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

January 23, 2017

Contents

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.

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 (2013) uses a similar strategy (see Appendix C of that paper).

FIGURE I

PRICE COMPUTED FOR INPUT CHLOROPHOS (ASICC 31611)

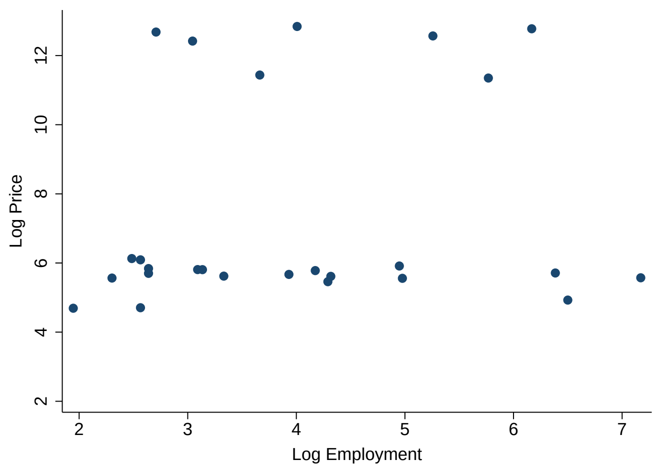

Figure I shows the price of input good Chlorophos (ASICC 31611), computed as value over quantity consumed, against the size of the firm. The change in scale for some prices shows the unit misreporting in some observations.

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.

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.

Figure II

Estimation of Transportation Costs

Accounting for Foreign Competition

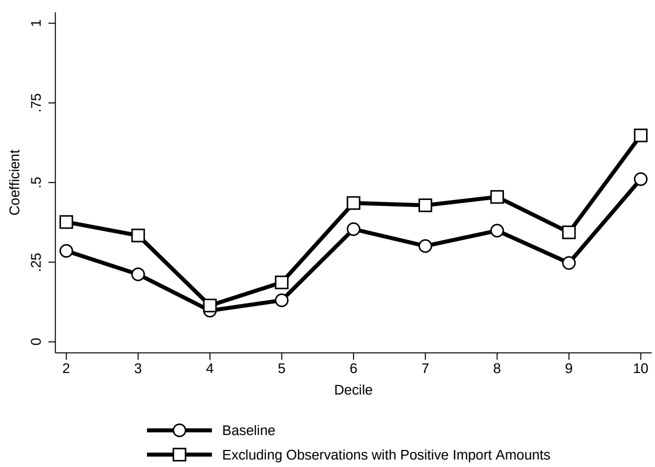

Figure II shows the coefficients of the estimation of equation (17) in the baseline specification (column (1) of Table I) and excluding those observations in which imported inputs account for more than 5% of total input usage (around 11 percent of observations).

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

In the differences-in-differences specification of prices (Section 8.1), 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 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:

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

References

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

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

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