist in the data, we use district-level data in the estimates of transportation costs, $\theta$, and $\gamma$. It is necessary to aggregate the district-to-district transportation costs to state-to-state transportation costs since the model that we simulate is based on interstate trade. We do so in two steps. In the first step, for every district we find the average transportation costs to the districts located in a given destination state. This average is weighted by the population of the destination districts. This yields a measure of district-to-state transportation costs. In the second step, we aggregate the district-to-state measure to obtain state-to-state transportation costs. To do so, we find the average transportation costs from the origin districts of the origin state to a given destination state. This average is weighted by the population of the origin districts.

Given this new set of transportation costs, we repeat the exercise above in which we map the transportation costs from the National Capital Territory of Delhi to all of the states in India. Panel C of Figure II shows a map of these transportation costs. The pattern of faraway states having higher transportation costs that we observed at the district level is also visible in this figure. Panel D of Figure II represents transportation costs in 2006. The fact that the colors are lighter means that there was a decline in transportation costs to most regions.

Importantly, there is a high variation in the decline of transportation costs across locations. As an illustration, Figure III shows the percentage decline in transportation costs from Delhi. As in the previous figure, the colors of the states represent the quartile in terms of decline in transportation costs. States in the top quartile tend to be far from Delhi and close to the GQ upgrades. The states in the top quartile underwent a decline of 3.9 to 8.7%. The states with the smallest decline in transportation costs are those far from the GQ upgrades. For example, this is the case in the northern state of Jammu and Kashmir and in the northeastern states of Arunachal Pradesh, Assam, Manipur, Tripura, and Mizoram. The percentage decline in transportation costs for the bottom quartile ranges from 0 to 0.59%.

## 6.5 Calibrating the Remaining Parameters

**Labor endowment** For the labor endowments of each state, $L_n$, we first normalize the labor endowment of the smallest state to 1. We then set the labor endowments of the remaining states so that the model matches the ratio of manufacturing value-added observed in the data across states.

## TABLE IV

### 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 | 1.84 |
| $\gamma$ | Elasticity of substitution within sector | 19.77 |
| (B) Parameters taken directly from data | | |
| $K_{ij}$ | Number of firms operating in sector $j$ of country $i$ | varies by state |
| (C) Parameters calibrated in equilibrium | | |
| $L_i$ | Labor endowment of the states | varies by state |
| $\alpha$ | Shape parameter Pareto | 2.55 |

Notes: Table IV refers to a our benchmark calibration. We explain how we estimate the parameters $\tau_d^o$, $\theta$, and $\gamma$, in sections 6.1, 6.2, and 6.3 respectively. We set the value $K_{ij}$ to match the number of plants observed in the data. We calibrate $L_n$ to the relative manufacturing value added across states $n$. We calibrate $\alpha$ to to match the fact that the top 5% of plants in manufacturing account for 89% of value-added in this sector (see section 6.5 for details).

**Parameters that govern within-industry productivity across states**  We will now calibrate the parameters that relate to the number of firms that operate and the productivity distribution. These parameters are crucial for the size of the Ricardian and pro-competitive effects of reducing transportation costs.

The number of firms in sector $j$ of state $n$, $K_{nj}$, is set to match the number of plants observed in the data.[18] Since there is no operating cost in the model, all firms operate and there is no entry and exit of firms even after changes in transportation costs. Abstracting from firm entry and exit in these kinds of models does not quantitatively affect the final results, as discussed by Atkeson and Burstein (2008) and Edmond, Midrigan, and Xu (2014). The reason is that a reduction in transportation costs will lead to the entry and exit of low-productivity firms, which do not significantly affect the markups that large firms charge. Furthermore, the data does not show a significant change in terms of firms across sectors in each state. For example, the auto-correlation of the number of producers per sector-state between 2001 and 2006 is 0.98.[19]

---

[18]In order to reduce the computational burden, we limit the number of firms operating in each sector of a given state to 50. This means that we set the maximum number of producers in a given sector to 1,450 (29x50).

[19]The number of active sectors across states remained stable over this period. The change in the percentage of active sectors within states is around 3% on average. The total number of firms did not vary significantly either. The percentage change in the total number of firms within states is around 2% on average.

The distribution of the number of firms across state-sectors is important in determining the nature of gains from lower transportation costs. As a simple example to illustrate this idea, consider a two-state example. Suppose that these two states go from autarky to trading with each other. If there is no overlap in the sectors that these two states produce in, the effects from trade will be purely Ricardian. This is true since trading with another state will not change markups. However, if two states produce very similar goods, then there is room for pro-competitive effects from trade. Note that this can greatly limit the scope of pro-competitive gains if states produce different types of goods in the calibrated data.

Another important factor to consider is the correlation of productivity draws across regions. The correlation determines the extent to which local firms with market power face new competition when the economy opens to trade. Edmond, Midrigan, and Xu (2014) show that the correlation in productivity draws is important to determine the size of pro-competitive gains from trade. In a situation in which productivity draws across states are independent, the pro-competitive gains from trade are zero or even negative. Furthermore, they infer a very high correlation (0.93) for the Taiwanese case.

We assume that firms across states, conditional on operating in a given sector, have perfectly correlated draws in our benchmark calibration. To generate productivity draws for an industry, we generate productivity draws equal to the maximum number of firms present in any state. We sort the productivities in descending order. If a state has one firm, then we select the first productivity on the sorted list. If a state has ten firms, then we select the first ten productivities on the sorted list. This setup ensures that the firms with the highest productivity face high levels of head-to-head competition. Importantly, note that this does not imply that the sectors are symmetric across states. The reason is that states have a different number of firms operating in a given sector. Furthermore, locations have different wages and transportation costs, which affect the marginal cost that firms face.

It is important to determine whether the model generates reasonable levels of head-to-head competition given the assumption of perfectly correlated productivity draws. It is especially important to match the level of head-to-head competition facing the largest firms since they are the ones that drive the dispersion in markups as we will show in Section 7.

We create an index that measures the similarity in size of the largest firms in India. This similarity index measures the degree to which firms face head-to-head competition since the size of firms is related to their productivity. To calculate the index, for each sector and state we identify the firm with the largest value-added. Then, we regress the log of the value-added of these firms on sector dummies. The R squared of that regression, which we use as our index, indicates the extent to which large firms in each sector are of similar size. For example, an R squared of one indicates that the largest firms in each state are exactly the same size. We calculate the index in the data as well as in the data generated by the model and find an index of 0.44 in the data

and 0.37 in the model. This exercise indicates that the model generates levels of head-to-head competition that are in line with the data.

Finally, we use a Pareto distribution for the productivity draws. The tail parameter, $\alpha$, is calibrated in equilibrium to match the fact that the top 5% of firms in manufacturing value-added account for 89% of value-added in this sector.

## 6.6 Discussion of the Distribution of Markups

Table IX summarizes the distribution of markups by destination market. The average markup implied by the model is approximately 1.06 across all markets. The median markup is also very similar. This result comes from the fact that the vast majority of firms have small market shares and thus charge markups close to $\gamma/(\gamma-1)$, which is the lowest possible markup. Markups do not become significantly larger until the 99th percentile of the distribution, which ranges from 1.23 to 1.31 across states. It is important to note that the model predicts a convex relationship between sectoral shares and markups. The few firms that get very high productivity draws (given the Pareto distribution of productivity draws) are those that have nontrivial sectoral shares and hence can charge high markups.

This dimension of the distribution of markups implied by the model is consistent with the distribution of markups estimated in the IO literature. These papers find that most firms have small markups and a handful of firms have large markups. For example, De Loecker, Goldberg, Khandelwal, and Pavcnik (2014) estimate the markups of medium and large Indian firms using the Prowess data set. They find a median markup of 1.18 and an average of 2.24 across sectors. Loecker and Warzynski (2012) estimate the markups of Slovenian firms and find a median markup of 1.17-1.28 across sectors. They also find a large standard deviation of markups of 0.50 across all specifications.

However, the markups implied by our model are lower than those estimated using empirical methods, especially on the high end of the distribution. As mentioned above, the 99th percentile of markups implied by the model ranges from 1.24 to 1.32 across states. Furthermore, the highest possible markup that firms can charge is $\theta/(\theta-1)$, which in our case is 2.20. These markups are lower than those estimated by De Loecker, Goldberg, Khandelwal, and Pavcnik (2014). They find that the average is 2.24, which is above the maximum markups the firms can charge in the model.

It is also interesting to note that firms which account for a large sectoral share are usually far from being monopolists. For instance, for the destination market of Aruchanal Pradesh, the average maximum sectoral share across industries is 35%. The equivalent number is higher in bigger states, 49% in Maharasthra for instance. This distribution of market shares predicted by the model is consistent with other empirical work. Hottman, Redding, and Weinstein (2014) use the Nielsen HomeScan database for the US. They find that half of all output in a product group is produced by five firms on average. Furthermore, they find that 98% of firms have market shares

of less than 2 percent.[20]

# 7 Quantifying the Impact of the GQ

In this section, we quantify the aggregate and state-level effects of the construction of the GQ. To this end, we compare the outcomes from our calibrated model in 2006 with the outcomes when we remove the GQ. To remove the GQ, we use the estimates from Section 6.1 to determine the changes in transportation costs. For illustrative purposes, we present all the results as changes from before to after the construction of the GQ (2001 to 2006). Lastly, we use a difference-in-difference strategy to estimate the decline in prices for districts close to the GQ compared to those that are further away. We compare these results with the predictions of the model.

## 7.1 Simulating the Construction of the GQ

In order to quantify the effects of the GQ, we begin with our baseline calibration described in Table IV. We change the transportation costs to reflect the absence of the GQ. To do so, we change the cost to travel on roads that were upgraded by the GQ as described in equation (16). Given these new costs, we re-compute the shortest path using Dijkstra's algorithm. Finally, we re-aggregate the district-to-district transportation costs to state-to-state transportation costs as described in Section 6.4.

**Benefits from the GQ** First, we consider the aggregate change in real income resulting from the GQ. Table V shows that real income increases by 2.05% for India. Changes in aggregate real income are calculated as the mean percentage change of all states weighted by real income. The increase in real income is in terms of manufacturing value-added, since this is the only sector considered in our model. The value-added of the manufacturing sector was $152.8 billion in 2006. This implies that the static benefit of the construction of the GQ is $3.1 billion. These are the benefits that accrue to India *each year* as a result of the construction of the GQ. We can compare these benefits to the cost of the construction of the GQ. Estimates indicate that the government spent approximately $5.6 billion on the GQ. Thus, the benefits over a two-year period exceed the initial construction costs.

**A framework to decompose the Ricardian and pro-competitive effects of the GQ** We apply the framework developed by Holmes, Hsu, and Lee (Forthcoming) to decompose the changes in real income in a way that highlights the various mechanisms at work in the model. The

---

[20]This comparison has to be taken with caution for a main reason. They look at much more narrowly defined products. While our definition of sector is based on the 5-digit Annual Survey of Industries Commodities Classification (ASICC), they use 12-digit Universal Product Codes (UPCs).

framework allows us in particular to distinguish between Ricardian, pro-competitive, and terms of trade effects from lowering transportation costs.

We now introduce some notations for the purpose of the decomposition. We define the aggregate markups on the goods sold. This reflects how much market power firms producing in a state have when selling to other states. First, the revenue-weighted mean labor cost share for the products sold by state $n$ is:

$$c_n^{sell} = \int_0^1 \left( \sum_{d=1}^N \sum_{k=1}^{K_{nj}} c_d^n(j,k) s_d^n(j,k) \right) dj,$$

where $s_d^n(j,k)$ is the share of state $n$'s revenue that comes from goods produced by firm $k$ in sector $j$ and sold in state $d$. The aggregate markup on the goods sold can be expressed:

$$\mu_n^{sell} = \frac{R_n}{W_n L_n} = \frac{1}{c_n^{sell}},$$

where $R_n = W_n L_n + \Pi_n$, which is the state's total revenue. Note that there is an analogous expression at the firm level which is that the firm's markup is equal to the reciprocal of the labor share.

We next define the aggregate markups on the goods purchased by state $n$, which reflect how much market power firms located in other states have when selling to state $n$. The revenue-weighted mean labor cost for the products purchased by state $n$ is:

$$c_n^{buy} = \int_0^1 \left( \sum_{o=1}^N \sum_{k=1}^{K_{oj}} c_n^o(j,k) b_n^o(j,k) \right) dj.$$

where $b_n^o(j,k)$ is the share of expenditures in state $n$ on goods produced by firm $k$ in sector $j$ located in state $o$. The aggregate markups on the goods purchased are:

$$\mu_n^{buy} = \frac{1}{c_n^{buy}}.$$

Lastly, let $P_n^{pc}$ be the aggregate price of state $n$ if every firm engages in marginal cost pricing. $P_n^{pc}$ is the aggregate price index that would emerge in a context of perfect competition. This price index depends on the factors that determine the marginal cost of firms: the distribution of firm productivity, the wages paid by firms, and the transportation costs that these firms face.

Using this notation, the real income in state $n$ can be rewritten into the following components:

$$Y_n = \underbrace{W_n L_n}_{\text{Labor income}} * \underbrace{\frac{1}{P_n^{pc}}}_{\text{Prod. efficiency}} * \underbrace{\frac{\mu_n^{sell}}{\mu_n^{buy}}}_{\text{Markup ToT}} * \underbrace{\frac{P_n^{pc}}{P_n} \mu_n^{buy}}_{\text{Allocative efficiency}} \tag{22}$$

The first component is the aggregate *labor income*. The second component is the *productive efficiency* component of welfare. This component is simply the inverse of the price index if all firms charge the marginal cost. The third component is the *terms of trade*. This component

24

compares the aggregate markups charged for the goods a state sells with those that it purchases. The last component is *allocative efficiency*. It can be shown that this term is equal to the cost of one unit of utility under marginal cost pricing divided by the cost of acquiring one unit of utility with the equilibrium bundle under marginal cost pricing. In a situation with no misallocation, i.e. no variations on markups across firms, this index is equal to one. As misallocation increases, this index decreases.

Combining the first two terms leads to an expression that is equal to real income if firms charge the marginal cost. This expression maps back to welfare in the large class of models considered by Arkolakis, Costinot, and Rodriguez-Clare (2012), in which the markups of firms remain unchanged. Thus, we consider changes in this component to be Ricardian effects. We consider changes in the allocative efficiency to be pro-competitive effects as this directly maps back to the welfare losses due to dispersion in markups. Given the expression in equation (22), we decompose the changes in real income into the following terms:

$$\Delta \ln Y_n \quad = \underbrace{\Delta \ln W_n L_n + \Delta \ln \frac{1}{P_n^{pc}}}_{\text{Ricardian}} \quad + \quad \underbrace{\Delta \ln \frac{P_n^{pc}}{P_n} \mu_n^{buy}}_{\text{Pro - competitive}} \quad + \quad \underbrace{\Delta \ln \frac{\mu_n^{sell}}{\mu_n^{buy}}}_{\text{Markup ToT}}$$

**Quantifying the decomposition**   Table V shows these three components at the aggregate and state level. We find that, for India as a whole, the pro-competitive component accounts for approximately 17% of the aggregate gains (0.35% of the 2.05% total gains). The pro-competitive component can be up to 26% of the gains at the state level (0.41% of the 1.60% of the gains for Maharashtra).

The welfare effects of the GQ are very heterogeneous across states. Overall, large states gain more from the reduction in transportation costs. Small states see modest gains and in some cases even lose. This is driven by the fact that, due to its placement, the GQ has lowered transportation costs primarily for large states. Many of the small states are located in northeastern India, which is far from the GQ.[21] The states in the Northeast that gained less than 1% include: Assam and Meghalaya. The states in the Northeast that experienced losses include: Arunachal Pradesh, Manipur, Mizoram, Nagaland, and Tripura. Figure IV shows a map of the welfare effects across states, including the states that lost.

First, we turn to the Ricardian components across states. These terms are generally positive and large across all states. This term also explains the modest or negative effects for the states in the Northeast. The only two factors that affect a firm's marginal cost to serve a destination are the transportation costs that it faces and local wages. First, we know that the GQ lowers transportation costs for some destinations and leaves the transportation costs for others unchanged. Thus,

---

[21]Northeastern Indian states include: Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura.

## Table V

### Quantitative Results

| state | size | income change | Descomposition | | | |
|---|---|---|---|---|---|---|
| | | | $\eta_w$ | $\eta_{PE}$ | $\eta_{ToT}$ | $\eta_{ae}$ |
| India | | 2.05 | 1.70 | -0.00 | 0.00 | 0.35 |
| Maharashtra | 100.00 | 1.60 | 1.53 | -0.50 | 0.16 | 0.41 |
| Gujarat | 65.12 | 2.70 | 2.18 | -0.05 | 0.08 | 0.49 |
| Tamil Nadu | 41.72 | 1.35 | 1.43 | -0.36 | -0.01 | 0.29 |
| Uttar Pradesh | 29.89 | 2.23 | 1.75 | 0.10 | -0.06 | 0.45 |
| Karnataka | 26.95 | 2.91 | 2.23 | 0.22 | -0.07 | 0.54 |
| Andhra Pradesh | 21.26 | 1.81 | 1.46 | 0.09 | -0.10 | 0.35 |
| West Bengal | 18.91 | 3.70 | 2.62 | 0.67 | -0.12 | 0.53 |
| Haryana | 18.81 | 1.03 | 1.18 | -0.21 | -0.15 | 0.21 |
| Jharkhand | 17.15 | 5.01 | 3.10 | 1.66 | 0.01 | 0.24 |
| Rajasthan | 12.63 | 2.89 | 2.07 | 0.81 | -0.25 | 0.26 |
| Madhya Pradesh | 11.32 | 1.40 | 1.19 | 0.17 | -0.13 | 0.18 |
| Orissa | 10.78 | 2.33 | 1.71 | 0.41 | -0.02 | 0.23 |
| Punjab | 10.13 | 0.44 | 0.58 | -0.30 | 0.05 | 0.10 |
| Himachal Pradesh | 9.47 | 0.32 | 0.71 | -0.31 | -0.17 | 0.09 |
| Chattisgarh | 9.29 | 0.30 | 0.69 | -0.44 | -0.05 | 0.11 |
| Kerala | 7.47 | 0.79 | 0.72 | -0.20 | 0.06 | 0.20 |
| Uttaranchal | 4.57 | 0.78 | 0.77 | -0.18 | 0.03 | 0.15 |
| Delhi | 4.54 | 1.01 | 0.78 | -0.13 | 0.15 | 0.21 |
| Assam | 3.75 | 0.30 | 0.52 | -0.32 | 0.01 | 0.09 |
| Goa | 3.51 | 7.82 | 4.49 | 3.86 | -0.58 | 0.06 |
| Bihar | 2.73 | 5.38 | 3.16 | 2.50 | -0.35 | 0.07 |
| Jammu and Kashmir | 2.59 | -0.37 | 0.13 | -0.38 | -0.09 | -0.03 |
| Meghalaya | 0.63 | 0.55 | 0.65 | -0.26 | 0.03 | 0.13 |
| Tripura | 0.26 | -1.58 | -0.57 | -1.25 | 0.21 | 0.03 |
| Manipur | 0.11 | -1.43 | -0.45 | -1.31 | 0.28 | 0.05 |
| Nagaland | 0.07 | -0.99 | -0.24 | -1.06 | 0.24 | 0.06 |
| Sikkim | 0.03 | 3.98 | 2.08 | 1.27 | 0.41 | 0.22 |
| Mizoram | 0.01 | -0.91 | 0.27 | -1.28 | 0.07 | 0.04 |
| Arunachal Pradesh | 0.01 | -1.13 | 0.00 | -1.41 | 0.24 | 0.04 |

Table V shows the % change in real income and the decomposition of the Holmes, Hsu, and Lee (Forthcoming) index for the 29 Indian states; $\eta_w$ represents the % change in labor income component of the index; $\eta_{PE}$ represents the % change in the *Prod. efficiency* component; $\eta_{ToT}$ represents the % change in the *terms of trade* component; and $\eta_{ae}$ represents the % change in the *allocative efficiency* component.

reductions in transportation costs increase the productive efficiency component. Transportation costs declined relatively more in states closer to the GQ, and hence these states benefited from a

higher increase in productive efficiency. These states started trading more intensively with each other, which increased their degree of openness and relative wages. The fact that the Ricardian term is negative for the Northeastern states implies that the effect of wages in the states close to the GQ outweighed the benefits of the GQ in terms of lower transportation costs. We discuss this *"trade diversion"* in detail in Section 7.2.

Next, we examine the pro-competitive and terms of trade effects across states. These two terms are the result of the variable markup feature of the model. First, we find that the pro-competitive effects are positive across all but one state. This means that lower transportation costs generally led to welfare-enhancing changes in markups. As mentioned before, theory is ambiguous as to whether declines in transportation costs lead to gains in allocative efficiency. The range of gains from changes in allocative efficiency is -0.03 to 0.54%. Larger states also gained the most in terms of allocative efficiency.

In order to understand the changes in allocative efficiency and terms of trade, it is necessary to think about the asymmetry of states. It is necessary in particular to think about the markups charged by firms based on their location. As can be seen in Table VIII, there are significant variations in wages across states. Large states tend to have lower wages, which means that, everything else equal, firms located in those states are relatively competitive. Conditional on the productivity draws, the competitiveness of firms located in state $o$ serving market $d$ is $(W_o \tau_d^o)^{-1}$, which is directly related to the marginal cost of serving that market from state $o$. Thus, firms located in large states have more market power than firms located in small states. This holds true even though firms have exactly the same productivity draws.

Figure V shows the distribution of the location of the firms whose markups on goods purchased in Arunachal Pradesh and Maharashtra are in the top 1%. We see that in both cases, these firms are primarily located in Maharashtra and other big states, where wages tend to be lower. For instance, in Maharashtra, around 58% of the firms charging the 1% biggest markups are domestic firms. It is important to note that, in Arunachal Pradesh, the competitiveness of firms located in Maharashtra decreases due to transportation costs. This implies that, for goods sold in Arunachal Pradesh, only around 23% of the firms charging the highest markups are from Maharashtra.

Thus, under our benchmark calibration, any force that increases the competitiveness of firms in large states will *decrease* allocative efficiency. This is the case because increasing the competitiveness of firms in large states will increase the market power of firms that already charge high markups. Likewise, any force that decreases the competitiveness of firms in large states will *increase* allocative efficiency.

Now we will examine changes in allocative efficiency for the states in the Northeast. The lower transportation costs make firms located in large states more competitive in the Northeast, lowering allocative efficiency. On the other hand, Table V shows that large states had the biggest wage increases. Higher wages lower the competitiveness of firms in large states, raising allocative

efficiency. In this case, the effect of rising wages is stronger than the decline in transportation costs. For example, the competitiveness of firms located in Maharashtra serving Arunachal Pradesh decreases by 1.44%. This loss in competitiveness is reflected in the changes in markups of those firms with strong market power. Table IX shows the changes in markups charged by destination market. We see that for Arunachal Pradesh, the markups of the 99th percentile decreases by 0.05%. This fact also explains why small states experienced significant increases in terms of trade. Note that the re-adjustment of wages is critical in thinking about changes in misallocation. If wages did not adjust, we would see *increases* in misallocation in the Northeast.

For large states, we know that local firms have relatively high markups. Thus, increases in the wages of large states improve allocative efficiency, just as in the case of the Northeastern states. However, a decline in transportation costs has the opposite effect on allocative efficiency. This is because it forces local firms to lower their high markups due to the increased competitiveness of firms located in other states, which have relatively low markups. Thus, the forces of changing wages and transportation costs work towards increasing allocative efficiency in large states. Large states experienced the biggest increase in allocative efficiency. Furthermore, large states generally experienced a decline in terms of trade.

These findings highlight the importance of considering asymmetric economies in models of variable markups. We find in particular that including variable markups in the model generates gains that mitigate the negative Ricardian terms for the states in the Northeast. For example, the state of Manipur gained 0.33% due to changes in markups. This helped reduce the loss of -1.76% from the Ricardian term.

It is interesting to note that there is a wide dispersion in the effects of the terms of trade component. For example, Goa lost 0.58%, while Sikkim gained 0.41%. Thus, although allocative efficiency improves for all but one state, the changes in the terms of trade can result in some states suffering losses due to changing markups. For example, Goa lost more from the changing terms of trade than it did from the improved allocative efficiency. Thus, although almost all states gain from increases in allocative efficiency, the changes in markups lead to a significant re-shuffling of income across states through changes in the terms of trade.

Figure VI highlights the main pattern of the change in markups after the reduction in transportation costs. Changes in markups are only quantitatively relevant for firms which charged markups above the 90th percentile of the (before) distribution of markups. Interestingly, the 90th and 95th percentiles of the markup distribution actually increase. More importantly, Figure VI shows that the 1% firms in the distribution substantially decrease their markups. The fact that reductions in markups after policy reforms are more pronounced among firms with initially high markups has already been documented in the literature. For instance, De Loecker, Goldberg, Khandelwal, and Pavcnik (2014) find that, for high-markup products (above the 90th percentile), the same reduction in tariffs results in an additional 4.40 percent decline in markups.

It is also important to measure the importance of misallocation generated by the model under our benchmark calibration. The allocative efficiency component ranges from 0.937 to 0.968, meaning that real income would increase by 3.2-6.3% under constant markups across firms. Furthermore, we find that larger states consistently have more misallocation than smaller states. Overall, the levels of misallocation that the model generates are low compared to those found by Hsieh and Klenow (2009).

## Table VI

### % Change in Total Trade between $i$ and $j$ (Model)

|     |                                  | mean | median | sd/mean | N   |
|-----|----------------------------------|------|--------|---------|-----|
|     | both $i$ and $j$ in GQ           | 4.91 | 4.28   | 1.63    | 78  |
| (A) | either $i$ or $j$ in GQ          | 1.04 | -1.16  | 9.35    | 208 |
|     | neither $i$ nor $j$ in GQ        | 2.60 | 2.37   | 4.21    | 120 |
|     | both $i$ and $j$ in (N-EAST+JM)  | 3.41 | 2.45   | 2.64    | 190 |
| (B) | either $i$ or $j$ in (N-EAST+JM) | 0.85 | -0.10  | 13.24   | 180 |
|     | neither $i$ nor $j$ in (N-EAST+JM) | 3.08 | 3.01 | 1.78    | 36  |

Table VI shows the mean, median, and coefficient of variation of the % change in total trade between states $i$ and $j$ after the construction of the GQ; "both $i$ and $j$ in GQ" refers to state pairs in which both of them are crossed by the GQ; "either $i$ or $j$ in GQ" refers to state pairs in which only one of them is crossed by the GQ; "neither $i$ nor $j$ in GQ" refers to state pairs in which none of them are crossed by the GQ; "both $i$ and $j$ in (N-EAST+JM)" refers to state pairs in which both of them belong to the group (N-EAST+JM); "either $i$ or $j$ in (N-EAST+JM)" refers to state pairs in which one of them belong to the group (N-EAST+JM); "neither $i$ nor $j$ in (N-EAST+JM)" refers to state pairs in which none of them belong to the group (N-EAST+JM). N-EAST+JM includes Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura, and Jammu and Kashmir. N is the number of state pairs that fall into the six different categories.

## 7.2 The GQ and Trade Diversion and Creation within India

We now study changes in state-to-state trade patterns induced by the construction of the GQ. The fact that reductions in transportation costs are not uniformly distributed across states leaves room for trade diversion and creation. To study this possibility, we compute the whole set of bilateral trade flows across Indian states before and after the GQ. We define the total trade between state $i$ and state $j$ as

$$\text{Total Trade}_{i,j} = \text{exports}_j^i + \text{exports}_i^j,$$

where $\text{exports}_j^i$ and $\text{exports}_i^j$ are the total exports from state $i$ to state $j$ and the total exports from state $j$ to state $i$ respectively.

Table VI shows the percentage changes of this variable for different sets of state pairs. In Panel (A), we divide the state pairs according to their access to the GQ.[22] In the first row, we include state pairs in which both of the states are crossed by the GQ. In the second row, we include state

---

[22]This analysis is similar in spirit to papers that study trade diversion such as Krueger (1999) and Bayoumi and Eichengreen (1998).

pairs in which only one of the states is crossed by the GQ. In the third row, we include state pairs in which none of the states are crossed by the GQ. Panel (B) is equivalent to Panel (A), but it classifies the state pairs based on whether they are far from the GQ. This set of states include the northeastern states (Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura), plus Jammu and Kashmir. With this additional classification, we try to capture the fact that, even if some states were not crossed by the GQ, they were relatively close so that their market access increased. Examples of this include Goa and Madhya Pradesh.

We find that, on average, trade increases considerably more between state pairs in which both states are either crossed by the GQ or not crossed by the GQ. In other words, the construction of the GQ resulted in an increase in trade flows between states located far away from it. It also resulted in increases in trade flows between states located close to the GQ. For instance, trade flows between state pairs crossed by the GQ increased on average by 4.91%. For state pairs not crossed by the GQ, the increase in trade was of 2.60%. On the other hand, trade between state pairs in which only one of the states was crossed by the GQ increased much less, 1.04% on average. In fact, the median percentage change in trade between these states is negative: -1.16%.

This means that trade flows in India have become more regionally concentrated. For instance, the trade between northeastern states increased by 2.98%. The pattern is similar for non-northeastern states, where trade increased by 4.91% on average.

## 7.3 The GQ and the Evolution of Prices

The model has an important implication which is behind the differential effect of the construction of the GQ across regions: prices fell more in states crossed by the GQ. We use the time dimension of the data to evaluate the ability of the model to predict the different responses of prices in GQ vs non-GQ locations. In order to do this, we examine the impact of the GQ on prices using a reduced-form approach and we compare it with the outcome of the model.[23] In the model, the prices paid for goods are endogenous and depend on changes in transportation costs, wages, and markups. These changes in wages and markups depend on complex general equilibrium effects. Thus, it is necessary to simulate the model in order to compare the changes in prices paid for goods with the data.

One of the major issues to tackle when studying the impact of transportation infrastructure is the fact that the placement of infrastructure is not random. In the case of India, the GQ was built with the goal of connecting the major urban centers. In order to deal with this identification problem, we use the strategy adopted by several authors such as Atack, Bateman, Haines, and

---

[23]The difference-in-difference strategy we use to identify how the GQ affected prices is similar to the one used in recent work investigating the impact of transportation infrastructure in development. These papers include Atack, Bateman, Haines, and Margo (2010), Banerjee, Duflo, and Qian (2012), and Faber (Forthcoming). This strategy has also been applied to investigate the impact of the GQ by Datta (2012) and Ghani, Goswami, and Kerr (Forthcoming).

Margo (2010) and Banerjee, Duflo, and Qian (2012) who have exploited the fact that the goal of infrastructure projects is usually to connect historical cities or large economic centers. The causal effect of transportation infrastructure is identified by applying a difference-in-difference approach comparing non-nodal areas that differ in their distance to the transportation network before and after the infrastructure was constructed. We follow this approach in order to study the impact of transportation costs on prices, making use of the natural experiment provided by the GQ. We run the following difference-in-difference regression in particular:

$$\Delta \log P_{jd} = \sum_j \alpha_j + \beta_1 \Delta \mathrm{GQ}_d + \sum_s \delta_s + \epsilon_{jd}, \tag{23}$$

where $P_{jd}$ is the price of input $j$ in district $d$ between 2001 and 2006, and $\mathrm{GQ}_d$ is a dummy variable taking the value 1 if district $d$ is within a certain distance of the GQ, and $\epsilon_{jd}$ is an error term. Distance is calculated as the shortest straight-line distance between the district and a treated portion of the GQ. Thus, $\Delta \mathrm{GQ}_d = 1$ if a district was within the specified distance of a treated portion of the GQ in 2006 and not in 2001. We use the following categories for distance: 15, 50, 100, 150, 200, and 300 km. 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 clustered at the district level in order to account for the possible serial correlation of price shocks within districts.

The estimates of equation (23) can be found in Table VII. We find that districts located within 15 km and 50 km of the GQ in 2006 experienced a statistically significant decline in input prices. For districts located within 15 km, input prices were 33 percentage points lower relative to districts located further away from the GQ. The first coefficient of column (1) includes nodal districts and column (2) excludes nodal districts. For districts within 50 km of the GQ in 2006, we find an even stronger effect, a decrease of 36 percentage points in input prices relative to districts further away. Extending the treatment group beyond 50 km makes the effect disappear. The evolution of input prices was not significantly different between districts within 100 km of the GQ and those beyond 100 km. This implies that prices in "GQ" districts decreased 1.57 times as much as in the average district. The model predicts a stronger effect. The decrease in prices charged in states through which the GQ passes are 1.96 times bigger than in the average state. Although we find that this evidence supports the differential evolution in prices predicted by the model, these comparisons have to be taken with caution for several reasons. First, while we calibrate the model at the state level, we carry out the difference-in-difference exercise at the district level. We do this in order to exploit the variation in the data as much as possible. Second, we are looking at a short time span in the data (2001-2006). However, one could think that it would take longer for the Indian economy to converge to the new equilibrium after the construction of the GQ.

TABLE VII

PRICES AND THE GOLDEN QUADRILATERAL: DIFFERENCES-IN-DIFFERENCES

|  | (1) | (2) |
|---|---|---|
| Dep. Variable: Log change in input prices between 2001 and 2006 | | |
| District within 15 km from GQ | -0.3219** | -0.3293** |
|  | (0.1395) | (0.1406) |
| District within 50 km from GQ | -0.3484** | -0.3604*** |
|  | (0.1363) | (0.1367) |
| District within 100 km from GQ | -0.2036 | -0.2171 |
|  | (0.1659) | (0.1697) |
| District within 150 km from GQ | 0.0711 | 0.0768 |
|  | (0.1357) | (0.1416) |
| District within 200 km from GQ | 0.0916 | 0.0973 |
|  | (0.1591) | (0.1767) |
| Input fixed-effects | YES | YES |
| State fixed-effects | YES | YES |
| Nodal districts | YES | NO |
| Observations | 5,123 | 5,037 |
| Average R-squared | 0.44 | 0.44 |
| Number of products | 929 | 912 |

Table VII shows the estimation of equation (23). The dependent variable is the log change in the price of input $j$ between 2001 and 2006 in district $d$. The variable of interest is the connectivity of the district, defined as whether the district is within a certain distance from the GQ in 2006 and 2001. Each row correspons to a different regression, where different distances are considered. The treatment variable at distance x takes value 1 if district $d$ is within x km from the GQ in 2006 and was not in 2001, and zero otherwise. Columns (1) includes all districts whereas column (2) excludes nodal districts. Input and state fixed effects are included in all specifications. Robust standard errors are in parenthesis, clusterd at the district level. Significance levels: *: 10%; **: 5%; ***: 1%.

## 8 Sensitivity Analysis

We now examine the sensitivity of our results by considering versions of our model in which we change the value of some of the crucial parameters. We first examine the implications of setting a lower value for the elasticity of substitution within sectors, $\gamma$. We next look at the effects of decreasing the value of the elasticity of substitution across sectors, $\theta$. Lastly, we study a version of the model in which firms' productivity shocks are uncorrelated across states.

For all these cases, we keep the rest of parameters which we estimate outside the model constant, and re-calibrate those that we calibrate in equilibrium (the labor endowment for each state $i$, $L_i$, and the shape parameter of the Pareto distribution, $\alpha$). Since the top 5% of plants in manufacturing account for 89% of value-added, the model requires a shape parameter of the Pareto distribution $\alpha$ of 2.12 in the case of $\theta = 1.28$, 2.00 in the case of $\gamma = 10$, and 5.75 in the case of uncorrelated draws (vs 2.55 in our benchmark calibration).

Overall, we find that the aggregate gains are remarkably stable across specifications. The share of gains that are pro-competitive remains quantitatively relevant in the case of a lower $\gamma$ and lower

$\theta$. However, pro-competitive gains disappear in the case of uncorrelated productivity draws across states. Section (E) of the Online Appendix provides a further description of the numbers generated by the model in the three different scenarios.

**A lower elasticity of substitution across sectors**  We set $\theta = 1.28$, which is the value estimated by Edmond, Midrigan, and Xu (2014) using Taiwanese data. In this economy, there is more misallocation than in the benchmark economy: the allocative efficiency index ranges from 0.87 to 0.91 across states, whereas in the benchmark calibration it ranged from 0.94 to 0.97. The reason is that the lower $\theta$ implies that firms with large market shares charge higher markups, increasing the dispersion of markups.

In this specification, pro-competitive gains increase to 0.39% (vs 0.35%). Interestingly, the share of pro-competitive gains declines to 16.5% of the gains (vs 17.1%). At the state-level, pro-competitive gains represent up to 26% of the overall gains (vs 26%). However, one has to be cautious when comparing these numbers since some crucial parameters are not constant across the two economies. For instance, the Pareto shape parameter becomes 2.12 (vs 2.55 in the baseline case).

A value of 1.28 for $\theta$ would imply a too low elasticity for monopolists compared to the one we estimate using our Indian data.

**A lower elasticity of substitution within sectors**  We set $\gamma = 10$, which is the value used by Atkeson and Burstein (2008) and Edmond, Midrigan, and Xu (2014). In this specification, the gains from the GQ increase to 2.32% (vs 2.05% in the benchmark). However, pro-competitive gains decline to 0.24% (vs 0.35%) and the share of gains that are pro-competitive drop to 10% (vs 17%). At the state-level, pro-competitive gains represent up to 21% of the gains (vs 26%).

When setting $\gamma = 10$, the model generates a too weak cross-sectional relationship between markups and sectoral shares compared to the one we measure in our data.

**Uncorrelated productivity draws**  Finally, we examine how our results change if we have uncorrelated productivity draws across locations. We find that aggregate gains increase to 2.56% (vs 2.05%). In this economy, pro-competitive gains are zero, which is consistent with the findings of Edmond, Midrigan, and Xu (2014).

When we calculate our similarity index for this economy, we find a value of 0.25. This is lower than the 0.37 we obtained in our baseline calibration, and much lower than the 0.44 that we measure in the data. Thus, the degree of face-to-face competition that firms are confronted with is too low in the case of uncorrelated draws.

# 9 Conclusions

The goal of this paper is to quantitatively evaluate the welfare effects of improving transportation infrastructure in a setting of internal trade and variable markups. Hence, we determine the extent to which misallocation is driven by high transports costs and decompose the welfare effects into Ricardian and pro-competitive gains, and we can thus gauge the distribution of gains across locations. We apply this framework to the construction of the Golden Quadrilateral in India, a major highway project spanning 5,800 km. We find large gains from the infrastructure project, amounting to more than 2% of real income, approximately 1/5 of which come from pro-competitive gains. Nevertheless, there is wide variation in income gains across Indian states, and even some negative effects after the project. These findings highlight the quantitative significance of the aggregate gains from the GQ, the size of pro-competitive gains, as well as the heterogeneous effects across states.

TABLE VIII

Quantitative Results

| state | $w_n$ | | $\frac{1}{P_n^{pc}}$ | | $\frac{\mu_n^{sell}}{\mu_n^{buy}}$ | | $\frac{P_n^{pc}}{P_n}\mu_n^{buy}$ | |
|---|---|---|---|---|---|---|---|---|
| | before | after | before | after | before | after | before | after |
| Maharashtra | 0.3850 | 0.3910 | 0.0097 | 0.0097 | 1.0194 | 1.0211 | 0.9247 | 0.9285 |
| Gujarat | 0.3842 | 0.3927 | 0.0093 | 0.0093 | 1.0105 | 1.0112 | 0.9302 | 0.9348 |
| Tamil Nadu | 0.3949 | 0.4006 | 0.0091 | 0.0091 | 0.9980 | 0.9979 | 0.9354 | 0.9381 |
| Uttar Pradesh | 0.4180 | 0.4254 | 0.0089 | 0.0089 | 1.0022 | 1.0016 | 0.9465 | 0.9508 |
| Karnataka | 0.4143 | 0.4236 | 0.0090 | 0.0090 | 0.9844 | 0.9837 | 0.9458 | 0.9509 |
| Andhra Pradesh | 0.4296 | 0.4359 | 0.0088 | 0.0088 | 0.9938 | 0.9928 | 0.9503 | 0.9536 |
| West Bengal | 0.4221 | 0.4333 | 0.0086 | 0.0087 | 0.9919 | 0.9907 | 0.9451 | 0.9501 |
| Haryana | 0.4230 | 0.4280 | 0.0089 | 0.0089 | 0.9863 | 0.9848 | 0.9559 | 0.9579 |
| Jharkhand | 0.3567 | 0.3680 | 0.0085 | 0.0087 | 0.9928 | 0.9928 | 0.9522 | 0.9545 |
| Rajasthan | 0.4384 | 0.4476 | 0.0087 | 0.0088 | 0.9931 | 0.9906 | 0.9573 | 0.9598 |
| Madhya Pradesh | 0.4328 | 0.4380 | 0.0087 | 0.0087 | 0.9803 | 0.9790 | 0.9568 | 0.9585 |
| Orissa | 0.4047 | 0.4117 | 0.0085 | 0.0085 | 0.9721 | 0.9720 | 0.9537 | 0.9559 |
| Punjab | 0.4527 | 0.4553 | 0.0086 | 0.0086 | 1.0018 | 1.0024 | 0.9580 | 0.9589 |
| Himachal Pradesh | 0.4139 | 0.4168 | 0.0085 | 0.0085 | 0.9865 | 0.9849 | 0.9553 | 0.9562 |
| Chattisgarh | 0.3907 | 0.3934 | 0.0085 | 0.0085 | 0.9648 | 0.9643 | 0.9550 | 0.9560 |
| Kerala | 0.4568 | 0.4602 | 0.0085 | 0.0084 | 0.9967 | 0.9973 | 0.9537 | 0.9556 |
| Uttaranchal | 0.4455 | 0.4489 | 0.0085 | 0.0085 | 0.9967 | 0.9970 | 0.9611 | 0.9625 |
| Delhi | 0.4895 | 0.4934 | 0.0089 | 0.0089 | 0.9797 | 0.9811 | 0.9590 | 0.9610 |
| Assam | 0.4241 | 0.4263 | 0.0079 | 0.0079 | 0.9872 | 0.9873 | 0.9561 | 0.9570 |
| Goa | 0.4471 | 0.4677 | 0.0084 | 0.0087 | 0.9736 | 0.9680 | 0.9575 | 0.9581 |
| Bihar | 0.4467 | 0.4611 | 0.0083 | 0.0085 | 0.9769 | 0.9735 | 0.9591 | 0.9598 |
| Jammu and Kashmir | 0.4555 | 0.4561 | 0.0080 | 0.0080 | 1.0192 | 1.0183 | 0.9614 | 0.9611 |
| Meghalaya | 0.4297 | 0.4325 | 0.0078 | 0.0078 | 0.9242 | 0.9244 | 0.9604 | 0.9616 |
| Tripura | 0.5083 | 0.5054 | 0.0076 | 0.0075 | 1.0456 | 1.0478 | 0.9653 | 0.9656 |
| Manipur | 0.5972 | 0.5945 | 0.0076 | 0.0075 | 1.0902 | 1.0932 | 0.9650 | 0.9654 |
| Nagaland | 0.6121 | 0.6106 | 0.0077 | 0.0076 | 1.0997 | 1.1024 | 0.9638 | 0.9644 |
| Sikkim | 0.6343 | 0.6476 | 0.0079 | 0.0080 | 1.0076 | 1.0117 | 0.9635 | 0.9657 |
| Mizoram | 1.2045 | 1.2077 | 0.0076 | 0.0075 | 1.1280 | 1.1287 | 0.9660 | 0.9664 |
| Arunachal Pradesh | 1.0000 | 1.0000 | 0.0076 | 0.0075 | 1.4402 | 1.4437 | 0.9657 | 0.9661 |

Table V shows the level of the four different components of the Holmes, Hsu, and Lee (Forthcoming) index for the 29 Indian states for before and after the construction of the GQ; $w_n$ is the relative wage (note that we have excluded labor endowment for presentation purposes); $P_n^{pc}$ is the aggregate price index in state $n$ if all firms charged marginal cost; $\mu_n^{buy}$ represents the expenditure-weighted average markup charged on goods purchased in state $n$; $\mu_n^{sell}$ represents the revenue-weighted average markup charged on goods produced in state $n$; $P_n$ is simply the aggregate price index in state $n$.
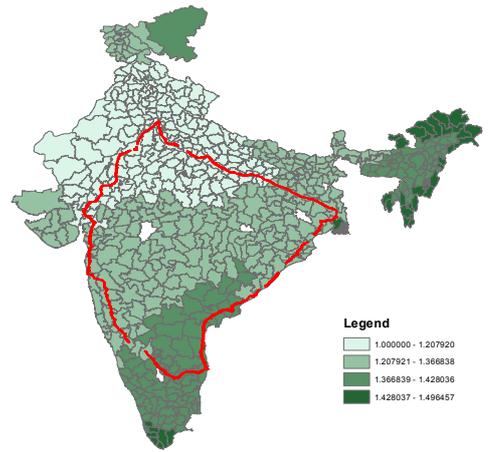
FIGURE II

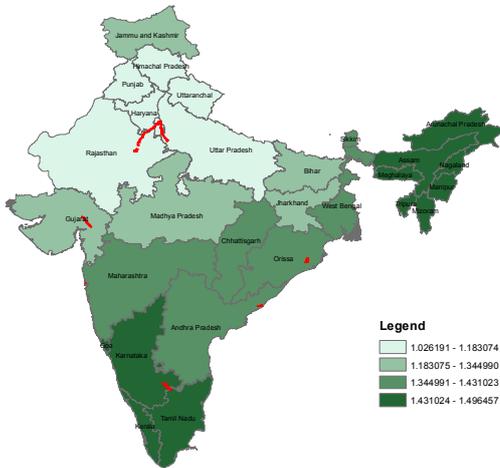## ESTIMATED TRANSPORTATION COSTS FROM DELHI
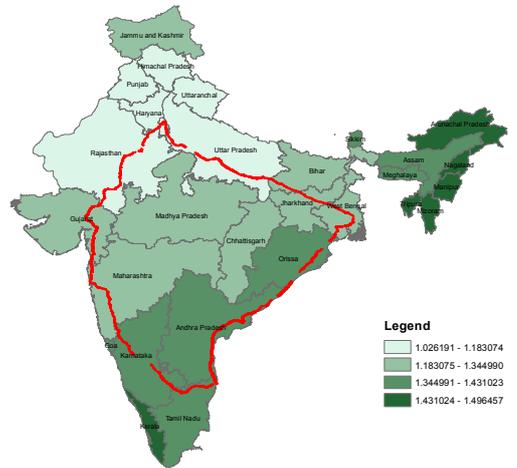
A: 2001 (District level)
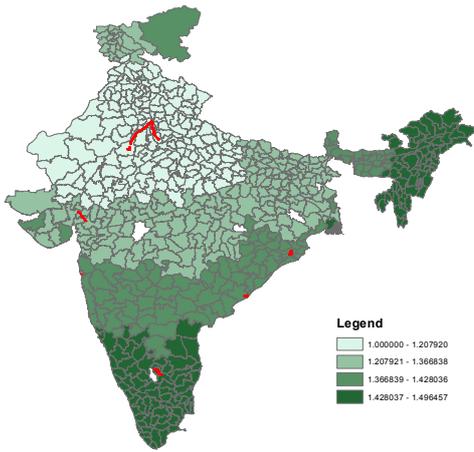
B: 2006 (District level)



C: 2001 (State level)

D: 2006 (State level)



Panel A of Figure II shows the estimated transportation costs from Delhi at the district level for 2001; Panel B of Figure II shows the estimated transportation costs from Haryana at the district level for 2006; Panel C of Figure II shows the estimated transportation costs from Delhi at the state level for 2001; Panel D of Figure II shows the estimated transportation costs from Delhi at the state level for 2006. The transportation costs have been estimated as explained in section 6.1.

FIGURE III

PERCENTAGE CHANGE IN TRANSPORTATION COSTS FROM DELHI
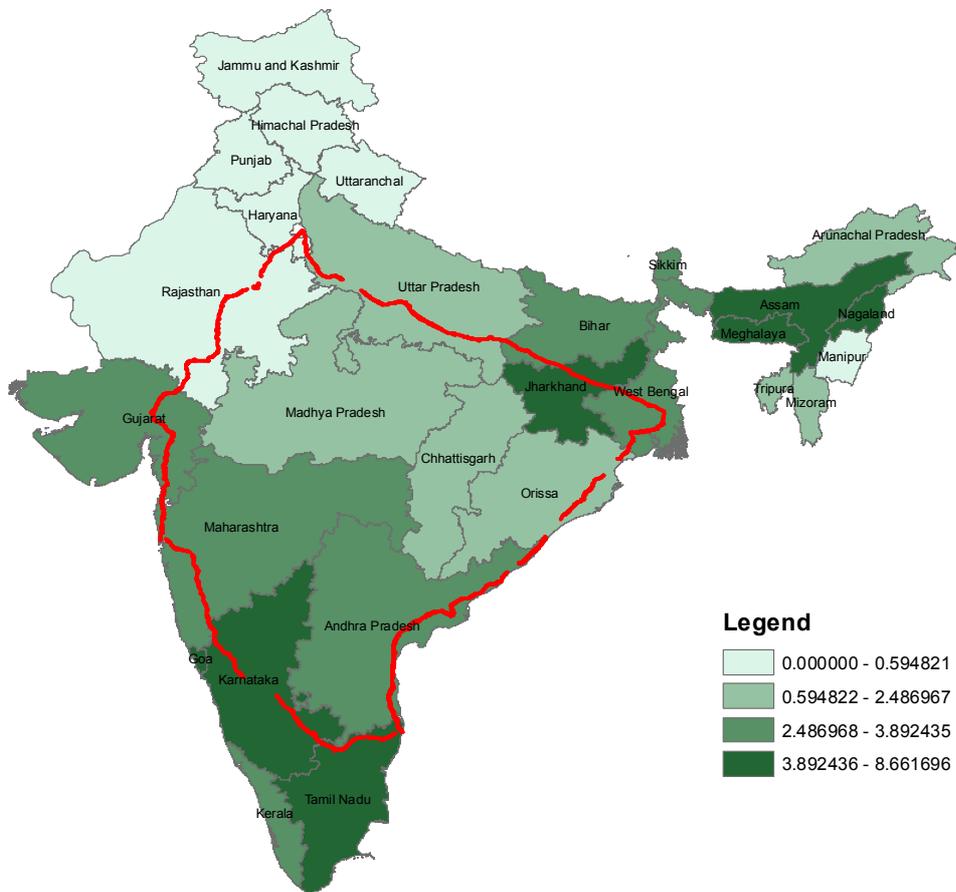


Figure III shows the % change in transportation costs due to the construction of the GQ at the state level.

FIGURE IV

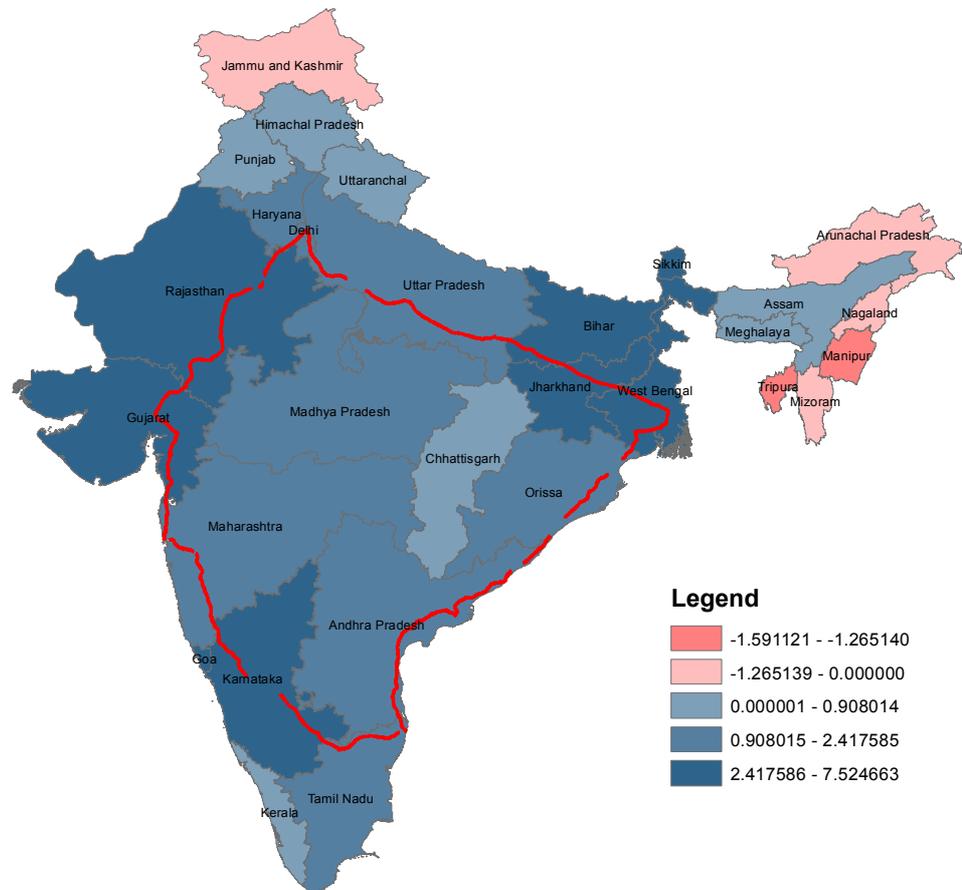PERCENTAGE CHANGE IN REAL INCOME AFTER GQ



Figure IV shows the % change in real after the decrease in transportation costs due to the construction of the GQ. The numbers represented in this map correspond to the ones presented in column 2 of Table V.

## FIGURE V

### SPATIAL DISTRIBUTION OF THE TOP 1% FIRMS IN TERMS OF MARKUPS (MODEL)

(A) Firms selling to Arunachal Pradesh
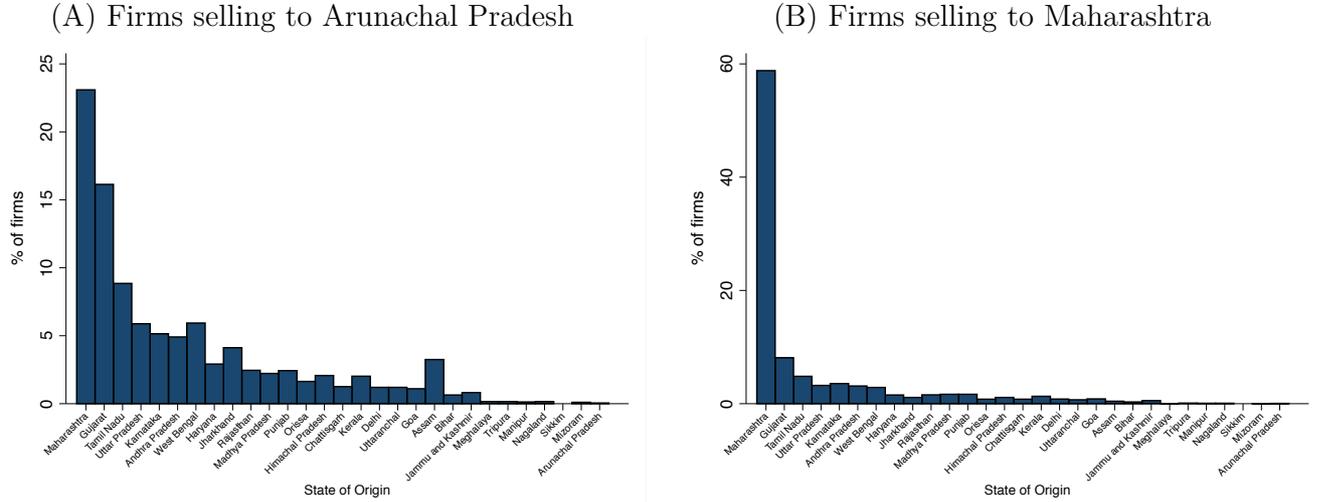
(B) Firms selling to Maharashtra



Figure (V) shows the distribution of states in which the top 1% of firms in terms of markups operate. Panel (A) refers to the markups charged on goods sold in Arunachal Pradesh. Panel (B) refers to the markups charged on goods sold in Maharashtra.

## FIGURE VI

### DISTRIBUTION OF THE CHANGE IN MARKUPS (MODEL)

(A) Firms selling to Arunachal Pradesh
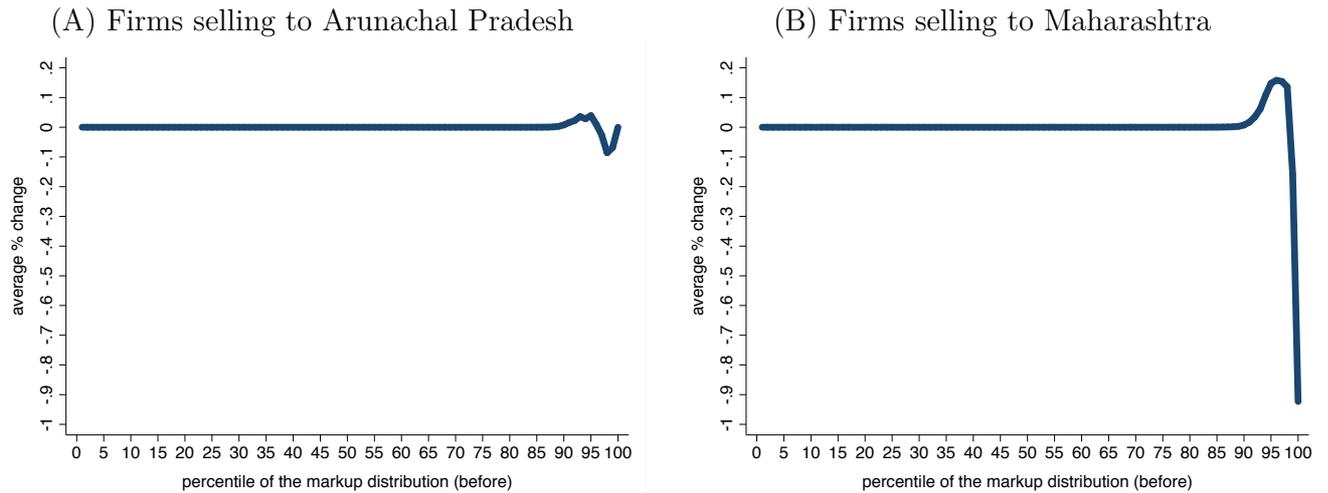
(B) Firms selling to Maharashtra



Figure (VI) shows the average percentage change in markups across firms across different percentiles of the distribution of markups before the construction of the GQ in the model. Panel (A) refers to the markups charged on goods sold in Arunachal Pradesh. Panel (B) refers to the markups charged on goods sold in Maharashtra.

TABLE IX

MARKUPS IN THE MODEL (BY DESTINATION MARKET)

| state | std | | | mean | | | p95 | | | p99 | | | log p99/p50 | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | before | after | % change | before | after | % change | before | after | % change | before | after | % change | before | after | % change |
| Maharashtra | 0.0587 | 0.0578 | -1.5421 | 1.0616 | 1.0615 | -0.0058 | 1.071 | 1.072 | 0.137 | 1.307 | 1.300 | -0.591 | 0.216 | 0.210 | -2.772 |
| Gujarat | 0.0577 | 0.0567 | -1.6595 | 1.0615 | 1.0615 | -0.0060 | 1.073 | 1.075 | 0.166 | 1.292 | 1.287 | -0.338 | 0.204 | 0.201 | -1.673 |
| Tamil Nadu | 0.0571 | 0.0565 | -1.0900 | 1.0615 | 1.0615 | -0.0041 | 1.073 | 1.075 | 0.157 | 1.288 | 1.284 | -0.307 | 0.201 | 0.198 | -1.540 |
| Uttar Pradesh | 0.0550 | 0.0546 | -0.8576 | 1.0614 | 1.0613 | -0.0031 | 1.079 | 1.080 | 0.147 | 1.266 | 1.259 | -0.607 | 0.184 | 0.178 | -3.350 |
| Karnataka | 0.0560 | 0.0553 | -1.2812 | 1.0614 | 1.0614 | -0.0048 | 1.074 | 1.076 | 0.185 | 1.279 | 1.268 | -0.840 | 0.194 | 0.186 | -4.424 |
| Andhra Pradesh | 0.0551 | 0.0548 | -0.6398 | 1.0614 | 1.0613 | -0.0024 | 1.077 | 1.078 | 0.095 | 1.265 | 1.259 | -0.453 | 0.183 | 0.179 | -2.502 |
| West Bengal | 0.0555 | 0.0548 | -1.2893 | 1.0614 | 1.0613 | -0.0047 | 1.077 | 1.079 | 0.175 | 1.271 | 1.263 | -0.654 | 0.188 | 0.181 | -3.546 |
| Haryana | 0.0546 | 0.0542 | -0.7539 | 1.0613 | 1.0613 | -0.0028 | 1.081 | 1.083 | 0.195 | 1.255 | 1.249 | -0.423 | 0.175 | 0.171 | -2.450 |
| Jharkhand | 0.0549 | 0.0545 | -0.7323 | 1.0613 | 1.0613 | -0.0024 | 1.081 | 1.082 | 0.066 | 1.249 | 1.249 | 0.048 | 0.170 | 0.171 | 0.282 |
| Rajasthan | 0.0543 | 0.0540 | -0.5574 | 1.0613 | 1.0613 | -0.0021 | 1.081 | 1.081 | 0.082 | 1.251 | 1.246 | -0.404 | 0.172 | 0.168 | -2.381 |
| Madhya Pradesh | 0.0541 | 0.0540 | -0.2828 | 1.0613 | 1.0613 | -0.0010 | 1.082 | 1.082 | 0.028 | 1.251 | 1.248 | -0.233 | 0.172 | 0.169 | -1.369 |
| Punjab | 0.0543 | 0.0541 | -0.4201 | 1.0613 | 1.0613 | -0.0016 | 1.080 | 1.081 | 0.089 | 1.250 | 1.247 | -0.272 | 0.171 | 0.169 | -1.602 |
| Orissa | 0.0545 | 0.0542 | -0.5899 | 1.0613 | 1.0613 | -0.0021 | 1.080 | 1.081 | 0.115 | 1.257 | 1.254 | -0.282 | 0.177 | 0.174 | -1.604 |
| Himachal Pradesh | 0.0544 | 0.0543 | -0.2947 | 1.0613 | 1.0613 | -0.0011 | 1.081 | 1.082 | 0.074 | 1.253 | 1.251 | -0.194 | 0.174 | 0.172 | -1.123 |
| Chattisgarh | 0.0544 | 0.0542 | -0.3137 | 1.0613 | 1.0613 | -0.0012 | 1.081 | 1.082 | 0.077 | 1.252 | 1.250 | -0.170 | 0.173 | 0.171 | -0.987 |
| Kerala | 0.0550 | 0.0547 | -0.4782 | 1.0614 | 1.0613 | -0.0018 | 1.077 | 1.078 | 0.131 | 1.260 | 1.255 | -0.377 | 0.179 | 0.175 | -2.127 |
| Delhi | 0.0544 | 0.0540 | -0.7932 | 1.0613 | 1.0613 | -0.0029 | 1.081 | 1.083 | 0.179 | 1.251 | 1.245 | -0.449 | 0.172 | 0.168 | -2.646 |
| Uttaranchal | 0.0540 | 0.0537 | -0.4693 | 1.0613 | 1.0613 | -0.0017 | 1.082 | 1.084 | 0.111 | 1.245 | 1.241 | -0.326 | 0.167 | 0.164 | -1.974 |
| Goa | 0.0548 | 0.0547 | -0.1597 | 1.0613 | 1.0613 | -0.0006 | 1.077 | 1.077 | -0.010 | 1.255 | 1.252 | -0.225 | 0.175 | 0.173 | -1.290 |
| Assam | 0.0541 | 0.0539 | -0.2976 | 1.0613 | 1.0613 | -0.0011 | 1.082 | 1.083 | 0.078 | 1.249 | 1.247 | -0.208 | 0.171 | 0.169 | -1.226 |
| Bihar | 0.0539 | 0.0538 | -0.1016 | 1.0613 | 1.0613 | -0.0004 | 1.083 | 1.083 | 0.031 | 1.248 | 1.245 | -0.187 | 0.169 | 0.168 | -1.108 |
| Jammu and Kashmir | 0.0535 | 0.0536 | 0.0992 | 1.0613 | 1.0613 | 0.0004 | 1.084 | 1.084 | -0.013 | 1.241 | 1.242 | 0.047 | 0.164 | 0.165 | 0.285 |
| Meghalaya | 0.0537 | 0.0535 | -0.3735 | 1.0613 | 1.0612 | -0.0014 | 1.084 | 1.085 | 0.097 | 1.246 | 1.242 | -0.275 | 0.168 | 0.165 | -1.652 |
| Tripura | 0.0532 | 0.0532 | -0.0977 | 1.0612 | 1.0612 | -0.0003 | 1.086 | 1.086 | 0.028 | 1.235 | 1.234 | -0.062 | 0.159 | 0.158 | -0.389 |
| Manipur | 0.0532 | 0.0531 | -0.1054 | 1.0612 | 1.0612 | -0.0004 | 1.086 | 1.086 | 0.036 | 1.235 | 1.234 | -0.047 | 0.159 | 0.159 | -0.295 |
| Nagaland | 0.0533 | 0.0532 | -0.1916 | 1.0612 | 1.0612 | -0.0007 | 1.085 | 1.086 | 0.037 | 1.237 | 1.235 | -0.144 | 0.161 | 0.159 | -0.900 |
| Sikkim | 0.0534 | 0.0532 | -0.2768 | 1.0612 | 1.0612 | -0.0011 | 1.086 | 1.087 | 0.118 | 1.238 | 1.235 | -0.235 | 0.162 | 0.159 | -1.465 |
| Mizoram | 0.0531 | 0.0531 | -0.1070 | 1.0612 | 1.0612 | -0.0004 | 1.086 | 1.087 | 0.038 | 1.234 | 1.233 | -0.103 | 0.158 | 0.157 | -0.655 |
| Arunachal Pradesh | 0.0531 | 0.0531 | -0.1085 | 1.0612 | 1.0612 | -0.0004 | 1.086 | 1.087 | 0.039 | 1.234 | 1.233 | -0.054 | 0.158 | 0.158 | -0.342 |

Notes: Table IX shows some moments of the unconditional markup distribution generated by the model; std, mean, p95, p99, and log p99/p50 refer to the simple mean, standard deviation, 95th percentile, 99th percentile and the percentage difference between the 99th percentile and the median of the markups charged to the goods purchased by each state respectively.

# References

ADAMOPOULOS, T. (2011): "Transportation Costs, Agricultural Productivity, and Cross-Country Income Differences," *International Economic Review*, 52(2), 489–521.

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*.

ALLEN, T., AND C. ARKOLAKIS (Forthcoming): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*.

ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2012): "The Elusive Pro-Competitive Effects of Trade," *Working Paper*.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (2012): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.

ASTURIAS, J., AND S. PETTY (2013): "Endogenous Transportation Costs," Discussion paper, University of Minnesota.

ATACK, 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.

ATKIN, D., AND D. DONALDSON (2014): "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," *Working Paper*.

BANERJEE, A. V., E. DUFLO, AND N. QIAN (2012): "On the Road: Access to Transportation Infrastructure and Economic Growth in China," pp. 1–50.

BAYOUMI, T., AND B. EICHENGREEN (1998): "Is Regionalism Simply a Diversion? Evidence from the Evolution of the EC and EFTA," in *Regionalism Versus Multilateral Trade Arrangements*, ed. by T. Ito, and A. Kruger. University of Chicago Press.

BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *The American Economic Review*, 4, 1268–1290.

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.

COSTINOT, A., AND D. DONALDSON (2012): "How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-2002," *Working Paper*.

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 (Forthcoming): "Understanding Markups in the Open Economy," *American Economic Journal: Macroeconomics*.

DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNIK (2014): "Prices, Markups and Trade Reform," *Working Paper*, pp. 1–59.

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*.

DONALDSON, D., AND R. HORNBECK (2014): "Railroads and American Economic Growth: A "Market Access" Approach," *Working Paper*.

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2014): "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*.

FABER, B. (Forthcoming): "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies.*

FEENSTRA, R. C., AND D. E. WEINSTEIN (2010): "Globalization, Markups and U.S. Welfare," *NBER Working Paper 15749.*

GABLER, A., AND M. POSCHKE (2013): "Experimentation by Firms, Distortions, and Aggregate Productivity," *Review of Economic Dynamics*, 16(1), 26–38.

GARCIA-SANTANA, M., AND J. PIJOAN-MAS (2014): "The reservation laws in India and the misallocation of production factors," *Journal of Monetary Economics*, 66(0), 193 – 209.

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 (Forthcoming): "Highway to success in India: the impact of the golden quadrilateral project for the location and performance of manufacturing," *Economic Journal.*

GOURIO, F., AND N. ROYS (2013): "Size-Dependent Regulations, Firm Size Distribution, and Reallocation," *Working Paper.*

GUNER, N., G. VENTURA, AND Y. XU (2008): "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 11(4), 721–744.

HERRENDORF, B., J. A. SCHMITZ, JR., AND A. TEIXEIRA (2012): "The Role of Transportation in U.S. Economic Development: 1840-1860," *International Economic Review*, 53(3), 693–716.

HOLMES, T. J., W.-T. HSU, AND S. LEE (Forthcoming): "Allocative Efficiency, Markups, and the Welfare Gains from Trade," *Journal of International Economics.*

HOPENHAYN, H. (2014): "Firms, Misallocation, and Aggregate Productivity: A Review," *The Annual Review of Economics*, 6(1).

HOTTMAN, C., S. J. REDDING, AND D. E. WEINSTEIN (2014): "What is 'Firm Heterogeneity'in Trade Models? The Role of Quality, Scope, Markups, and Cost," *NBER Working Paper.*

HSIEH, C.-T., AND P. KLENOW (2009): "Misallocation and Manufacuring TFP in China and India," *Quaterly Journal of Economics*, 124(4), 1403–1448.

KOTHARI, S. (2013): "The Size Distribution of Manufacturing Plants and Development," *Working Paper.*

KRUEGER, A. O. (1999): "Trade Creation and Trade Diversion Under NAFTA," *National Bureau of Economic Research Working Paper 7429*, (7429).

LICANDRO, O., AND G. IMPULLITTI (2013): "Trade, firm selection, and innovation: the competition channel," *Working Paper*, pp. 1–54.

LOECKER, J. D., AND F. WARZYNSKI (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102(6), 2437–2471.

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., AND G. I. P. OTTAVIANO (2008): "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 75(1), 295–316.

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.

REDDING, S., AND M. A. TURNER (2014): "Transportation Costs and the Spatial Organization of Economic Activity," NBER 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.

SOTELO, S. (2014): "Trade Frictions and Agricultural Productivity: Theory and Evidence from Peru"," *Working Paper*.

WAUGH, M. E. (2010): "International Trade and Income Differences," *American Economic Review*, 100(5), 2093–2124.

WORLD-BANK (2002): "India's Transport Sector: The Challenges Ahead," *World Bank Reports Series*, 1, 1–65.