Online Appendix for:

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

UPF, Barcelona GSE and CEPR

Roberto Ramos

Bank of Spain

September 14, 2017

Contents

A Unit Misreporting 2

B Details on the Construction of the Graph 2

C Accounting for Foreign Competition When Estimating Transportation Costs 3

D Details on Data Preparation for the Diff-in-Diff Specifications 3

E The Firm-Level Linear Relationship Between Labor and Sectoral Shares 5

F Productivity Levels Across States in the Calibration 6

G Labor migration across states after the construction of the GQ 8

This online appendix provides supplementary material to the main article. It is organized as follows. Section A discusses the issue of unit misreporting. Section B gives additional information on the construction of the graph. Section C addresses the issue of foreign competition in the estimation of transport costs. Section D provides details of the diff-in-diff specifications. Section E derives the relationship implied by the model between labor and sectoral shares in the case of goods that are produced only in one state. Section F discusses an alternative calibration that explicitly takes into account the cross-state heterogeneity in productivity. Finally, Section G addresses the issue of labor migration across states.

A Unit Misreporting

We identify some firms that report quantities in units different from what they should, which generates big changes in the scale of prices for some goods. Figure I shows the example of Chlorophos (ASICC 31611). The average log price of this input is 5.6 for some firms and 12.3 for others. This is due to the fact that some firms report quantities of this input in tons, as intended by the survey, whereas others do so in kilograms. Hence, the difference in the average log price of 6.7 can be explained by a denominator multiplied by 1,000 ($ln(1000) = 6.9$).

In order to address this issue, we first identify large changes in prices within each product. Then, we treat each set of prices having a different scale as a different product. Specifically, we sort every product by price (from low to high) and assign a discrete jump in prices if the ratio of one price over the previous is larger than 20. If this happens, we consider the set of prices to the left and right of the large price change as different goods, by using separate fixed effects when estimating equation (17). Kothari (2014) uses a similar strategy (see Appendix C of that paper).

B Details on the Construction of the Graph

The geospatial data on the national highway system of India was provided to us by ML Infomap. In order to convert the national highway system into a graph, we used Network Analyst in ArcGIS. A node on the graph is the most populous city in a district. Cities that are not immediately on the road are mapped to the closest straight-line point to a National Highway. In addition, the graph has nodes at any point where the road changed from being treated to non-treated (upgraded vs. not upgraded). The nodes are connected by an arc if it is possible to travel from one node to the other without passing through another node.

Figure I

Price Computed for Input Chlorophos (ASICC 31611)

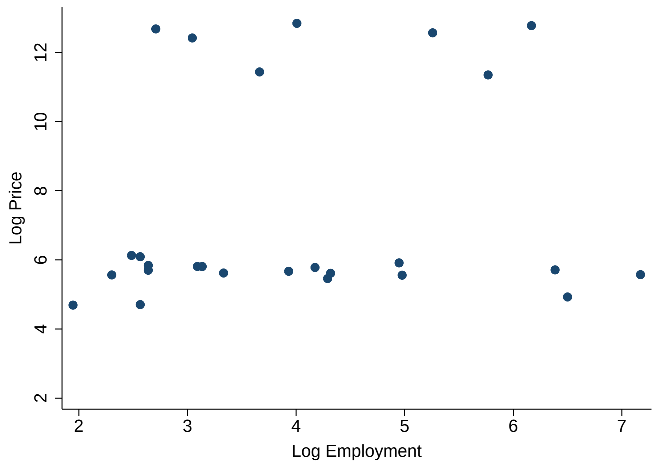

Figure I shows the price of input good Chlorophos (ASICC 31611), computed as value over quantity consumed, against the size of the firm. The change in scale for some prices shows the unit misreporting in some observations.

C Accounting for Foreign Competition When Estimating Transportation Costs

As discussed in the main text, our strategy to identify transportation costs relies on the identification of plants that act as monopolists in their sector. To account for a possible bias induced by foreign competition, we also run the regression described in equation (17) of the main text excluding those districts in which foreign inputs account for more than 5% of total input usage of the good in the district. In Figure II we plot the estimates of this regression compared to the baseline reported in the main text. We find a similar pattern of price changes across deciles of effective distance.

D Details on Data Preparation for the Diff-in-Diff Specifications

In the differences-in-differences specification of prices (Section 8.2), for each round and district, we compute the price of each product as a weighted average of the prices paid by the plants using that product as an intermediate input in that district. Each price is calculated as the total input value over the total input quantity. We exclude the goods with unit misreporting within a round as well as outliers in the change of prices across rounds (2% of observations). We have information of 920 ASICC products consumed in 325 districts in both 2001 and 2006.

In the differences-in-differences specification of allocative efficiency (Section 8.2), for each 4-digit industry and district, we compute the changes in the Olley and Pakes (1996) within-industry cross-sectional covariance between size and productivity. We restrict the sample to districts containing

Figure II

Estimation of Transportation Costs Accounting for Foreign Competition

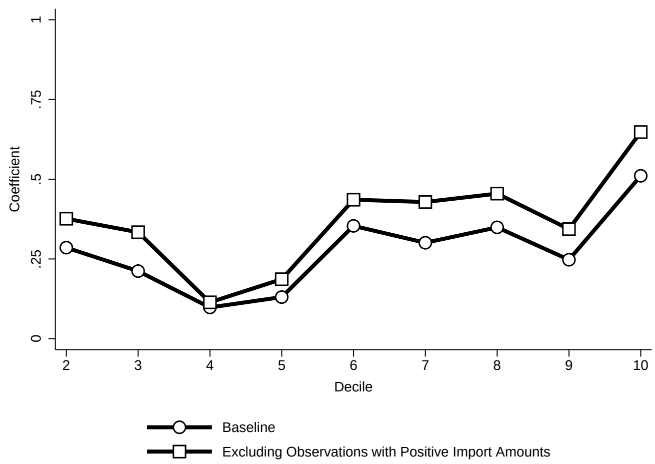

Figure II shows the coefficients of the estimation of equation (17) in the baseline specification (column (1) of Table I) and excluding those observations in which imported inputs account for more than 5% of total input usage (around 11 percent of observations).

at least 10 plants and trim 1% of the distribution of changes in the covariance term at each tail. Our final sample have 117 industries and 466 districts.

Additionally, using ArcGIS, we compute the shortest straight-line distance from each district to a completed stretch of the GQ in March 2001 and March 2006. We then compute several treatment dummies taking the value 1 if the district is within a certain distance of the GQ and zero otherwise. The treated districts are those for which the treatment dummy changes between 2001 and 2006. The control districts are those that did not gain further access to the highway between 2001 and 2006.

In the specifications excluding nodal districts, we exclude Delhi, Mumbai, Chennai, and Calcutta, as well as a few contiguous suburbs identified by Datta (2012) which were on the GQ as a matter of design rather than fortuitousness. These districts include Gurgaon, Faridabad, Ghaziabad, Gautam Buddha Nagar, and Thane. Finally, we exclude the few districts that were within 50 kilometers of an upgraded portion of the GQ in 2001. The reason is that we want to compare the evolution of outcomes in districts that were treated in 2006 with districts that were not treated in 2006.

Our benchmark administrative division is that of 2001, hence districts in 2006 that were carved out from existent districts in 2001 are assigned to their 2001 district. Using geospatial data along with ArcGIS, we compute the shortest straight-line distance from every district to the nearest completed stretch of the GQ in March 2001 and March 2006.

E The Firm-Level Linear Relationship Between Labor and Sectoral Shares

The optimal pricing decision of the firm is given by:

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

where

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

and

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

Multiplying by $l_d^o(j,k)$ on both sides of the equation and re-ordering the terms:

$$\frac{\tau_d^o W_o l_d^o(j,k)}{p_d^o(j,k)c_d^o(j,k)} = \frac{\epsilon_d^o(j,k)-1}{\epsilon_d^o(j,k)}.$$

We now introduce additional notation to define the price that the firm sets before charging transportation costs. Let the price set by the firm at the gate of the factory be denoted:

$$\tilde{p}_d^o(j,k) = \frac{p_d^o(j,k)}{\tau_d^o}.$$

This is the price that we can compute in the data when using firms' reported sales and physical units. Using this definition, we can write the firm's inverse of the markup as:

$$\frac{W_o l_d^o(j,k)}{\tilde{p}_d^o(j,k)c_d^o(j,k)} = \frac{\epsilon_d^o(j,k)-1}{\epsilon_d^o(j,k)},$$

where $\frac{W_o l_d^o(j,k)}{\tilde{p}_d^o(j,k)c_d^o(j,k)}$ is the labor share of firms' total revenue at destination d before transportation costs are charged. Using the expression for the firm's elasticity:

$$\begin{aligned}\frac{W_o l_d^o(j,k)}{\tilde{p}_d^o(j,k)c_d^o(j,k)} &= \left[\left(\omega_d^o(j,k)\tfrac{1}{\theta} + (1-\omega_d^o(j,k))\tfrac{1}{\gamma}\right)^{-1} - 1\right]\left(\omega_d^o(j,k)\tfrac{1}{\theta} + (1-\omega_d^o(j,k))\tfrac{1}{\gamma}\right)\\ &= 1-\omega_d^o(j,k)\tfrac{1}{\theta} - (1-\omega_d^o(j,k))\tfrac{1}{\gamma} = 1-\omega_d^o(j,k)\tfrac{1}{\theta} - \tfrac{1}{\gamma} + \tfrac{1}{\gamma}\omega_d^o(j,k),\end{aligned}$$

which yields the following linear relationship between the firms' labor share and sectoral share:

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

Goods produced only in one state For those goods that are produced only in one location (location o for instance), the expression for firms' market share becomes:

$$\omega_d^o(j,k) = \frac{p_d^o(j,k)^{1-\gamma}}{\sum_{k=1}^K p_d^o(j,k)^{1-\gamma}} = \frac{(\tau_d^o)^{1-\gamma}\tilde{p}_d^o(j,k)^{1-\gamma}}{\sum_{k=1}^K (\tau_d^o)^{1-\gamma}\tilde{p}_d^o(j,k)^{1-\gamma}} = \frac{(\tau_d^o)^{1-\gamma}\tilde{p}_d^o(j,k)^{1-\gamma}}{(\tau_d^o)^{1-\gamma}\sum_{k=1}^K \tilde{p}_d^o(j,k)^{1-\gamma}} = \frac{\tilde{p}_d^o(j,k)^{1-\gamma}}{\sum_{k=1}^K \tilde{p}_d^o(j,k)^{1-\gamma}}.$$

Note that $\omega_d^o(j,k)$ will be constant across different destinations. Then, summing equation (1) across destinations we get:

$$\frac{W_o l^o(j,k)}{\tilde{p}^o(j,k)c^o(j,k)} = 1 - \frac{1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right)\omega^o(j,k)$$

where:

$$\begin{array}{rcl} l^o(j,k) & = & \sum_{d=1}^{N} l_d^o(j,k) \\ \tilde{p}^o(j,k)c^o(j,k) & = & \sum_{d=1}^{N} \tilde{p}_d^o(j,k)c_d^o(j,k) \\ \omega^o(j,k) & = & \frac{\sum_{d=1}^{N} \tilde{p}_d^o(j,k)c_d^o(j,k)}{\sum_{k=1}^{K}\sum_{d=1}^{N} \tilde{p}_d^o(j,k)c_d^o(j,k)} \end{array}$$

F Productivity Levels Across States in the Calibration

Proposition 1. *Consider the calibrated economy in Section 6. Now consider an alternative calibration in which we raise the productivities of all firms in one economy by a common factor and re-calibrate the labor endowments of all economies to match the total income of each state. In the new calibration, the equilibrium distribution of prices, markups, market shares, value of sales, and quantities sold across destinations remain the same for all firms. Furthermore, the price index and aggregate output for each state remain the same. Finally, total labor income and profits for each state also remain unchanged.*

Proof. Suppose there is a calibration in which the labor endowment is set to match the GDP of each economy, where the GDP of economy n is defined to be $W_nL_n + \Pi_n$. In the initial calibration we normalize the labor endowment of economy 1, $\tilde{L}_1$, such that its GDP is equal to 1. The labor endowment of economy n, $\tilde{L}_n$, is calibrated so that economy n's GDP matches the ratio of its own GDP to that of economy 1's, which we observe in the data. The following are the endogenous variables from the initial calibration: $\tilde{p}_d^o(j,k)$, $\tilde{\epsilon}_d^o(j,k)$, $\tilde{\omega}_d^o(j,k)$, $\tilde{c}_d^o(j,k)$, and $\tilde{\pi}_d^o(j,k)$ for a firm k in sector j in state o serving destination d; $\tilde{P}_n(j)$ and $\tilde{C}_n(j)$ for sector j in state n; $\tilde{P}_n$, $\tilde{C}_n$, $\tilde{\Pi}_n$, and $\tilde{W}_n$ in state n. The productivities are characterized by $\tilde{a}_n(j,k)$. We suppose that the wage in economy 1 is the numeraire.

Consider an alternative calibration in which the productivities of all firms in economy q are raise by a factor f and the productivities in all other economies remain the same. Furthermore, suppose that in the new calibration we again normalize the GDP of economy 1 to 1. The labor endowment of economy n, $\hat{L}_n$, is calibrated to match the relative size of economy n's GDP to that of economy 1's. The endogenous variables in the new calibration are: $\hat{p}_d^o(j,k)$, $\hat{\epsilon}_d^o(k,j)$, $\hat{\omega}_d^o(j,k)$, $\hat{c}_d^o(j,k)$,$\hat{\pi}_d^o(j,k)$, $\hat{P}_n(j)$, $\hat{C}_n(j)$, $\hat{P}_n$, $\hat{C}_n$, $\hat{\Pi}_n$, and $\hat{W}_n$. The new productivities are characterized by $\hat{a}_n(j,k)$. Given the setup, we know that $\hat{a}_n(j,k) = \tilde{a}_n(j,k)$ if $n \neq q$ and $\hat{a}_q(j,k) = f * \tilde{a}_q(j,k)$. We suppose that the wage in economy 1 is the numeraire and that $q \neq 1$.

We follow a guess and verify procedure. We guess that in the new calibration the wage for economy q will be higher by a factor f and the wages of all other economies remain the same.

Thus, we have that $\hat{W}_n(j,k) = \tilde{W}_n(j,k)$ if $n \neq q$ and $\hat{W}_q(j,k) = f * \tilde{W}_q(j,k)$. We will later verify that the guess satisfies the labor clearing condition of each economy.

We begin by characterizing the industry-level equilibrium. We know that prices are characterized by:

$$p_d^o(j,k) = \underbrace{\frac{\epsilon_d^o(j,k)}{\epsilon_d^o(j,k)-1}}_{\text{Markup}} \underbrace{\frac{W_o}{a_o(j,k)}}_{\text{Marginal cost}} \tau_d^o,$$

We see that the marginal cost component remains the same for all firms $\left(\frac{\tilde{W}_o}{\tilde{a}_o(j,k)}\tau_d^o = \frac{\hat{W}_o}{\hat{a}_o(j,k)}\tau_d^o\right)$. This is true in economy q since both wages and productivity increase by a factor f. In other economies, this is true since wages and productivities remain the same. The markup portion of price is characterized by:

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

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

Since the marginal cost remains the same, the market shares across destination also remains unchanged. Thus, we conclude that: $\tilde{p}_n^o(j,k) = \hat{p}_n^o(j,k)$, $\tilde{\epsilon}_d^o(j,k) = \hat{\epsilon}_d^o(j,k)$, and $\tilde{\omega}_d^o(j,k) = \hat{\omega}_d^o(j,k)$.

We now find the industry-price index and aggregate price index under the new calibration. We know that these variables are characterized by:

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

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

Thus, it must be that $\hat{P}_n(j) = \tilde{P}_n(j)$ and $\hat{P}_n = \tilde{P}_n$ since $\hat{p}_n^o(j,k) = \tilde{p}_n^o(j,k)$.

We will now show that the quantity produced by firms remains the same across the two calibrations. We know that the following conditions characterize the quantity produced by firms

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

We re-arrange this condition to find that

$$c_n^o(j,k) = \frac{P_n C_n}{p_n^o(j,k)^{\gamma} P_n(j)^{\theta-\gamma} P_n^{1-\theta}}.$$

We conclude that $\hat{c}_n^o(j,k) = \tilde{c}_n^o(j,k)$ since $\hat{p}_n^o(j,k) = \tilde{p}_n^o(j,k)$, $\hat{P}_n(j) = \tilde{P}_n(j)$, $\hat{P}_n = \tilde{P}_n$, and $\hat{P}_n\hat{C}_n = \tilde{P}_n\tilde{C}_n$. The last condition holds since the GDP of all economies remains the same in both calibrations. The reason is that the GDP of economy 1 is 1 in both calibrations and the relative GDP's for the other economies does not change. We also know that

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

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

Because $\hat{c}_n^o(j,k) = \tilde{c}_n^o(j,k)$, we conclude that $\hat{C}_n(j) = \tilde{C}_n(j)$ and $\hat{C}_n = \tilde{C}_n$

Next, we show that the profits of all firms remain the same. The expression for profits is

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

We conclude that $\hat{\pi}_d^o(j,k) = \tilde{\pi}_d^o(j,k)$ since $\hat{p}_n^o(j,k) = \tilde{p}_n^o(j,k)$, $\hat{c}_n^o(j,k) = \tilde{c}_n^o(j,k)$ and $\frac{\hat{W}_o \tau_d^o}{\hat{a}_o(j,k)} = \frac{\tilde{W}_o \tau_d^o}{\tilde{a}_o(j,k)}$. We also conclude that $\hat{\Pi}_n = \tilde{\Pi}_n$ since the expression for aggregate profits is

$$\Pi_n = \int_0^1 \left(\sum_{d=1}^{N} \sum_{k=1}^{K_{nj}} \pi_d^n(j,k) \right) dj.$$

Because both GDP and aggregate profits remain the same for all economies across the two calibrations, then it must be that $\hat{W}_n \hat{L}_n = \tilde{W}_n \tilde{L}_n$, implying that total labor income remains unchanged. Thus, $\hat{L}_n = \tilde{L}_n$ if $n \neq q$. It also implies that $\hat{L}_q = \tilde{L}_q / f$ since $\hat{W}_q = f * \tilde{W}_q$.

We will now verify that labor demand equals labor supply given our guess of w. The condition for state n is

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

We know that the quantity produced by all firms remains the same ($\hat{c}_n^o(j,k) = \tilde{c}_n^o(j,k)$). In economy q, the productivity of all firms increases by factor f. Thus, all firms use $1/f$ as much labor, which is consistent with the market clearing condition. In all other economies, the labor demand and labor endowment remain the same, which means that the market clearing condition holds. Thus, the guess has been verified.

□

G Labor migration across states after the construction of the GQ

The model presented in the main text does not allow for labor mobility across states. Yet if the construction of the GQ resulted in an increase in inter-state migration, the estimated effects of the GQ on welfare could change. In this regard, it must be noted that Mahapatro (2012) shows that most of Indian internal migration happens within the state boundaries, and hence the estimated welfare effects would hardly be affected.

To address this issue formally, we estimated a differences-in-differences regression of state internal migration changes between 1999/2000 and 2007/2008 against a dummy capturing whether the state is crossed by the GQ. Internal migration rates are defined as immigrants - emigrants from other states over population. We obtained the data from Table 4 of Mahapatro (2012). We find that the GQ dummy is not statistically significant in explaining changes in migration rates. Its p-value is .32 if we include Delhi, which underwent a disproportionally high influx of migrants during this period and constitute an outlier, and 0.84 if we exclude it. Therefore, we do not find evidence that the GQ triggered a large influx of migrants toward the states connected by this infrastructure, at least to a first approximation.

References

Datta, S. (2012): "The Impact of Improved Highways on Indian Firms," *Journal of Development Economics*, 99(1), 46–57.

Kothari, S. (2014): "The Size Distribution of Manufacturing Plants and Development," *IMF Working Paper No. 14/236*.

Mahapatro, S. R. (2012): "The Changing Pattern of Internal Migration in India," *Working Paper*, http://epc2012.princeton.edu/papers/121017.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.