How Productive is Public Investment? Evidence from Indian Manufacturing*

Santanu Chatterjee† — *University of Georgia*
Thomas Lebesmuehlbacher‡ — *Xavier University*

Abhinav Narayanan§
Reserve Bank of India

Abstract

This paper uses firm-level data on formal and informal production in the manufacturing sector in India to examine the sectoral consequences of government investment in public infrastructure. On average, public investment has a strong and positive association with the productivity of formal sector firms, with its output elasticity ranging between 0.08-0.17. By contrast, there is no systematic association between public investment and the output of the average firm in the informal sector. Using a major highway construction project in India as a natural experiment, we show that the complementarities generated by public investment accrue mainly to larger firms, leading to a crowding out of output for smaller informal firms. Consequently, this mitigates the over all benefits of public investment for the informal sector.

Keywords: Informal sector, formal sector, public investment, output elasticity, quantile regression, Golden Quadrilateral, Infrastructure, Manufacturing, India.

JEL Classification: E2, H4, H5

*We thank Shankha Chakraborty, Ejaz Ghani, David Mustard, Chris Papageorgiou, Nishith Prakash, Erwan Quintin, Jagadish Sivadasan, Ian Schmutte, and Meghan Skira for constructive comments on an earlier draft. The paper has also benefited from presentations at Florida International University, the University of Georgia, the World Bank-IHD Global Conference on Prosperity, Equality, and Sustainability in New Delhi, the Annual Growth and Development Conference at the Indian Statistical Institute in New Delhi, the SIDE Workshop at the Federal Reserve Bank of Atlanta, the Midwest Macroeconomics Meetings at Louisiana State University, and the Southern Growth workshop in Washington, DC. Research for this paper was supported by the Graduate School Dean's Award at The University of Georgia.

†Department of Economics, University of Georgia, Athens, GA 30602, USA. Email: schatt@uga.edu

‡Department of Economics, Xavier University, Cincinnati, OH 45027, USA. Email: lebesmuehlbachert@xavier.edu

§Strategic Research Unit, Reserve Bank of India, Mumbai 400001, INDIA. Email:abhinavnarayanan@rbi.org.in

1 Introduction

Informal production is a pervasive feature of most developing countries. As such, this sector consists of small, unregistered firms that typically produce labor intensive non-traded goods and services, with little or no access to capital markets, and limited outward labor mobility to the formal or organized sector (La Porta and Shleifer, 2014). However, this sector plays an important role in the structural evolution of these countries, accounting for about 42 percent of GDP, and absorbing between $48-54$ percent of the labor force (Schneider et al. 2010). Given underlying capital and labor market rigidities, informal sector firms may have to rely heavily on government-provided investment goods such as transportation, power, water, etc. for production purposes. This is especially relevant, with many public goods and services being non-excludable in developing countries. However, very little, if anything, is known about the quantitative benefits of government investment (and the resulting stock of public capital) for informal production in developing countries. In this paper, we use two large firm-level datasets on formal and informal production in the manufacturing sector in India and a natural experiment based on a major highway construction project to examine the sectoral consequences of government investment on firm-level productivity.

Despite being a high-growth emerging market, the Indian economy is largely informal, with this sector contributing to 55 percent of GDP and employing about 84 percent of the non-agricultural labor force in 2010 (ILO, 2013).[1] Figure 1 shows the substantial variation in the share of formal and informal manufacturing across Indian states in 2010, with 14 of 23 states having more than 50 percent of their manufacturing output generated by informal production. Figures 2 and 3 depict the average firm-level capital intensity and output-labor ratio for cross-sections of manufacturing firms in the formal and informal sectors for 1999 and 2010, respectively. For example, in 2010 the capital intensity of formal sector firms exceeded that of informal firms by a factor of 5, while output per worker was higher by a factor of about 10. Interestingly, however, these gaps were smaller in 2010 than they were in 1999, suggesting that during this period, informal sector firms have indeed been able to improve both their relative usage of capital as well as labor productivity. This point is further underscored in Figure 4, which shows that the output share of the informal sector, though quite substantial, has been on a downward trend, declining from about 60 percent of GDP in 1999 to 55 percent in 2010.

One factor that may affect the output of both formal and informal sector firms is the

[1] Mehrotra et al. (2014) document that between 2004-2012, a period of relatively high economic growth for India, the share of informal employment in the manufacturing sector was very large and persistent, at around 89 percent. Informal employment is a job-based concept, comprising of workers who lack access to basic legal protection, social security, and employment benefits (ILO, 2013).

government's provision of public infrastructure, which may serve as an input in the firm's production process. Essentially, public spending on roads, power, water, sanitation, communications, healthcare, and education may have complementary spillovers for private factors of production in both sectors. As such, public investment may help alleviate the credit and labor market constraints that firms typically face, especially in the informal sector. Indeed, infrastructure investment has been a centre-piece of public policy in India over the past two decades or so.[2] As shown in Figure 5, the share of total infrastructure spending in GDP increased from 6.4 percent in 2008 to about 9 percent in 2017, with more than 70 percent of this spending coming from the public sector. A critical consideration here is the effect of the rising share of infrastructure spending in India on the productivity of formal and informal sector firms. Given the relative magnitude of public investment and the share of the informal sector in India, their underlying relationship (if any) is of critical importance for the design and implementation of public policy.

In this paper, we attempt to bridge a gap between two strands of research that have evolved largely independently of each other. On the one hand, starting with the work of Aschauer (1989), a voluminous empirical literature has explored the productivity benefits of public investment in infrastructure, with a rich diversity of results.[3] However, these studies have, without exception, considered either industrialized countries (where the share of informal production is relatively small), or only for the formal sector in developing countries. On the other hand, the literature on the informal sector has mainly focused on issues of measurement of its output share (Schneider and Enste 2000, La Porta and Shleifer 2008, 2014, and Gomis-Porqueras et al. 2014), or issues pertaining to tax policy and enforcement (Rauch 1991, Ihrig and Moe 2004, Turnovsky and Basher 2009, Prado 2011, and Ordonez 2014). The quantitative importance of public investment for this type of production has generally been ignored. Consequently, in the context of a developing economy, the quantitative role of public investment for firm-level productivity cannot be well understood unless its effects on the informal sector are accounted for. Therefore, the first contribution of this paper is to estimate the output elasticity of public investment for private production, both in the formal and informal sectors. Second, while most studies on public investment are conducted at a fairly aggregated level (at the level of a country, state or region), we attempt to estimate its sectoral productivity benefits at the level of the individual firm. In the case of India,

[2]See, for example, two recent reports by the McKInsey Global Institute (2013) and the Urban Land Institute and Ernst & Young (2013) on trends in public infrastructure spending in emerging markets like India.

[3]See, for example, Munnell and Cook (1990), Lynde and Richmond (1992), Gramlich (1994), and Holtz-Eakin and Schwartz (1995), and Devarajan et al. (1996) for some early contributions. Bom and Ligthart (2014) provide an excellent survey and meta-analysis of the recent empirical literature.

for example, while Binswanger et al. (1993), Lall (1999), Mitra et al. (2002), Zhang and Fan (2004), and Hulten et al. (2006), among others, have examined the effects of public infrastructure for the formal sector at the state, district, or industry level, there is no current evidence of its sectoral importance at the level of the firm. The firm-level datasets we use for our study enable us to shed light on the role of public investment and infrastructure at a much more disaggregated level than previously studied. We view this as an additional contribution to the literature.[4] Finally, from the perspective of designing public policy, it is important to know how the spillovers from public investment are dispersed over the size distribution of firms in each sector. In other words, do larger firms tend to benefit more or less relative to their smaller counterparts from government spending on public goods? This may help determine how public goods should be targeted to firms in each sector. To the best of our knowledge, our analysis is the first to shed light on this issue.

In India, the main source of information at the firm level for the formal sector is the Annual Survey of Industries (ASI), while for the informal sector it is the surveys conducted by the National Sample Survey Organization (NSSO). Though the ASI surveys firms on an annual basis, the NSSO survey is conducted once every 10 years. We use data from the 2010 round for each of these surveys, since that is the latest round for which firm-level information is currently available for *both* sectors. Restricting our coverage to only the manufacturing sector, we obtain a cross-section of $32,388$ formal-sector firms (from the ASI) and $82,748$ informal-sector firms (from the NSSO) for 2010.

We proxy public investment in two different ways for our analysis. First, we use state-level data on government *Development Expenditures*, obtained from the Reserve Bank of India (under the category of "Capital Expenditures"), which includes public expenditures on transport, communications, and energy, healthcare, education, water, and sanitation, to construct measures of both the *flow* of public investment, using average annual expenditures over the $2006-2010$ period, as well as its accumulated *stock* for each state, using data over the period $2000-2010$. The flow measure is intended to capture the short-term effects of public investment, while the stock measure captures its effects over the longer term. Henceforth, we will interchangeably refer to the broad category of Development Expenditures as *public investment*, and the corresponding stock measure as *public capital.* Second, we use data from India's National Highway Development Program (NHDP) for a major highway upgrade project, namely the Golden Quadrilateral (GQ) and the North-South East-West (NS-EW) corridor, as a natural experiment to provide a causal interpretation of the effects of public

[4]Two recent studies, namely Datta (2011) and Ghani et al. (2016) examine the spatial role of India's recent expansion of its interstate system on plant-level production. These studies, however, do not distinguish between formal and informal production at the firm level.

investment for sectoral firm-level productivity.

Our empirical strategy involves the estimation of the output elasticity of public investment at the firm level in the formal and informal sectors. As such, this approach raises several econometric issues. First, the usage of private inputs like capital and labor may be endogenous to the firm's decision to produce output. We use methods suggested by Levinsohn and Petrin (2003), Sivadasan (2009), and Ackerberg et al. (2015), using past values of estimated productivity of intermediate inputs and exploiting the repeated cross-sectional nature of our dataset to control for the unobserved productivity shock at the firm-level in each sector. Second, it is plausible that the inclusion of public investment as an input generates a reverse causality problem with output in the firm's production function. To address this issue, we exploit two large infrastructure construction projects in India – the Golden Quadrilateral (GQ) and the North-South East-West (NS-EW) corridor as a natural experiment to identify the effect of infrastructure spending on firm-level output in each sector. This approach is related to recent contributions by Datta (2012) and Ghani et al. (2015), who have used the GQ as a natural experiment to identify the effect of infrastructure on formal-sector firm output. However, in contrast to these papers, our analysis involves the estimation of the output elasticity of public investment for both formal and informal manufacturing firms, and further examines how these sectoral output elasticities vary across the size distribution of firms.

Our results indicate that, on average, for formal firms, the output elasticity of public investment is about 0.08 when we consider the flow of public expenditures as the relevant input in production. However, when we use the stock specification of public investment in the production function, the corresponding output elasticity increases to about 0.17. Since the stock measure of public investment is intended to capture its long term productivity spillovers, these results suggest that the benefits accruing to formal sector firms from the accumulated stock of public capital are much larger relative to those from the flow of public investment. Within the sub-categories of public investment, we find that Economic Services is associated with higher productivity spillovers relative to Social Services, irrespective of whether we use the stock or flow specification. On the other hand, for the informal sector we find no systematic association between public investment and the output of the average firm, irrespective of whether we consider the flow or stock specification. This result is in sharp contrast to the effect of public investment and capital on the output of formal sector firms. While we use state-level government spending data for this exercise, we confirm these results using the GQ/NS-EW corridor construction as a natural experiment. Further, we do not find any evidence that the location choices for formal and informal sector firms are driven by the level of public investment in a given state.

Why does public investment not seem to influence the productivity of informal firms? To examine this further, we use quantile regressions to examine whether public investment has a differential impact along the size distribution of firms in each sector. Using the GQ/NS-EW corridor as a natural experiment, we find that the complementarities generated by an increase in public investment lead to large firms crowding out the output of small informal sector firms. This happens both within and across sectors: smaller informal sector firms tend to get crowded out by not only large firms within that sector, but also by formal sector firms. Intuitively, formals sector firms and larger informal sector firms tend to have a higher capital intensity in production than smaller informal firms. As such, public investment benefits not only larger firms in each sector, but also formal firms much more than informal ones. Therefore, informal firms, especially the smaller ones, are disproportionately hurt by the highway upgrades. This can help explain why we are unable to find any systematic association between public investment and production for the average informal sector firm in our sample. Our results thus have important implications for public policy: rather than a one-size-fits-all approach, more public investment goods might be targeted for the largest firms in each sector, especially those that are informal. This may not only help such firms appropriate the benefits of public investment, but also facilitate the transition of informal firms to formal production.

The rest of the paper is organized as follows. Section 2 discusses the data and summary statistics, while Section 3 describes the empirical specification and the identification strategy. Section 4 reports the results of the empirical analysis, and Section 5 concludes.

2 Data

The data collected for this paper spans India's formal and informal manufacturing sectors at the firm-level, and public investment at the state-level for aggregate expenditures, and at the district level for the National Highway Development Program.

2.1 Manufacturing: Formal and Informal Sectors

We use firm-level data from two sources, namely the (i) Annual Survey of Industries (ASI), and (ii) National Sample Survey Organization (NSSO). The ASI covers formal sector firms registered under Sections 2(m)(i)-(ii) of India's Factories Act of 1948, and reports annual data on firm-level receipts, expenses, and operational (firm-specific) characteristics. The data set is a repeated cross-section, where the sampling of firms changes in every round of the survey. The NSSO's "Survey of Unincorporated Non-Agricultural Enterprises" is the predominant source of firm-level information for the informal sector in India. The survey is conducted every ten years, and provides firm-level information on the ownership category, location, and other operational characteristics. Specifically, the NSSO survey includes household proprietary and partnership enterprises that are not registered under the Factories Act of 1948 or the Bidi and Cigar Workers (Condition of Employment) Act of 1966. Public sector enterprises and cooperatives are excluded from the survey. Since the ASI reports data on an annual frequency, while the NSSO does so on a ten-year frequency, we use the cross-sections from both surveys for 2010, which is the latest available survey round for the NSSO, in order to maintain compatibility between the two sectors.

The ASI survey covers $52,243$ formal sector firms in 2010. The coverage is skewed heavily towards manufacturing firms: 93.7 percent of the firms surveyed were engaged in manufacturing. The 2010 NSSO survey of the informal sector covers $334,474$ firms. Of these, only 30 percent are in the manufacturing sector, with trading activities (36 percent) and services (34 percent) making up the rest. To ensure that the sample of formal and informal sector firms are comparable, we restrict the coverage to only manufacturing firms in both sectors. This gives us a sample of $32,388$ formal-sector firms and $82,748$ informal-sector firms in 2010.

Output for both the formal and informal sector firms is measured by the gross value added (GVA; the value of total output net of total inputs). Private capital is given by the closing balance of gross fixed capital (owned and rented) at the end of the accounting year, and labor is measured by the average number of workers employed during the accounting year. An important consideration for our empirical strategy is the value of intermediate inputs. For the formal sector, we use the value of electricity consumed at the firm level as the proxy for an intermediate input. For informal sectors firms, the value of electricity usage has many missing values, as many informal sector firms do not report electricity consumed. Therefore, we use the value of total operating expenses for the firm, which includes the combined cost of fuel, electricity, repairs, and maintenance.[5] All monetary values are expressed in terms of $2004 - 2005$ Indian Rupees.

2.2 State-level Public Investment

Data on public investment have been collected from the State Finances Database of the Reserve Bank of India. We use state-level data on public expenditures (payments for accu-

[5]This could be due to informal sector firms using unauthorized or illegal sources of electricity, such as "borrowing" from a neighbor's or public power line. Reporting an aggregated number for operating expenses makes it difficult to distinguish different types of energy consumption. These costs are reported for the past-30 day reference period, which is then converted to an annual figure.

mulation of assets financed by borrowed funds) for two categories: (i) *Economic Services*, which include expenditures on transport, communications, and energy, and (ii) *Social Services*, which include expenditures on health, education, water and sanitation, and other welfare programs. The sum of these two categories is defined as *Total Development Expenditures*, and serves as a proxy for state-level public investment in our analysis. We scale each category of public expenditure by the population in each state, to obtain per-capita measures of government spending by state. To estimate the output elasticity of public investment for a firm's production function in 2010, we use average annual per-capita public expenditures at the state level for the past five years, i.e., for the period 2006-2010, to factor out any annual idiosyncratic changes to the level of public spending. This gives us an average *flow* measure for public investment.

In addition to the flow measure, we also construct a *stock* measure for public capital using the perpetual inventory method. Specifically, we use the year 2000 to pin down the initial stock of public capital, since some Indian states before 2000 were part of bigger states. The initial level of public capital stock is measured by

$$K_{G,0} = \frac{G_{I,0}}{g + \delta_G} \tag{1}$$

where $G_{I,0}$ is the flow of public investment in the initial period, g is the growth rate of public investment, and δ_G is the depreciation rate for public capital. We follow Gupta et al. (2014) and set the annual depreciation rate to 2.5 percent. The stock of public capital at the end of the time period is given by the following accumulation equation

$$K_{G,t} = K_{G,0} + \sum_{t=1}^{T} (1 - \delta_G)^t \, G_{I,t} \tag{2}$$

We compute the stock measure of public capital in 2010 by using the public expenditure flows for each year during 2000-2010 (measured at 2004-2005 prices), using the average growth rate of public investment across the sample as a lower bound to measure the initial stock. The total stock measure is then divided by the state-level population to obtain a per-capita estimate by state. Our analysis also uses several other state-level controls such as state GDP (Net State Domestic Product or NSDP), total labor force, literacy rate, dependency ratio, crime rate, and total number of enterprises. The data sources for these variables are provided in the appendix.

2.3 The National Highway Development Program

An important feature of our identification strategy is district and firm-level data on a major highway construction project started by the Government of India in 2001. In the early 2000s, India's National Highway System constituted merely 1.7% of India's total road network, yet carried 40% of the total traffic volume.[6] In fact, about a third of India's road infrastructure network consisted of single-laned roads, with a majority of the rest being low-quality two-lane highways (World Bank, 2007). To meet the needs of a rapidly expanding economy, India launched the National Highway Development Program (NHDP) in 2001. The project targeted connectivity of major ports and metropolitan cities. The NHDP upgraded 8,700 miles of roads to 4-lane highways, constructed about 900 miles of new six-lane expressways, and about 600 miles of other new national highways (Source: NHAI).

As part of the NHDP, the Golden Quadrilateral (GQ) Project aimed at improving connectivity between India's four major cities: Delhi, Mumbai, Chennai, and Kolkata. Due to delays in awarding contracts, problems with land acquisition, and zoning constraints, only 80% of the program was completed by the initial deadline in 2004, with the remaining 20% completed by 2012. In addition to the GQ, the NHDP also connected east and west India from Silchar to Porbandar, as well as north and south India from Srinagar to Kanyakumari. This North-South East-West (NS-EW) corridor upgraded about 4,400 miles of roads. Unlike the GQ, which was mostly completed by 2004, zoning problems led to massive delays with only a 4% completion rate by 2004, and 10% by 2006. These figures include overlaps with the GQ project which represented about 40% of total NS-EW construction in 2006. By 2017, however, 92% of the NS-EW corridor had been completed. Figure 6 provides a map of the GQ and NS-EW corridor, as well as markers indicating the construction of the individual sections that make up this project.

The data for this part of the analysis comes from three sources. First, we use geo-spatial data from the World Bank Urban Development Unit to identify the coordinates of the GQ/NS-EW corridor. Second, we use geo-spatial data from DIVA-GIS to match Indian districts with the GQ/NS-EW corridor. Third, data regarding the individual sections that make up the GQ/NS-EW corridor comes from annual reports of the National Highway Authority of India (NHAI). This data includes the start/stop location of a section, the highway number, length and construction cost of the section, as well as a section's start and completion date. We determine the coordinates of the start/stop locations of each section using Google Maps in order to match the sections with geo-spatial data from the GQ/NS-EW corridor. Finally, firm-level data comes from the ASI and NSSO surveys, as described above.

[6] Source: National Highway Administration of India (NHAI).

We use the district identifiers in these surveys to match the location of a firm to a district, thus determining its relative proximity to the GQ/NS-EW corridor. The appendix includes more detail regarding data sources, data preparation, the matching of different geo-spatial data, and the merging of geo-spatial data and the firm-level surveys.

Figure 7 shows districts on the GQ/NS-EW corridor by completion year. We measure "completion" as a discrete variable equal to the number of years a district has been located on a completed section of the GQ/NS-EW corridor. Figure 7 maps the years since completion in four categories: “completed before 2004”, “completed between 2004 and 2006”, “completed after 2006,” and “unfinished”. In general, Figure 7 suggests that a majority of the sections along the NS-EW corridor were not completed by 2009. Most sections along the GQ were completed by 2009, with the timing of the completion appearing random. In other words, it does not seem that one side of the quadrilateral was given construction preference over another. Since our firm-level data for formal and informal manufacturing are from 2010, we restrict the GQ/NS-EW data to range between 2001 and 2009.

2.3.1 Summary Statistics

Table 1 presents the summary statistics for firm-level characteristics for the formal and informal sectors, respectively, for 2010. Firms in the informal sector were much smaller in size (as measured by their GVA), with average capital-labor and output-labor ratios being significantly smaller than their formal-sector counterparts. For example, capital intensity (measured by the capital-labor ratio) in production was about 5 times higher for formal firms, while output per worker was higher by a factor of about 10. About 60 percent of formal sector firms were situated in urban areas, with a large majority being privately owned. About 50 percent of informal sector firms were in urban areas, with only 20 percent being registered with some government-level authority. About 70 percent of these firms were male-owned proprietary businesses.

Table 2 lists the average state-wise public development expenditures, along with its two sub-categories (social and economic services) (i) as a share of state GDP (Net State Domestic Product-NSDP), and (ii) in per-capita terms, for the period 2006-2010, for both the flow and stock measures. On average, Indian states spent about 4.9 percent of state GDP on development expenditures, with about 69 percent being allocated to expenditures on economic services (transport, communications, and energy). There is significant variation in public expenditures on development across Indian states: while the north-eastern state of Manipur spends the most, with about 13 percent of state GDP allocated to public investment, the southern state of Kerala spends the least, at about 1.3 percent. This comparison is also consistent for the per-capita measure of government expenditures. The average per-capita level of development expenditures across states between 2006-2010 was about Rs. $1,611(approximately $23 in current prices), with economic services again accounting for about 69 percent of per-capita development spending. Further, the stock of public capital represented about 37 percent of state GDP, with the economic services sub-category accounting for about 26 percent of state GDP. Figure 8 illustrates the variation in the average share of public investment spending across Indian states for the period 2006-2010.

Table 3 presents summary statistics for formal and informal sector manufacturing firms (from Table 1) that are geo-spatially matched to districts in five categories related to the GQ/NS-EW corridor in 2009, with the variables GQ and $Completion$ used as identifiers: (i) firms in districts along completed sections of the GQ/NS-EW corridor ($GQ = 1$, $Completion = 1$, column 1), (ii) firms in districts along the unfinished sections of GQ/NS-EW corridor ($GQ = 1$, $Completion = 0$, column 2), (iii) firms in districts through which the GQ/NS-EW does not pass ($GQ = 0$, column 3), (iv) firms in districts through which the GQ/NS-EW does not pass, but are within 30 miles of the corridor ($GQ = 0$, $dist < 30$ column 4), and (v) firms in districts through which the GQ/NS-EW does not pass, but are located between 30-50 miles of the corridor ($GQ = 0$, $30 < dist < 50$, column 5).

Table 3 suggests that both formal and informal manufacturing firms along completed sections the GQ/NS-EW corridor produce more output relative to firms off the corridor, as well as those along the corridor's unfinished sections. This higher firm output, in turn, is associated with larger GDP-per-capita and share of manufacturing in these districts. Moreover, production for formal and informal firms on the corridor is more capital intensive relative to firms without access to the new highway corridor. Finally, there is relatively little variation in the age structure of firms across locations.

3 Empirical Specification and Identification Strategy

The main objective of our empirical analysis is to estimate the output elasticity of public investment for the formal and informal sectors. To do this, we estimate a Cobb-Douglas production function without any *a priori* restriction on the returns to scale in production:

$$Y_{ist} = A_{ist} L_{ist}^{\alpha} K_{ist}^{\beta} \tag{3}$$

where the subscripts i refer to the firm, s to the state where the firm is located, and t denotes the time period.[7] Y_{ist} denotes the flow of output for a firm i in a given sector located in

[7] Since we use a cross-section data for the year 2010, the time subscript t denotes the year 2010. We cannot drop the time subscript at this point because we are going to refer to a previous period's average productivity in this section (for the year 1999) .

state s at time t. Similarly, L_{ist} is the labor input , private capital is given by K_{ist}, and A_{ist} represents a productivity shock. Assume that productivity at time t for a given firm i located in state s is given by

$$A_{ist} = \varepsilon_{ist} G_{st}^{\gamma} \tag{3.1}$$

where G_{st} denotes the state-level public investment, and ε_{ist} is an unobserved productivity shock specific to the firm. The specifications in (3) and (3.1) are consistent with the voluminous literature on the link between output and public investment, starting with Aschauer (1989) and Barro (1990). Taking logs and using firm-level Gross Value-Added (GVA) as a proxy for output, we can write the empirical specification as

$$\ln GVA_{ist} = \alpha \ln L_{ist} + \beta \ln K_{ist} + \gamma \ln G_{st} + \theta X_{ist} + \rho Z_{st} + \varepsilon_{ist} \tag{4}$$

In (4), output is measured by firm-level Gross Value-Added (GVA), and α, β, and γ are the output elasticities of labor, private capital, and public investment, respectively. Since the unit of observation is the firm, X is a vector of firm-level characteristics that includes age of the firm, type of ownership, industrial category (NIC 2-digit level), and geographical location (rural or urban). We use the same set of characteristics for both formal and informal sector firms, with the addition of registration status for informal sector firms. Additionally, we control for state-level variables (Z) to factor out any state-level factors other than public investment that may have an effect on the firm's output. Z includes state GDP (Net State Domestic Product or NSDP), total labor force, literacy rate, dependency ratio, crime rate, and total number of enterprises.

3.1 Econometric Issues

A common issue with the production function approach in (4) is that it may produce biased estimates of output elasticities if there exists reverse causality between the factors of production and output. Our empirical strategy addresses this concern on two fronts. First, there may exist an endogeneity problem with respect to the private inputs in production (labor and private capital). To address this issue, we use a method developed by Ackerberg, Caves, and Frazer (2015) (henceforth referred to as ACF), while also reporting results from earlier methods proposed by Levinsohn and Petrin (2003) and Sivadasan (2009) (henceforth referred to as LP-S). Second, there is a potential for reverse causality between a firm's output and public investment. For example, government expenditure might be allocated to certain states based on regional economic growth. In addition, if government spending is allocated to areas with little or no firm presence in a state, then its true effects on productivity at the firm-level may be understated. Finally, infrastructure resulting from government spending might take time to get built or may be subject to unanticipated delays, which would lead to deferred benefits not picked up by our estimates. We address these issues by using India's National Highway Development Program (NHDP) as a natural experiment to estimate the effect of exogenous shocks to public spending on firm output.

3.1.1 Endogeneity of Private Inputs

One source of endogeneity in specification (4) is the unobserved productivity shock that is observed by the firm, but not by the econometrician. This may induce the firm to choose private inputs (capital and labor) endogenously. Hence, the error term that contains the unobserved productivity shock may be correlated with the choice of private inputs. Levinsohn and Petrin (2003) develop a strategy that uses intermediate inputs to control for the unobserved productivity shock. In a nutshell, their approach uses information from an input choice equation to control for the endogenous productivity term. Unfortunately, this identification method relies on the availability of panel data, which is not available for our case. Sivadasan (2003) provides a solution to this problem. Rather than using the prior period productivity for the establishment, he uses the average productivity in the prior period for a matched industry-location combination to derive the predicted component of the current productivity shock. With repeated cross-section data, we can estimate the average productivity for a particular industry in a particular state for the previous time period, which in our case is the 1999 round for both surveys, and use that estimate in place of a particular firm's previous productivity.

One drawback of this method is that it is based on implicit timing assumptions about the employment choice of labor and materials input. Specifically, in the LP-S method, both labor and intermediate inputs are assumed to be variable inputs. However, if labor is chosen prior to other intermediate inputs, then labor should also be entering the intermediate input demand function. In other words, the firm's input choices will depend on labor inputs along with the productivity shock and capital. Ackerberg et al. (2015) build on this implicit assumption and suggest an alternative method to address this collinearity issue between labor and intermediate inputs. Their approach is based on the intuition that labor is "less variable" than materials as it takes time for firms to hire (and fire) workers. Thus, while LP-S invert the intermediate input demand functions that are unconditional on the labor input, ACF suggests inverting investment or intermediate demand functions that are conditional on the labor input. We will use both methods, LP-S and ACF, to estimate the production function in (4). We refer the interested reader to the respective papers for a more detailed discussion of the methodologies.[8]

3.1.2 Endogeneity of Public Inputs

Another source of endogeneity relates to the potential for reverse causality between a firm's output and public investment. Since the limitations of our dataset prevent us from fully addressing this issue, we use data from a natural experiment to examine the robustness of our results.[9] Specifically, we use two large infrastructure projects from India's national Highway Development Program (NHDP) – the Golden Quadrilateral (GQ) and the North-South East-West corridor (NS-EW) – as an exogenous public spending shock to areas that are located between the nodal points of the highway system. As Chandra and Thompson (2000) argue in the context of the US Interstate Highway Construction Program, while the nodal points of most large highway projects are selected endogenously, intermediate areas through which the highway passes are determined randomly. As such, a highway can be seen as an exogenous shock to areas between two nodal points. The fundamental assumption is that when a highway is built to connect two locations, the route is not specifically determined to pass through certain intermediate areas to the exclusion of others. Thus, some areas get better infrastructure not as a consequence of their economic characteristics, but merely because of where they happen to be located.

In the context of India, the NHDP provides a natural experiment that allows us to analyze the output of firms that are randomly placed along the new highway system (the GQ/NS-EW corridor). The underlying idea is that the government's decision to connect the largest cities in India affected smaller cities and villages in the country differently, depending on their location. Both the GQ and NS-EW corridors provide the most direct link between the chosen nodal cities, without being re-aligned to include some cities but not others. For example, Lucknow, the capital of the northern state Uttar Pradesh, did not benefit directly from the highway project, while Kanpur, another similar sized city in the same state, did.

The empirical strategy in this paper is to compare the outcomes of firms on completed sections of the GQ/NS-EW corridor with the outcomes of firms not near the updated highway

[8] It is important to note here that there are alternative approaches to estimating output elasticities of factors of production. For example, the cost function approach, based on duality theory, estimates a translog cost function where, in our specific case, public investment would be included as an unpaid factor of production. Direct estimation of this cost function would produce an estimate of the marginal benefit (or cost reduction) from public investment. The elasticity of public investment would then be backed out with the help of duality theory; See, for example, Lynde and Richmond (1992) and Binswanger et al. (1993).

[9] These limitations include the cross-sectional nature of our data, and the fact that informal sector firms are surveyed by the NSSO once every ten years.

system. Specifically, we estimate

$$ln(GVA_{id}) = \alpha ln(L_{id}) + \beta ln(K_{id}) + \gamma_1 GQ_{id} + \gamma_2 GQ_{id} * Completion_{id} + \theta X_{id} + \rho Z_d + \delta_s + \epsilon_{is} \quad (5)$$

where GVA, L, K, and X are as defined before, with the subscript id referring to firm i in district d in a given state. δ represents state-level dummies, and Z is a vector of district-level control variables including the log of district GDP, a district's literacy rate, rural population share, male-to-female ratio, share of population in casts or tribes, as well as a district's manufacturing gross value added (in log).

The coefficient of interest in (5) is γ_2. GQ indicates whether a firm is on the GQ/NS-EW corridor, while *Completion* indicates the number of years since the completion of a firm's nearest highway section. Intuitively, γ_2 shows the effect on firm output of being an additional year on a completed section of the GQ/NS-EW corridor, relative to firms that are not on completed sections of the highway system. We measure firm proximity to the highway system in two ways. First, $GQ = 1$ for firm i in district d if the highway system passes through the district in which the firm is located in. Second, we distinguish between firms that are (i) located in districts on the highway, (ii) located in districts off the highway but within 30 miles (geodesic distance from centroid), and (iii) located in districts between 30 - 50 miles from the highway. The control group therefore consists of firms located in districts more than 50 miles from the GQ/NS-EW corridor.[10] *Completion* measures the number of years a district's highway section has been completed prior to 2009. For example, the section from Khaga to Kokhraj in Uttar Pradesh was completed in 2005. Thus, *Completion* takes on a value of 4 (2009 - 2005) for firms near that highway section. If a district lies on a section of the highway that has been completed in 2001, *Completion* takes on the value of 8. On the other hand, it takes on a value of zero for firms located in (i) districts off of the GQ/NS-EW corridor, and (ii) districts on the GQ/NS-EW but whose sections have not yet been completed. As such, *Completion* is equal to $max[0, 2009 - completion\ year]$. This takes into account that firms in districts that have had access to better infrastructure longer may benefit differently than firms in districts whose section just recently got completed.[11]

[10] We measured GQ in several other ways with results being robust to the specification described above. For example, we defined $GQ = 1$ if a firm is located in a district whose (i) border is within 30 (50) miles from the highway, and (ii) center is within 30 (50) miles from the highway. We also used actual distance from the GQ/NS-EW corridor as a continuous variable as a replacement for GQ.

[11] If a district has more than one section and these sections were completed at different times, we tested our results for robustness using the completion date of the first completed section, as well as the completion date of the last completed section. Moreover, instead of taking the completion date, we confirmed the robustness of our results using the midpoint between start and the completion year. Using just the start year is not possible since each section was started before 2009. Hence there is no variation in *Completion*. Finally, we defined *Completion* as a dummy variable equal to 1 if a firm's district's highway section has been completed

Two important issues must be emphasized here: first, for this natural experiment to be valid, it must be the case that firms along the GQ/NS-EW corridor were not systematically different from firms that were not on the corridor *prior* to the commencement of the highway project (i.e., the highway upgrades were allocated randomly). Table 4 compares the characteristics of formal and informal manufacturing firms in 1999, two years before the highway construction project was announced, and who would eventually find themselves either on or off the GQ/NS-EW corridor.[12] The results suggest that there was no systematic difference between formal sector firms who were on or off the future highway system prior to its construction. This applies to informal sector firms as well, except for private capital and whether the firm was in a rural or urban area. We control for both these factors in our empirical specification. Second, while the intermediate areas between nodal points on the GQ/NS-EW corridor were determined randomly, new firms could have selected in to the corridor. We conduct several tests to shed light on this self-selection issue.

4 Results

This section reports the results of our empirical analysis in two parts. First, conditional on aggregate state-level public investment expenditures, we estimate the sectoral production functions in (4) to control for the endogeneity of private inputs at the firm level (capital and labor). Second, we address the potential reverse causality of public investment by using data from a natural experiment from India's NHDP. We also check for the robustness of our results by examining how the effect of public investment is distributed across firm size and age in each sector. These results are reported in Tables 5-8.

4.1 Output Elasticities

We begin our empirical analysis with an OLS estimation of the output elasticity of the private factors of production (capital and labor) and public investment for manufacturing firms in the formal and informal sectors. Tables 5 and 6 report the results of regressing firm-level GVA in each sector on the private and public inputs, along with controls at both the level of the firm and the state. Columns 1-3 in each table reports estimates from the OLS, LP-S, and ACF methods, respectively, for the flow specification of public investment, while columns 4-6 report the corresponding results for the for the stock specification. For public

by 2009. Again, results remain robust.

[12] Even though construction on the highway project did not begin until 2001, the most recent firm-level data prior to that date for the formal and informal sectors is 1999.

investment, we estimate the output elasticity both for the aggregate measure (development expenditures), as well as its sub-categories (social and economic services).

4.1.1 Formal Sector

We start with a basic OLS estimation of (4), reported in column 1 of Table 5. The output elasticities of labor and private capital for formal sector firms are about 0.78 and 0.33, respectively, reflecting the presence of increasing returns to scale in the private factors of production (note that the empirical specification does not impose any *a priori* restriction on returns to scale in the production function). As for the public input, neither the aggregated category of development expenditures, nor the sub-categories of economic and social services expenditures, have significant effects on firm-level output in the formal sector.

As mentioned in the previous section, the OLS estimates are most likely biased due to the endogeneity of private inputs in production. To address this issue, we use the strategy developed by LP-S and ACF to obtain more robust estimates of the output elasticities of the private and public inputs. Correcting for the endogeneity of private inputs alters the results significantly. For public investment, the estimated elasticities from both the LP-S and ACF methods are much larger than those suggested by the OLS estimation. Development expenditures are associated with an elasticity of about 0.08 (statistically significant), driven mainly by the sub-category of Economic Services.

Columns 4 - 6 in Table 5 presents the results from estimating the production function with government investment measured as a per-capita *stock* variable, rather than a flow. The estimated elasticities associated with the aggregated and sub-categories of government expenditure turn out to be much larger with the stock specification. For example, the output elasticity of Development Expenditures is now about 0.17, and that for Economic Services is about 0.16, indicating that the productivity benefits from the accumulated stock of public capital significantly exceed those from the flow of public investment for formal sector firms. As with the flow specification of public investment, the statistical significance of the public input in the firm's production function seems to be driven predominantly by the sub-category of expenditures on Economic Services.

4.1.2 Informal Sector

Table 6 reports the estimation results for the output elasticities of private and public inputs for informal sector firms, along with firm and state-level controls. Comparing the OLS results from Table 5, we see that informal sector firms have a higher (lower) output elasticity for labor (private capital) relative to the formal sector. As with the OLS results for the formal sector, the informal sector also exhibits increasing returns to scale in the private inputs. The productivity effect of the public input, including its subcategories, is not statistically significant.

Given the endogeneity issue with the OLS estimation, we turn our focus to columns 2 and 3. Using the ACF method (column 3), for example, we get output elasticities of 0.87 and 0.28 for labor and capital, respectively. The output elasticity of Development Expenditures is about 0.028, which is about three times smaller than the corresponding elasticity for formal sector firms, but is not statistically significant. Similarly, the output elasticity with respect to Economic Services expenditures for informal firms is lower than their formal counterparts by a factor of about two, but also remains statistically insignificant. Columns 4 - 6 of Table 6 presents the estimation of the informal sector production function, but with the stock measure of public investment. As with the flow measure, the results suggest that public expenditures have no significant impact on the output of informal sector firms.

In summary, the results reported in Tables 5 and 6 provide preliminary evidence that, on average, while public investment is positively associated with firm-level productivity in India's formal manufacturing sector, it has no systematic association with the output of informal manufacturing firms. The underlying factors driving this differential result will be the focus of the remainder of our analysis.

4.2 The GQ/NS-EW Corridor: A Natural Experiment

As mentioned before, the coefficients associated with diverse forms of state-wide government spending in Tables 5 and 6 are difficult to interpret causally, due to the potential for endogeneity associated with the public input. We use the construction of the GQ/NS-EW corridor in India between 2001-2009 as a natural experiment to identify the impact of government spending on firm-level productivity in the formal and informal manufacturing sectors. As such, we compare the output of formal and informal firms with and without access to the upgraded highway network. Therefore, we estimate (5) for the benchmark sample from section 3, as well as for the full sample excluding nodal districts that were endogenously selected to be on the highway. As in Datta (2012), we define these districts as Mumbai, Kolkata, Chennai, and Delhi, plus their suburbs Ghaziabad, Faridabad, Gurgaon, and Thane. We then address potential self-selection of firms on to the highway corridor, and finally complete the analysis with a discussion of the distributional impact of the highway upgrades.

Column 1 in Table 7 shows the effect of being located an additional year in a district with a completed GQ/NS-EW section, relative to firms not located on the upgraded highway corridor. The sample excludes firms in states for which government spending data is not available. As such, this sample is comparable to the firms included in Tables 5 and 6 in Section 4.1. By contrast, column 3 includes firms in all states minus firms in nodal districts and their suburbs, since these nodal districts are located on the GQ/NS-EW corridor by design rather than coincidence. GQ in columns 1 and 3 is simply an indicator equal to 1 if a firm is located in a district that is directly affected by highway upgrades. In columns 2 and 4 we additionally include two dummy variables: one for firms located in districts off of the upgraded highways but whose centroid is within 30 miles from the GQ/NS-EW corridor, and one for firms located in districts within 30 to 50 miles from an upgraded highway.

The results in Table 7 (columns 1 and 3) show that formal sector firms in districts along the planned route of the GQ/NS-EW corridor ($I(onGQ)$) are, on average, 9% - 10% more productive relative to firms not on the planned route. However, the completion of the respective highway upgrades does not significantly increase firm output ($I(onGQ) \times Completion$). By contrast, informal sector firms in districts along the planned highway upgrades are no more productive than their off-route counterparts. However, there is some evidence that the completion of a highway upgrade in the district of an informal firm has small negative effects on output.

In columns 2 and 4 in Table 7 we further separate firms off the highway into firms within 30 miles of the upgraded highway, 30-50 miles from the highway, and more than 50 miles from the highway (the control group). As before, formal sector firms along the planned route of the upgraded highway are more productive than their off-route counterparts, while there is no difference between informal firms on and off the highway corridor. More importantly, however, results indicate that the completion of highway upgrades significantly increases output of firms in districts within 0-30 miles, and 30-50 miles from the GQ/NS-EW corridor. Specifically, formal firms located in districts 0-30 miles from an upgraded highway section produce 2 - 4% more relative to their off-highway counterparts for every additional year since the completion of the highway . Similarly, being an additional year on an upgraded section of the highway increases output for firms in districts 30-50 miles from the highway by 3%. The trend for informal sector firms is similar, albeit the quantitative effect being a little smaller, at only around 1%. These findings support the results from Section 4.1: public investment has a larger impact on formal firms relative to informal firms in India's manufacturing sector, with geographical proximity to a completed infrastructure project being an important determinant.[13]

[13] To test the robustness of our results, we also included additional regressors such as distance to railroads, port connectivity, percentage of paved roads, and other controls that proxy for a district's infrastructure quality. We used these measures for 2001 (a census year), which marked the beginning of the GQ project. The results remain robust to the inclusion of infrastructure quality and access controls, and are available from the authors on request.

4.2.1 Self-selection

Almost half of the firms in our sample were established after the announcement of the NHDP in 2001. This causes concern if these younger firms selected to be located near the upgraded GQ/NS-EW highway system, rather than the highway being randomly assigned to their location. We examine this potential self-selection issue in two steps. First, we estimate the following regression

$$Age_{id} = \gamma_1 GQ_{id} + \gamma_2 GQ_{id} * Completion_{id} + \theta X_{id} + \rho Z_d + \delta_s + \epsilon_{is} \tag{6}$$

where Age is a firm's age, and the rest of the variables are as explained in (5). The parameters γ_1 and γ_2 will be negative if new firms indeed choose to be located near the GQ/NS-EW corridor. Columns 5 and 6 in Table 7 show no systematic evidence of younger firms selecting onto the GQ/NS-EW corridor. Particularly, for informal firms there is no statistically significant difference between firm age of treatment and the control group. Formal sector firms located near the planned route of the upgraded highway tend to be 1-6 years older relative to firms more than 50 miles from the GQ/NS-EW corridor. There is, however, evidence that more formal sector start-ups are occurring in districts 30-50 miles from an upgraded highway.

To ensure that our benchmark results are not sensitive to the potential selection of firms onto the GQ/NS-EW corridor, we re-estimate (5) for firms founded before the year 2001 (columns 7-8 in Table 7). These firms already existed before the plans of the NHDP were announced and thus were randomly selected to receive an infrastructure upgrade. The results from the benchmark case are widely robust to the exclusion of young firms (established post-2001). Formal firms located in districts 0-30 miles from an upgraded highway section for an additional year produce 5% more relative to their off-highway counterparts. Similarly, being an additional year on an upgraded section of the highway increases output for firms in districts 30-50 miles from the highway by 4%. The trend for informal sector firms is similar, albeit the quantitative effect is smaller, at around 2%.

4.2.2 Distributional Effects and Crowding Out

Figure 9 plots the coefficient γ_2 from a quantile regression of (5) for formal (blue) and informal (maroon) firms for different percentiles based on the distribution of GVA across firms. While there is not much variation in γ_2 for formal sector output across the distribution, for informal firms γ_2 is negative for (small) firms at the low end of the output distribution and positive for (large) firms at the top of the distribution. To shed some light on this result, we re-estimate

(5) using quantile regression excluding young firms that were founded after the year 2001.[14] Figure 10 plots the coefficient γ_2 of this regression in the right panel. Compared to the benchmark case (on the left) γ_2 has a slightly less negative effect on informal firm output for the smallest firms. This indicates that the self-selection of younger informal firms on the GQ/NS-EW corridor cannot account for a large part of the adverse impact of infrastructure upgrades on small informal firms.

Our next step is to understand how the productivity spillovers or complementarities generated by public investment accrue to small and large firms. Small informal firms are likely to be characterized by low levels of capital intensity. Therefore, when public investment increases, it benefits larger informal firms more (who have higher capital intensity), which in turn increases their productivity and market share, thereby crowding out production of small informal firms (*within-sector* crowding out). Further, as the summary statistics in Table 1 suggest, informal firms over all tend to have lower levels of capital intensity than formal sector firms. As such, an increase in public investment may benefit formal firms more than informal firms, further contributing to the crowding out phenomenon (*across-sector* crowding out).

To test the crowding-out hypothesis, we compare the effect of the GQ/NS-EW corridor on small informal firms in districts with few large informal firms, to its effects on small firms in districts with many large informal firms. Intuitively, production of smaller informal firms is more likely to be crowded out in districts that are characterized by more larger, capital intensive firms. If these larger firms indeed crowd out small firm production, the effect of highway completion should be less negative in districts with fewer large informal firms. Table 8 shows the effect of this treatment on small (25th percentile and 50th percentile) informal firms in districts with many large firms relative to fewer large firms. The variable $I(\# large\, firms > Mean)$ is an indicator equal to one if firm i is located in a district with many large informal firms (more than the average district). A large firm is defined as a firm whose output is above the 50th percentile in the GVA distribution. The coefficient of interest is the one associated with $I\left(\#\ large\ firms > Mean\right) \times Completion$: the difference in the output of small firms (25th and 50th percentile) in districts with many large firms relative to districts with fewer large firms, after a completed highway upgrade. Results support the hypothesis of crowding out both within and across sectors: being an additional year on an upgraded highway section has twice the negative effect on the output of small firms located in districts with many large firms (both formal and informal) relative to small firms located

[14] As start-ups are smaller in size and potentially less productive than their more established and older counterparts (e.g. because of learning by doing, or securing market share and supply chains), the negative effect of the highway upgrade might be an artifact of self-selection.

in districts with few large firms ($I\left(\# \; large \; firms > Mean\right) \times Completion + I(on \; GQ) \times Completion$).

5 Conclusions

Government investment in infrastructure goods such as roads, transportation, water and sanitation, and energy is a key element of public policy is developing countries. At the same time, these countries are, on average, characterized by a significant amount of production that takes place in the informal sector. Firms in this sector are often small, unregistered, and produce non-traded goods characterized by low capital intensity in production, relative to the formal sector. One possible way in which productivity may be influenced in this sector is through government provision of public goods such as infrastructure, which are often non-excludable in developing countries. However, very little is known about the spillovers generated by public investment for firms in the informal sector. In this paper, we use two firm-level datasets from India's manufacturing sector to estimate the output elasticities of public investment for firms in the formal and informal sector. We also use data from a major highway construction project in India as a natural experiment to provide a better causal understanding of the channels through which public investment might affect firm-level productivity.

Our results indicate that, on average, for formal firms, the output elasticity of public investment varies between 0.08-0.17, depending on whether we consider the flow or stock specification of public expenditures as the relevant input in production. For the informal sector, however, we find no systematic association between public investment and the output of the average firm, irrespective of whether we consider the flow or stock specification. In estimating these sectoral output elasticities, we use methods proposed by Levinsohn and Petrin (2003), Sivadasan (2009), and Ackerberg et al. (2015), to control for firm-level endogeneity in the usage of private factors of production. To control for the potential endogeneity of public investment, we use a large infrastructure project in India – the Golden Quadrilateral (GQ) and North-South East-West (NS-EW) corridor as a natural experiment to identify the causal effect of public investment on firm production in the formal and informal manufacturing sectors. Here, we confirm our results on the positive association between public investment and firm productivity in the formal sector, and the lack of such a systematic association for the informal sector. We also do not find any evidence that the location choices for formal and informal sector firms are driven by the level of public investment in a given state.

Why does public investment not seem to influence the productivity of informal firms? To provide some understanding of this question, we use quantile regressions to examine whether public investment has a differential impact along the size distribution of firms in each sector. Here, we find evidence that the complementarities generated by an increase in public investment lead to large firms crowding out the output of smaller firms, both within and across sectors: large firms in the informal sector tend to crowd out smaller firms within that sector and, formal sector firms also tend to crowd out small informal firms. Intuitively, large informal firms tend to have a higher capital intensity in production than their smaller counterparts, and formal sector firms also tend to have higher capital intensity than their informal counterparts over all. As such, public investment benefits not only larger firms in each sector, but also formal firms much more than informal ones. This can help explain why we are unable to find any positive and significant association between public investment and informal production for the average firm in our sample.

From a policy perspective, our results suggest that the size distribution of firms in the formal and informal sectors are an important factor in understanding how public investment affects firm-level productivity in India's manufacturing sector. Consequently, an effective way to increase the productivity and capital usage of informal sector firms might be to send more public investment goods to the largest firms in that sector. This may have the added advantage of lowering the relative size of the informal sector, by helping to formalize the largest and most productive firms, rather than a one-size-fits-all approach.

References

[1] Ackerberg, D., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica* 83 (6):2411{2451.

[2] Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics* 23 (2), 177–200.

[3] Barro, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy* 98 (5 pt 2).

[4] Binswanger, H. P., S. R. Khandker, and M. R. Rosenzweig (1993). How infrastructure and financial institutions affect agricultural output and investment in India. *Journal of Development Economics* 41 (2), 337–366.

[5] Bom, P. R. and J. E. Ligthart (2014). What have we learned from three decades of research on the productivity of public capital? *Journal of Economic Surveys* 28 (5), 889–916.

[6] Chandra, A. and E. Thompson (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics* 30 (4):457–490.

[7] Datta, S. (2012). The impact of improved highways on Indian firms. *Journal of Development Economics* 99 (1), 46–57.

[8] Devarajan, S., V. Swaroop, and H.-f. Zou (1996). The composition of public expenditure and economic growth. *Journal of Monetary Economics* 37 (2), 313–344.

[9] Ghani, E., A. G. Goswami, and W. R. Kerr (2015). Highway to success: The impact of the golden quadrilateral project for the location and performance of Indian manufacturing. *The Economic Journal* 126, 317-357.

[10] Gomis-Porqueras, P., A. Peralta-Alva, and C. Waller (2014). The shadow economy as an equilibrium outcome. *Journal of Economic Dynamics and Control* 41, 1–19.

[11] Gramlich, E. M. (1994). Infrastructure investment: a review essay. *Journal of Economic Literature*, 1176–1196.

[12] Gupta, S., A. Kangur, C. Papageorgiou, and A. Wane (2014). Efficiency-adjusted public capital and growth. *World Development* 57, 164–178.

[13] Holtz-Eakin, D. and A. E. Schwartz (1995). Infrastructure in a structural model of economic growth. *Regional Science and Urban Economics* 25 (2), 131–151.

[14] Hulten, C. R., E. Bennathan, and S. Srinivasan (2006). Infrastructure, externalities, and economic development: a study of the Indian manufacturing industry. *The World Bank Economic Review* 20 (2), 291–308.

[15] Ihrig, J. and K. S. Moe (2004). Lurking in the shadows: the informal sector and government policy. *Journal of Development Economics* 73 (2), 541–557.

[16] ILO (2013). *Women and men in the informal economy: A statistical picture*. Second Edition, International Labour Office-Geneva .

[17] La Porta, R., and A. Shleifer (2008). The unofficial economy and economic development. *Brookings Papers on Economic Activity* (2), 275–363.

[18] La Porta, R. and A. Shleifer (2014), Informality and development. *Journal of Economic Perspectives* 28, 109-126.

[19] Lall, S. V. (1999). The role of public infrastructure investments in regional development: Experience of Indian states. *Economic and Political Weekly* , 717–725.

[20] Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70 (2), 317–341.

[21] Lynde, C. and J. Richmond (1992). The role of public capital in production. *The Review of Economics and Statistics* , 37–44.

[22] McKinsey Global Institute (2013). *Infrastructure productivity: How to save $1 trillion a year*. New York: McKinsey Global Institute .

[23] Mehrotra, S., A. Gandhi, P. Saha, and B. Sahoo (2013). Turnaround in India's employment story-silver lining amidst joblessness and informalization? *Economic and Political Weekly* 48 (35).

[24] Mitra, A., A. Varoudakis, and M.-A. Veganzones-Varoudakis (2002). Productivity and technical efficiency in Indian states' manufacturing: The role of infrastructure. *Economic Development & Cultural Change* 50 (2), 395.

[25] Munnell, A. H. and L. M. Cook (1990). How does public infrastructure affect regional economic performance? *Proceedings of Conference Series no. 34*, Federal Reserve Bank of Boston.

[26] Ordonez, J. C. L. (2014). Tax collection, the informal sector, and productivity. *Review of Economic Dynamics* 17 (2), 262–286.

[27] Prado, M. (2011). Government policy in the formal and informal sectors. *European Economic Review* 55 (8), 1120–1136.

[28] Rauch, J. E. (1991). Modelling the informal sector formally. *Journal of Development Economics* 35 (1), 33–47.

[29] Schneider, F., A. Buehn, and C. E. Montenegro (2010). New estimates for the shadow economies all over the world. *International Economic Journal* 24 (4), 443–461.

[30] Schneider, F. and D. H. Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature* 38, 77–114.

[31] Sivadasan, J. (2009). Barriers to competition and productivity: Evidence from India. *The BE Journal of Economic Analysis & Policy* 9 (1).

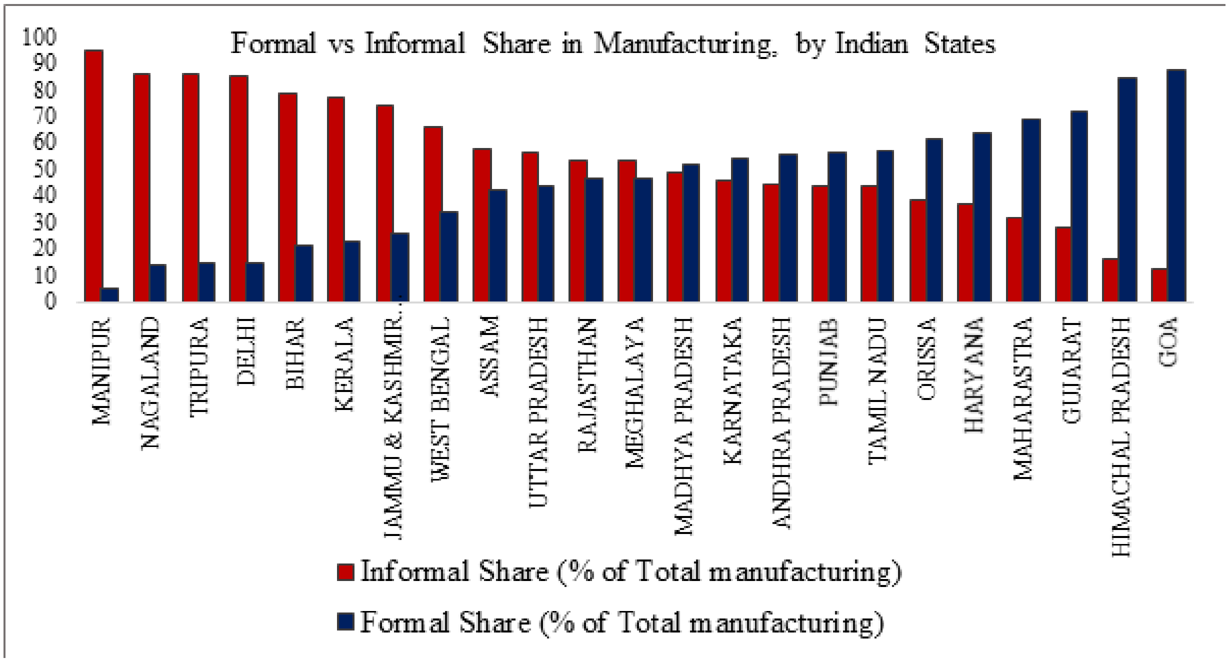
[32] Turnovsky, S. J. and M. A. Basher (2009). Fiscal policy and the structure of production in a two-sector developing economy. *Journal of Development Economics* 88 (2), 205–216.

[33] Urban Land Institute and Ernst &Young (2013). *Infrastructure 2013: Global Priorities,Global Insights*. Washington, DC: Urban Land Institute.

[34] World Bank (2007). *India's Transportation Sector: The Challenges Ahead*. The World Bank Group. Washington, DC .

[35] Zhang, H. (2004). Self-selection and wage differentials in urban china: a polychotomous model with selectivity. Massachusetts Institute of Technology, Boston, mimeo.

Figure 1. Share of Formal and Informal Production in Manufacturing: Indian States, 2010



Source: ASI, NSSO

Figure 2. Sectoral Capital Intensity: 1999 and 2010

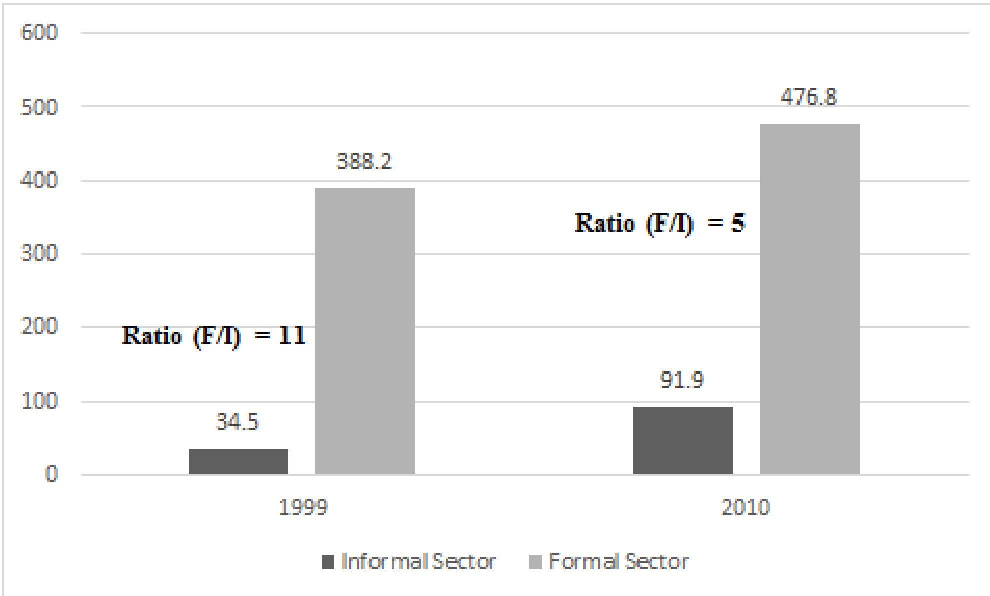
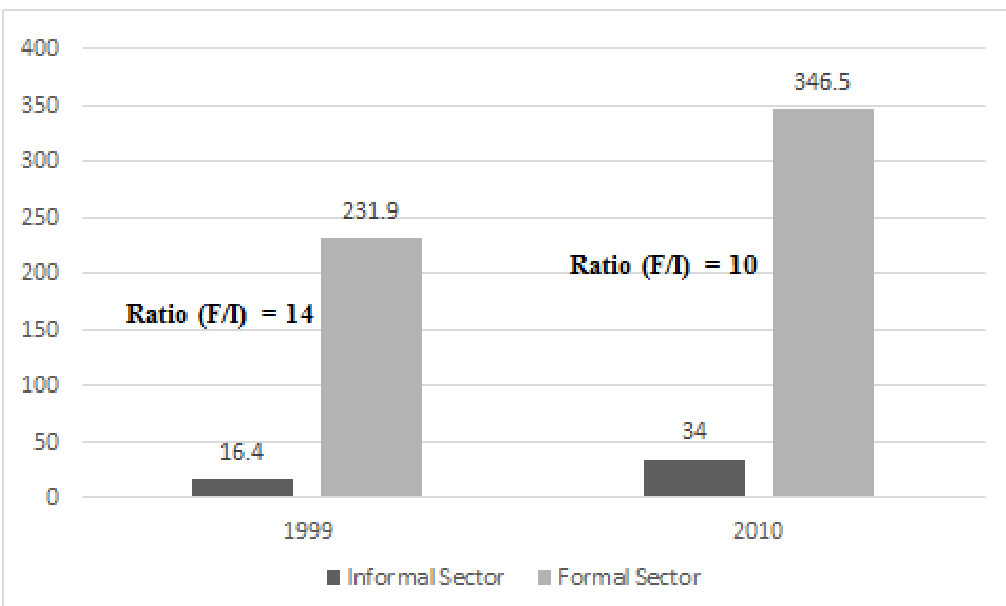


Figure 3. Sectoral Output per Worker: 1999 and 2010



Source: ASI, NSSO

Figure 4. Share of Informal Sector in GDP: India, 1999-2010

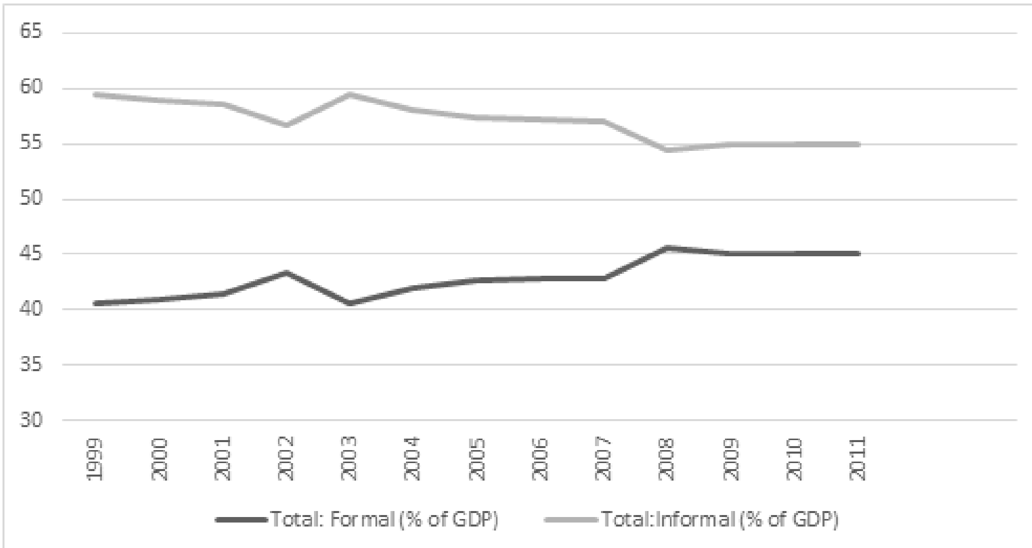
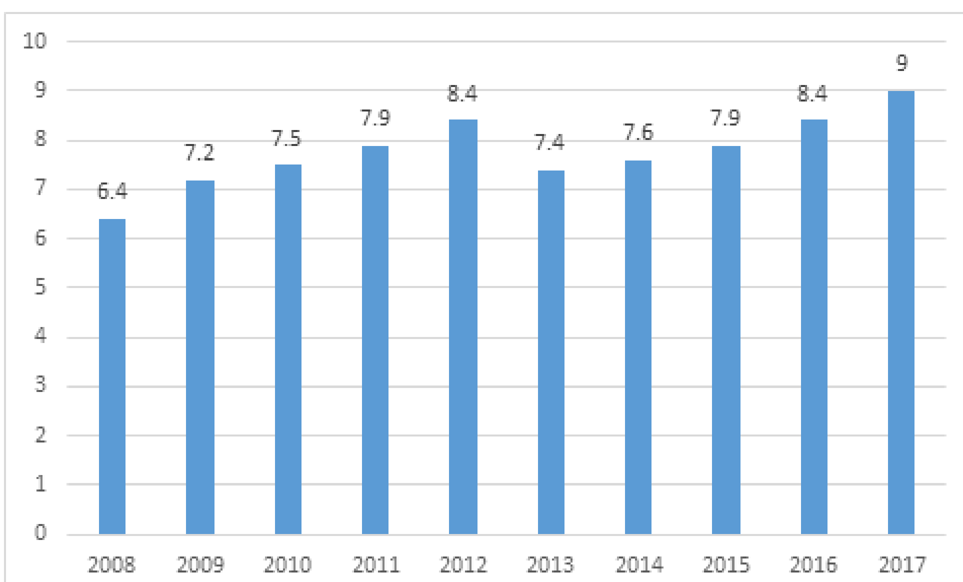


Figure 5. Share of Infrastructure Spending in GDP, 2008-2017



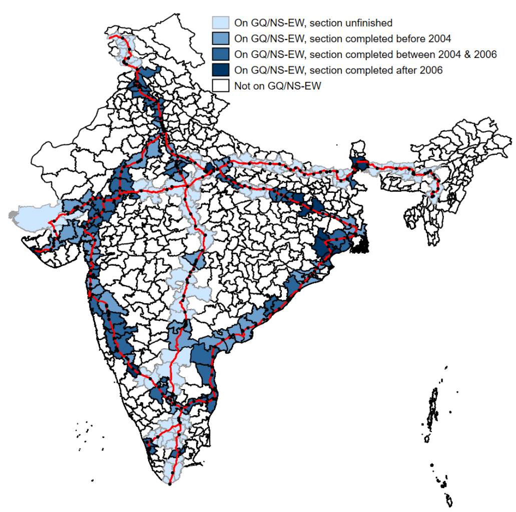
Source: ASI, NSSO, Statista

Figure 6. Map of the Golden Quadrilateral (GQ) and North-South-East-West (NS-EW) Corridor

Source: NHAI

The nodal cities of the GQ include Delhi, Kolkata, Chennai, and Mumbai. The start/end point of the NS section of the NS corridor are Srinagar and Kanyakumari, respectively. The start/end point of the EW section are Silchar and Porbandar. The black markers represent the start/end points of the individual construction sections that make up the GQ/NS-EW corridor.

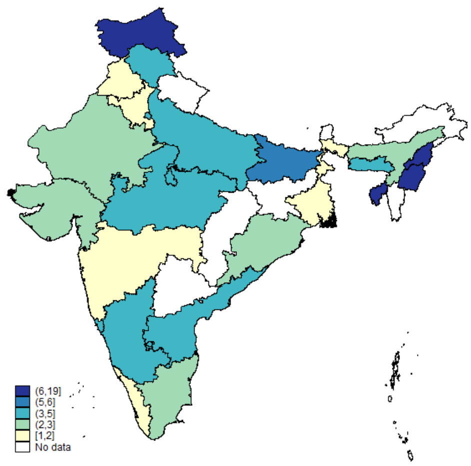
Figure 7. Map of GQ/NS-EW Corridor by Year of Completion



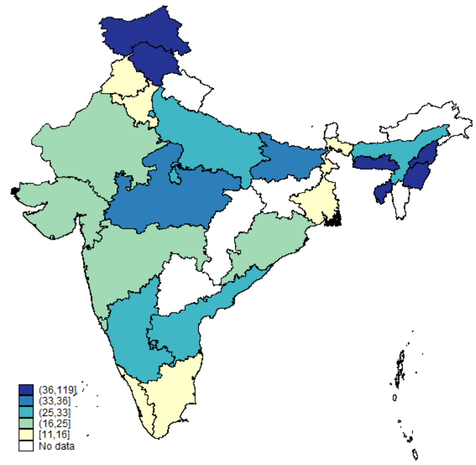
Source: NHAI

Figure 8. Average Public Investment Expenditures Across Indian States, 2006-2010

A. Public Investment (share of State GDP)

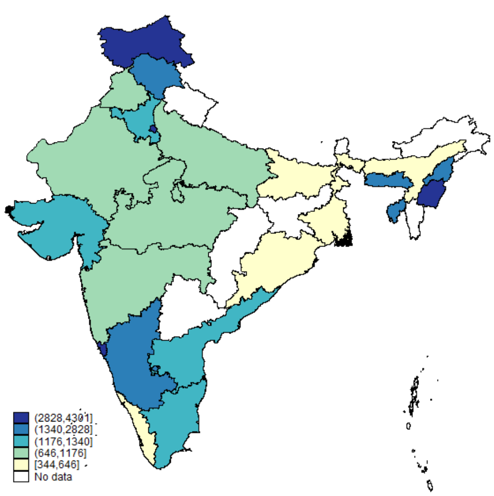


i. Flow Measure

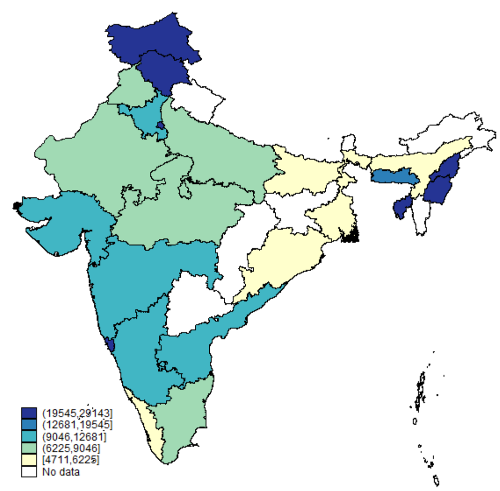


ii. Stock Measure

B. Public Investment per-capita

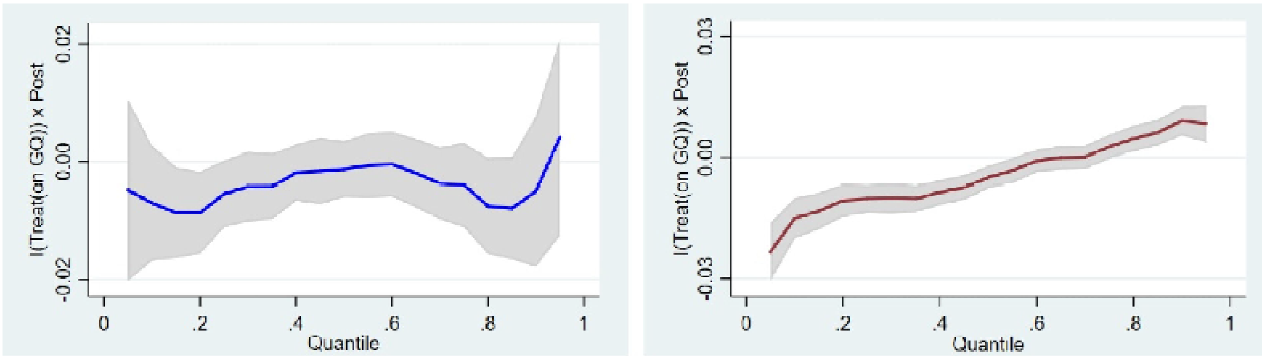


i. Flow Measure



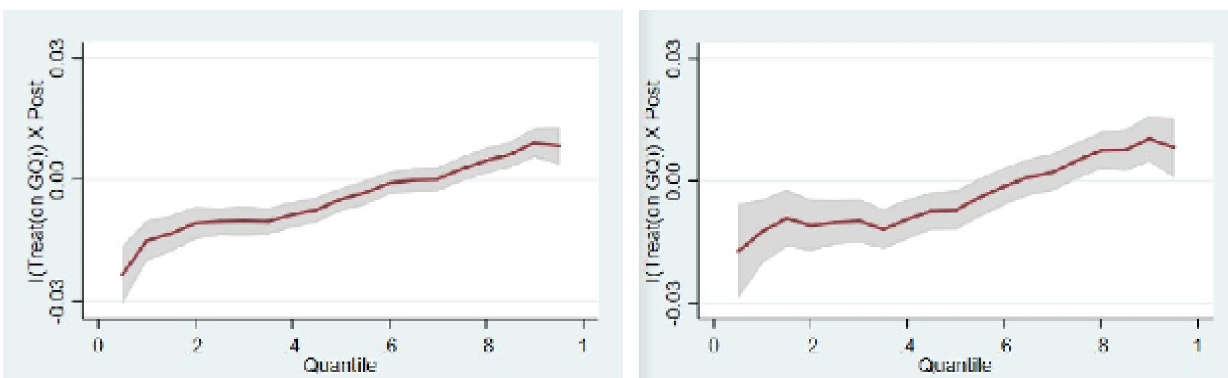
ii. Stock Measure

Figure 9. Impact of GQ/NS-EW Corridor by Size Distribution of Firms



The figure plots the coefficient $\gamma 2$ from a quantile regression of equation (5). The blue line represents the effect of being an additional year on a completed section of the GQ/NS-EW for formal firms. The maroon line represents the effect of being an additional year on a completed section of the GQ/NS-EW for informal firms. The sample excludes firms in non-nodal districts. 95% confidence bands are plotted in grey.

Figure 10. Impact of GQ/NS-EW Corridor by Size Distribution of Informal Sector Firms: Selection of Start-ups



The figure plots the coefficient $\gamma 2$ from a quantile regression of equation (5). The panel on the left is the same as the right panel in Figure 9. The panel on the right reduces the sample to exclude firms that were founded after the announcement of the GQ/NS-EW corridor in 2000. 95% confidence bands are plotted in grey.

Table 1. Summary Statistics: Formal and Informal Sector Firms in India, 2010

	Formal Sector		**Informal Sector**	
	Mean	Std dev.	Mean	Std. dev.
Gross value-added (GVA, thousand Rs.)	97603.0	677048.7	86.7	158.0
Net fixed assets (K, thousand Rs.)	169607.2	2021480.7	231.8	840.7
Total workers (L)	192.2	697.1	2.2	1.7
Capital-labor ratio (K/L, thousand Rs.)	476.8	2771.8	91.9	221.1
Output-labor ratio (Y/L, thousand Rs.)	346.5	3029.7	34.0	33.9
Rural	0.4	0.5	0.5	0.5
Age of firm	17.1	13.0	12.3	9.4
Registered under any act/authority?			0.2	0.4
Ownership				
Wholly central government	0.002	0.05		
Wholly state and/or local government	0.007	0.09		
Central government and state jointly	0.002	0.04		
Joint sector public	0.007	0.08		
Joint sector private	0.009	0.09		
Wholly private ownership	1.0	0.2		
Proprietary (male)			0.7	0.4
Proprietary (female)			0.3	0.4
Partnership with members of same household			0.02	0.1
Partnership with members of different households			0.005	0.07
Not known				
Self-help group			0.0008	0.03
Trusts			0.00007	0.009
Others			0.0001	0.01
Observations	32388		82748	

Table 2. Average State-wise Public Development Expenditures (2004-2005 INR)

	Flow measure, 2006-2010						Stock measure, 2006-2010					
	Share (% of NSDP)			Average per capita			Share (% of NSDP)			Average per capita		
States	Develop	Social	Economic	Develop	Social	Economic	Develop	Social	Economic	Develop	Social	Economic
JAMMU AND KASHMIR	13.0	4.8	8.2	3259.2	1191.5	2067.7	86.6	27.7	58.9	24831.3	7940.3	16891.0
HIMACHAL PRADESH	4.9	1.9	2.9	2008.9	784.0	1224.9	41.4	17.6	23.8	19594.7	8311.7	11283.1
PUNJAB	1.8	0.5	1.2	800.4	252.7	547.7	16.5	3.4	13.1	8486.8	1747.2	6739.6
HARYANA	2.4	0.7	1.7	1324.9	406.5	918.4	16.2	4.1	12.1	10843.5	2767.3	8076.2
DELHI	3.6	1.1	2.5	3031.7	899.3	2132.4	29.4	8.9	20.5	28272.2	8548.5	19723.7
RAJASTHAN	2.9	1.3	1.6	744.3	336.6	407.7	23.5	9.7	13.7	7390.6	3059.6	4331.0
UTTAR PRADESH	5.1	1.0	4.1	849.6	165.6	684.0	33.4	5.5	27.8	6284.0	1042.8	5241.2
BIHAR	5.6	0.7	4.9	618.1	78.9	539.1	35.9	4.5	31.4	4856.5	612.1	4244.4
NAGALAND	9.0	3.5	5.5	2828.0	1109.4	1718.7	83.3	34.5	48.8	27375.2	11334.4	16040.8
MANIPUR	19.0	7.2	11.8	4300.6	1630.7	2669.9	118.8	46.1	72.7	27782.9	10780.4	17002.4
TRIPURA	7.6	2.9	4.6	2188.0	845.4	1342.6	75.0	28.6	46.3	23749.5	9075.0	14674.5
MEGHALAYA	4.5	1.8	2.8	1435.0	553.1	881.8	47.5	18.3	29.2	16247.4	6260.1	9987.3
ASSAM	3.1	0.5	2.6	622.2	91.9	530.2	26.6	3.6	22.9	6225.4	846.4	5379.0
WEST BENGAL	1.2	0.3	0.9	344.3	81.8	262.5	14.2	2.0	12.2	4711.1	651.9	4059.2
ORISSA	2.6	0.6	2.0	645.8	149.2	496.5	21.7	5.1	16.6	6075.6	1432.2	4643.5
MADHYA PRADESH	5.3	0.9	4.4	1075.1	174.7	900.4	36.5	5.8	30.7	8280.3	1320.8	6959.5
GUJARAT	2.7	0.6	2.1	1247.1	296.9	950.2	21.5	5.8	15.7	11935.2	3239.0	8696.2
MAHARASHTRA	2.3	0.3	2.0	1176.1	139.7	1036.4	20.4	2.0	18.4	11968.2	1193.8	10774.3
ANDHRA PRADESH	3.7	0.6	3.1	1340.1	229.0	1111.1	29.8	5.3	24.5	12681.2	2267.5	10413.7
KARNATAKA	3.9	1.0	2.9	1468.8	378.6	1090.2	26.5	6.3	20.2	11592.6	2758.6	8834.0
GOA	3.7	1.0	2.7	3885.6	1028.2	2857.4	24.8	7.1	17.7	29143.2	8300.0	20843.2
KERALA	1.4	0.4	1.0	590.6	173.7	417.0	10.8	2.6	8.2	5246.2	1249.1	3997.1
TAMIL NADU	2.8	0.7	2.1	1267.6	314.1	953.5	16.5	5.3	11.2	9046.4	2913.9	6132.5
Mean	**4.9**	**1.5**	**3.4**	**1611.0**	**491.8**	**1119.1**	**37.3**	**11.3**	**25.9**	**14027.0**	**4245.8**	**9781.2**
S.D	**4.1**	**1.7**	**2.5**	**1119.6**	**432.9**	**716.7**	**27.6**	**11.9**	**16.4**	**8626.3**	**3611.3**	**5305.1**
Observations	23	23	23	23	23	23	23	23	23	23	23	23

Table 3: Descriptive Statistics for Firms for GQ/ES-NW Corridor Project

	$GQ = 1$		$GQ = 0$		
	$Completion = 1$	$Completion = 0$	Total	$Dist < 30$	$30 < Dist < 50$
Formal Sector Firms:					
Gross Value Added (Rs. '000)	116203.9	65487.5	97540.7	89213.4	82595.0
	(1928601.0)	(430852.5)	(737860.4)	(699159.1)	(553375.7)
Capital (Rs. '000)	190542.1	111030.8	185800.2	127472.8	112434.4
	(1829884.4)	(733432.8)	(2541505.5)	(1091825.1)	(615327.4)
Labor	185.7	209.4	174.1	177.8	159.8
	(468.1)	(1322.9)	(583.0)	(734.1)	(469.0)
Capital-Labor Ratio (Rs. '000)	492.0	390.0	477.3	361.5	434.3
	(2756.0)	(1256.8)	(1589.0)	(844.5)	(1278.8)
Age	17.23	16.41	16.14	15.12	17.53
	(12.86)	(12.58)	(13.21)	(12.23)	(13.70)
Rural	0.385	0.493	0.481	0.253	0.528
	(0.487)	(0.500)	(0.500)	(0.435)	(0.499)
District GDP-per-Capita (Rs. '000)	88.54	70.96	72.92	98.07	74.32
	(30.86)	(26.50)	(40.43)	(40.68)	(49.73)
District Manufact. GVA (in Bn Rs)	91.17	56.72	23.81	26.27	19.57
	(99.61)	(63.49)	(19.86)	(14.15)	(15.60)
N	**12973**	**4022**	**11778**	**1358**	**3758**
Informal Sector Firms:					
Gross Value Added (Rs. '000)	91.18	81.65	74.62	76.43	80.14
	(144.9)	(122.0)	(144.3)	(107.3)	(177.9)
Capital (in 1000)	241.9	181.7	179.8	265.6	179.6
	(784.3)	(484.4)	(674.1)	(1413.2)	(552.8)
Labor	2.268	2.227	2.035	2.080	2.153
	(1.742)	(1.693)	(1.485)	(1.473)	(1.622)
Capital-Labor Ratio (Rs. '000)	92.95	74.98	77.82	108.1	77.11
	(193.3)	(142.7)	(179.7)	(343.5)	(149.7)
Age	10.48	10.30	10.42	10.43	10.61
	(9.711)	(9.069)	(9.183)	(8.779)	(9.596)
Rural	0.474	0.529	0.582	0.564	0.579
	(0.499)	(0.499)	(0.493)	(0.496)	(0.494)
District GDP-per-Capita (Rs. '000)	75.54	49.24	51.84	50.49	52.02
	(37.51)	(24.07)	(30.49)	(44.58)	(32.30)
District Manufact. GVA (in Bn Rs)	57.15	22.50	9.858	7.036	10.85
	(77.51)	(37.85)	(13.02)	(9.716)	(11.09)
N	**23130**	**14385**	**48311**	**4797**	**14666**

Descriptive Statistics calculated from ASI (Formal Sector) and NSSO (Informal Sector). Firms are matched with geospatial data as explained in Section XXX. See data appendix for more detail. Column 1 (GQ=1, Completion=1) includes firms in districts along completed sections of the GQ/NS-EW. Column 2 (GQ=1, Completion=0) includes firms in districts along unfinished sections of the GQ/NS-EW. Column 3 (GQ=0, Total) includes firms in districts through which the GQ/NS-EW does NOT pass. Column 4 (GQ=0, Dist<30) includes firms in districts through which the GQ/NS-EW does NOT pass, but that are within 30 miles of the GQ/NS-EW (geodesic distance from district centroid to nearest section of the GQ/NS-EW). Column 5 (GQ=0, 30<Dist<50) includes firms in districts which are located 30-50 miles from the GQ (geodesic distance from district centroid to nearest section of the GQ/NS-EW). Nodal districts are excluded. These nodal districts include Delhi, Mumbai, Kolkata and Chennai and their contiguous suburbs (Gurgaon, Faridabad, Ghaziabad and Noida for Delhi; Thane for Mumbai).

Table 4. Firms on/off the GQ/NS-EW Corridor, 1999

	$GQ = 1$	$GQ = 0$	Difference	s.e.
Formal Sector Firms:				
Gross Value Added (in 1000 Rs)	55244.326	62872.532	7628.207	[4843.787]
Total Workers (L)	179.900	165.660	-14.240	[11.901]
Net Fixed Assets (K) (in 1000Rs)	126036.286	139393.466	13357.181	[21238.272]
Inputs	7960.922	8961.782	1000.859	[825.469]
Rural	0.337	0.346	0.008	[0.007]
Total Expenditure (in 1000 Rs)	13764.768	14526.208	761.440	[1168.735]
Total Receipts (in 1000 Rs)	10771.309	12539.672	1768.362	[1307.461]
Ownership:				
Wholly Central Government	0.007	0.007	-0.000	[0.001]
Wholly State and/or Local Govt	0.013	0.015	0.002	[0.002]
Central Government and State jointly	0.003	0.006	0.002**	[0.001]
Joint Sector Public	0.017	0.018	0.001	[0.002]
Joint Sector Private	0.009	0.010	0.001	[0.001]
Wholly Private Ownership	0.950	0.945	-0.005	[0.003]
N	**7,935**	**10,489**		
Informal Sector Firms:				
Gross Value Added (in 1000 Rs)	36276.347	36367.938	91.591	[1116.523]
Total Workers (L)	2.019	1.998	-0.021	[0.013]
Net Fixed Assets (K) (in 1000Rs)	72654.051	85571.744	12917.692***	[4103.787]
Inputs	88419.892	76736.748	-11683.143	[15620.301]
Rural	0.635	0.616	-0.019***	[0.005]
Total Expenditure (in 1000 Rs)	68592.687	59529.363	-9063.324	[12117.617]
Total Receipts (in 1000 Rs)	104869.035	95897.302	-8971.733	[12553.710]
Registered under any act/authority?	0.141	0.142	0.001	[0.003]
Ownership:				
Proprietary	0.983	0.983	-0.000	[0.001]
Partnership with mebers of same HH	0.012	0.011	-0.001	[0.001]
Partnership with mebers of different HH	0.005	0.006	0.001	[0.001]
N	**18,576**	**28,614**		

Table 5. Production Function Estimation: Formal Sector Firms

Sector: Formal Dependent variable: *ln GVA*	Public Investment Flow			Public Investment Stock		
	OLS	LP-S	ACF	OLS	LP-S	ACF
ln L	0.791***	0.687***	0.796***	0.790***	0.687***	0.796***
	(0.021)	(0.012)	(0.017)	(0.022)	(0.012)	(0.016)
ln K	0.334***	0.375***	0.331***	0.334***	0.375***	0.332***
	(0.016)	(0.010)	(0.012)	(0.016)	(0.010)	(0.012)
ln Development exp. per capita	0.023	0.077**	0.079**	0.120**	0.163***	0.171***
	(0.037)	(0.033)	(0.031)	(0.048)	(0.041)	(0.038)
ln Social Services exp. per capita	-0.006	0.041	0.032	0.005	0.038	0.031
	(0.036)	(0.029)	(0.028)	(0.034)	(0.027)	(0.026)
ln Economic Services exp. per capita	0.019	0.062**	0.068**	0.115**	0.143***	0.156***
	(0.032)	(0.029)	(0.028)	(0.044)	(0.039)	(0.037)
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes
State-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	30533	30533	30533	30533	30533	30533

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrap (1000 replications) standard errors (in parentheses) are clustered at the State - NIC-3 digit level.

Table 6. Production Function Estimation: Informal Sector Firms

Sector: Informal	Public Investment Flow			Public Investment Stock		
Dependent variable: *ln GVA*	OLS	LP-S	ACF	OLS	LP-S	ACF
ln L	0.820***	0.628***	0.866***	0.820***	0.628***	0.866***
	(0.017)	(0.015)	(0.025)	(0.018)	(0.016)	(0.025)
ln K	0.252***	0.317***	0.281***	0.252***	0.318***	0.282***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)
ln Development exp. per capita	-0.002	0.027	0.028	-0.020	0.024	0.024
	(0.027)	(0.031)	(0.031)	(0.052)	(0.046)	(0.044)
ln Social Services exp. per capita *ln*	-0.048	-0.026	-0.022	-0.033	-0.014	-0.011
	(0.028)	(0.032)	(0.030)	(0.030)	(0.030)	(0.030)
Economic Services exp. per capita	0.009	0.039	0.039	-0.012	0.036	0.035
	(0.027)	(0.030)	(0.031)	(0.042)	(0.047)	(0.046)
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes
State-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	82748	82748	82748	82748	82748	82748

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrap (1000 replications) standard errors (in parentheses) are clustered at the State - NIC-3 digit level.

Table 7. Impact of GQ/NS-EW corridor

	log(GVA)				Self-Selection			
	Benchmark Sample		No Nodal Districts		*Age*		*log(GVA)* of old firms	
Formal Sector Firms:								
$I(on\ GQ)$	0.10***	0.13***	0.09**	0.12***	0.57	1.47**	0.003	0.01
	(0.04)	(0.05)	(0.04)	(0.04)	(0.51)	(0.58)	(0.04)	(0.05)
$I(0 < Dist < 30)$		-0.06		-0.01		6.27***		-0.16*
		(0.07)		(0.07)		(1.52)		(0.09)
$I(30 < Dist < 50)$		-0.06		-0.06		3.23***		-0.17*
		(0.08)		(0.08)		(0.69)		(0.10)
$I(on\ GQ) \times Completion$	-0.01	-0.01	-0.01	-0.002	0.07	0.01	0.007	0.012**
	(0.01)	(0.01)	(0.01)	(0.005)	(0.08)	(0.08)	(0.005)	(0.005)
$I(0 < Dist < 30) \times Completion$		0.04***		0.02*		-1.23***		0.05***
		(0.01)		(0.01)		(0.20)		(0.01)
$I(30 < Dist < 50) \times Completion$		0.03**		0.03**		-0.06		0.04**
		(0.01)		(0.01)		(0.14)		(0.02)
N	**29923**	**29923**	**28766**	**28766**	**32849**	**32849**	**20852**	**20852**
Informal Sector Firms:								
$I(on\ GQ)$	-0.02	-0.02	-0.003	0.002	0.19	0.34	-0.02	-0.02
	(0.02)	(0.03)	(0.02)	(0.03)	(0.19)	(0.21)	(0.02)	(0.02)
$I(0 < Dist < 30)$		-0.05		-0.01		-0.16		-0.01
		(0.04)		(0.04)		(0.37)		(0.04)
$I(30 < Dist < 50)$		-0.04*		-0.03		0.09		-0.07**
		(0.02)		(0.02)		(0.19)		(0.03)
$I(on\ GQ) \times Completion$	-0.01*	-0.004	-0.01*	-0.005	0.02	0.03	-0.01	-0.01
	(0.004)	(0.004)	(0.003)	(0.003)	(0.03)	(0.03)	(0.004)	(0.004)
$I(0 < Dist < 30) \times Completion$		0.014*		0.01*		0.09		0.01
		(0.008)		(0.007)		(0.07)		(0.01)
$I(30 < Dist < 50) \times Completion$		0.01**		0.01***		0.09		0.02***
		0.005		(0.004)		(0.05)		(0.01)
N	**80985**	**80985**	**85660**	**85660**	**89686**	**89686**	**38659**	**38659**

The table presents results of estimating equation (5). The Benchmark Sample does not include the following districts: Chandigarh, Chattisgarh, Daman and Diu, D&N Haveli, Pondicherry, A&N Islands, Arunachal Pradesh, Mizoram, Sikkim Lakshadweep, Jharkhand, and Uttaranchal as they were excluded from Table 4 - 6 due to lack of public spending data. Columns 3 & 4 include all states, but exclude nodal districts due to endogeneity concerns. Columns 5 & 6 estimate equation (6) for all states, excluding nodal districts. Finally, Columns 7 & 8 estimate equation (5) for all states excluding nodal districts for firms that existed before the announcement of the GQ/NS-EW corridor in 2000. All regressions include labor and capital (in logs), firm and district level controls as explained in Section 2, as well as state and industry dummies. Standard errors are in parenthesis and clustered at the State-NIC3 digit level. ***, **, and * denote 1%, 5% and 10% significance level, respectively.

Table 8. Within and Across Sector Crowding-Out of Small Firms

	Large Informal Firms		Large Formal Firms		Formal and Informal	
	25th P-tile	50 P-tile	25th P-tile	50 P-tile	25th P-tile	50 P-tile
Informal Sector Firms:						
$I(on\ GQ)$	0.00	-0.01	0.01	0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$I\left(\#\ large\ firms > Mean\right)$	0.25***	0.21***	0.00	0.00	0.13***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$I(on\ GQ) \times Completion$	-0.01***	-0.01***	-0.01***	-0.01***	0.00	0.00
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
$I\left(\#\ large\ firms > Mean\right) \times Compl$	-0.01***	-0.002	-0.01***	0.001	-0.02***	-0.01***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
N	**85660**	**85660**	**85660**	**85660**	**85660**	**85660**