Three Essays on Local Economic Development in India

Three Essays on Local Economic Development in India

A dissertation presented

by

Samuel Edward Asher

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

May 2013

Dissertation Advisor:
Professor Andrei Shleifer

Author:
Samuel Edward Asher

Three Essays on Local Economic Development in India

Abstract

This dissertation examines the determinants of local economic and political development in India. In the first chapter, I study the impact that agricultural income shocks have on the local nonfarm economy. I find that positive rainfall shocks induce significant employment growth, not in the rural areas where agricultural production takes place but in the nearby towns. Manufacturing firms in particular respond to changes in agricultural production. Further investigation suggests that the most likely mechanism is a capital channel by which local agricultural surplus funds investments in urban manufacturing. In the second chapter, I examine the relationship between natural resource wealth and political outcomes. The interaction of mineral deposit locations and global price changes provide exogenous variation in the value of mineral wealth of state legislative assembly constituencies in India. I find that margins of victory, incumbency advantages and politician criminality are increasing in local mineral wealth. I test three channels for the criminality effect: (i) greater criminality in office; (ii) adverse selection of politicians into the political system; and (iii) greater success of criminal candidates in elections, finding the strongest evidence for the third effect. Finally, in the third chapter, I evaluate the importance of transportation costs to rural economic development. I take advantage of the allocation rules of a large-scale road construction program in India to estimate the impact of village roads on nonfarm economic activity. I find that new paved roads lead to large increases in village employment. Roads lead to an increase in firm size, suggesting that firms are inefficiently small when transport costs are high. Further, I find evidence that roads are most effective in the presence of electricity, suggesting complementarities between infrastructural investments.

Contents

List of Tables

List of Figures

Acknowledgments

This dissertation, while assembled hurriedly, was in fact very long in the making. In looking back, I received so much support that it is difficult to know where to start thanking people. To my sage advisers Asim Khwaja, Sendhil Mullainathan and Andrei Shleifer: your (usually) subtle blend of intellectual guidance, encouragement and urgency was invaluable in turning the many corners of the path from ideas to projects to papers. To all of my other professors and intellectual mentors, too numerous to enumerate: you have my eternal gratitude for your open doors, academic insight and personal support. My parents were always present and never pressuring. My siblings were always there to make sure my sense of self never got too high or too low. My brilliant classmates were all at once comrades in arms, sparring partners and role models. To all my friends and lovers: many thanks and much respect for accepting my absences and inconsistencies over these last six years. Mr. PC Mohanan of the Indian Ministry of Statistics and Programme Implementation was ever ready to help with the Economic Census. Finally, I give the greatest thanks to my coauthor and partner in crime, Paul Novosad, the single person without whom this dissertation would not have happened, whose companionship and good cheer made this process (almost) downright enjoyable.

Chapter 1

The Impact of Agricultural Output on Local Economic Activity[1]

1.1 Introduction

What consequences do sector-specific shocks have on other local economic activity? In a highly integrated world, with few barriers to the reallocation of resources, the answer should be little to none. Of course, the world is still characterized by limited mobility in many goods, both intermediate and final. These frictions apply both to productive inputs, such as labor (Behrman, 1999), capital (Banerjee, 2003; La Porta et al., 2002) and information (Jensen, 2007), as well as in output markets (Limao and Venables, 2001; Donaldson, 2012). A large "big push" literature shows how the decisions of firms can depend on the choices of other local firms (see, for example, Murphy et al. (1989)). Immobilities increase these interdependencies, meaning that the location decisions of firms will be determined more by the availability of inputs and local demand in low-income countries than in richer nations (Venables, 2005).

In this paper, we test for the impact of short-term agricultural output shocks on nearby firms in unrelated sectors in India. We utilize exogenous variation in agricultural output due to weather to estimate the dependence of firms on the local rural economy. In constrast to

[1]Co-authored with Paul Novosad.

recent work on the United States (Hornbeck and Keskin, 2012), we find that strong evidence that agricultural shocks have large and persistent consequences for local firms, particularly manufacturing firms in nearby urban areas.

We consider India to be an optimal setting for this research. By taking advantage of within-country variation in agricultural output, we are able to eliminate immobilities of goods and labor that exist at national borders. Furthermore, India is a large developing country with many of the constraints faced by other low-income countries. Finally, India is a large country with considerable heterogeneity in local economic conditions, such as infrastruture and industrial composition, that allow us to determine the characteristics of industries and locations that make shocks more or less important for local economic activity.

We provide evidence to differentiate between two major channels by which agricultural shocks may induce growth in nearby firms. In the first, local demand is increasing in agricultural income, inducing employment growth in nearby firms producing imperfectably tradable goods. This channel was explored at length by Mian and Sufi (2012), among others, who argue that increases in local unemployment in the United States following the financial crisis in 2008 were closely tied to shocks to household balance sheets and demand. Our findings support a somewhat different story. We find that increases in agricultural output cause significant employment growth in urban manufacturing firms, more than in nontradable industries such as retail or other services. We argue that this is evidence for immobilities in either inputs or outputs. A local demand channel does not fit our findings well: areas wtih low transportation costs exhibit weakly greater growth in manufacturing compared to areas where transportation costs are greater. If manufacturing employment were growing in response to increases in local demand, we would expect to see the greatest response to agricultural income in locations with high transportation costs. Instead, a capital channel best explains our results. A positive shock to agriculture causes an increase in employment in firms reporting a reliance on local, informal capital. This is consistent with recent evidence suggesting that local sources of savings matter even for financial access in the United States (Gilje, 2012), although credit constraints are widely considered to be more binding in developing countries. Reasons for this include contracting (Banerjee, 2003) and political interference (La Porta et al., 2002).[2]

Our findings support an intermediate level of localization of capital markets. Specifically, we reject both autarky and perfect mobility of goods. Changes to manufacturing employment in local urban areas suggest that capital flows occur between them and the rural areas in which agricultural shocks occur. This finding implies that it is essential to consider nearby urban areas when assessing the impact of agricultural performance on the non-farm economy and structural transformation, in contrast to the rural focus of much of the literature (Foster and Rosenzweig, 2004; Haggblade et al., 2010; Lanjouw and Lanjouw, 2001).

Finally, we contribute to a large literature on the role of agriculture in development. We provide evidence that weather shocks promote structural transformation: the number of workers in agriculture decreases following positive rainfall shocks. However, we find no evidence of permanent migration. Theoretically, the impact of shifts in agricultural output on the nonfarm economy are unclear: on the one hand, increases may stimulate demand, provide capital for investment and supply inputs to the non-agricultural economy. On the other hand, increased agricultural productivity increases the returns to labor and capital in agriculture, potentially crowding out other activities. Empirical evidence reflects this ambiguity. Some researchers have found that agricultural productivity gains facilitate economic growth and development (Gollin et al., 2002; Nunn and Qian, 2011), while others have found that they crowd out non-agricultural activities (Foster and Rosenzweig, 2004).[3] We should be clear that our shocks are best understood as windfalls, rather than the permanent shifts in agricultural productivity typically discussed in the literature. More closely related to this paper, and consistent with our results, Dercon (2004) provides evidence that temporary rainfall shocks can produce persistent differences in consumption

[2]Burgess and Pande (2005) find significant decreases in rural poverty as the result of a major banking expansion in India.

[3]For a comprehensive survey of the theory and evidence, see Douglas Gollin's chapter in the Handbook of Agricultural Economics (Gollin, 2010). For a more policy-focused survey, see the 2008 World Development Report: Agriculture for Development (World Bank, 2008).

in Ethiopian villages.

This paper proceeds as follows: Section 1.2 develops a theoretical framework to organize our examination of the relationship between agricultural productivity shocks and employment in the nonfarm economy. Section 1.3 explains the various sources of data used and how the key variables are constructed. Section 1.4 lays out the empirical strategy that is used to identify the impact of agricultural income on nonfarm employment, also providing summary statistics of the main variables. Section 1.5 presents the primary results. Section 1.6 concludes.

1.2 Theoretical framework

In this section we develop a theoretical framework for understanding how temporary shocks to agricultural productivity will affect firms in the non-farm economy. The purpose is to demonstrate how, depending on mobility of inputs and output, agricultural productivity shocks can either crowd-out or crowd-in employment and production in nonfarm sectors. Section 1.2.1 sets up the model, Section 1.2.2 derives the predictions that we will explore empirically and Section 1.2.3 discusses the implications.

1.2.1 Setup

Consider a small open economy (say, a village) with two productive sectors: agriculture and other. Each sector is modeled as a single, profit-maximizing firm. Agricultural production utilizes only labor, such that the equation representing agricultural profits is the following:

$$\Pi_A = \theta ln(L_A) - wL_A,$$

where $\theta > 0$ is stochastic and represents temporary weather-related agricultural productivity shocks.

Production of the other sector, which we call tradable, is CRS Cobb-Douglas in capital and labor, such that tradable profits are represented by the following equation:

$$\Pi_T = AK_T^{\alpha} L_T^{1-\alpha},$$

where capital share $\alpha \in (0,1)$ with permanent TFP parameter A.[4]

Agents are modeled as a single unitary household with Cobb-Douglas preferences over consumption of agricultural and tradable goods:

$$U(C_A, C_T) = C_T^{\gamma} C_A^{1-\gamma},$$

ensuring that that expenditure on consumption of tradables $p_T C_T$ is a constant share γ of total consumption $\overline{C}$. The price of agricultural goods is the numeraire, yielding:

$$C_A = (1-\gamma)\overline{C}.$$

For simplicity, intertemporal allocation of consumption is decided by a fixed savings rate s, meaning that current consumption is a fixed proportion of current income:

$$\overline{C_t} = (1-s)I_t.$$

Income not saved is added to the local capital stock $\overline{K}$, producing the following equation of motion for capital:

$$\overline{K_t} = sI_{t-1} + \overline{K}_{t-1}.$$

Labor is immobile across space, but not across sectors. The total labor supply available in the village is normalized to 1; therefore $L_A + L_T = 1$. Given labor mobility across sectors, the wage w must be equal in agriculture and the tradable sector.

We consider two market frictions that determine how shocks to the agricultural sector

[4] We model the economy with two sectors, although the analysis remains fundamentally similar if a third, nontradable sector is included in the analysis. Instead, in modeling transport costs, we are able to consider how the effect of temporary agricultural productivity shocks varies based on the tradability of the sector.

will affect the nonfarm economy. The first, transport costs, are modeled as iceberg costs $\tau \geq 1$: for one unit exported to the outside market, $\frac{1}{\tau}$ arrives at its destination. For simplicity, these transport costs apply only to the non-agricultural good, although applying them to agriculture does not fundamentally change the analysis. The price of the traded good in the outside market is p_T. Outside firms are indifferent between selling locally or to the outside market, thus fixing the local price of the tradable good at τp_T. Likewise, local producers selling to the outside market face an effective price of $\frac{p_T}{\tau}$.

Transport costs for capital, long a concern for development economists (Besley, 1995), are the other friction. These create a wedge between the local and outside opportunity cost of capital. As with τ above, $\phi \geq 1$ represents the cost of moving one unit of capital into or out of the local market. The interest rate r in the outside market applies to both savings and borrowing. Thus the cost of capital faced by firms borrowing locally is $\frac{r}{\phi}$, while the cost of capital borrowed from the outside market is ϕr.

Frictionless markets in capital and tradable goods is achieved by setting $\tau = \phi = 1$. For τ or ϕ greater than one, there is a wedge between the inside and outside prices of goods and capital, changing the impact that agricultural productivity shocks have on the tradable sector.

1.2.2 Production maximization problem

Solution of the model hinges on the source of marginal demand and marginal capital for the tradable industry. Due to the transport costs of tradable goods τ, the firm will first sell tradable goods to local market at τp_T. Only once local demand is fully satisfied will the firm sell to the outside market at effective price $\frac{p_T}{\tau}$. Likewise, due to capital transport costs ϕ, the firm will first rent local capital at $\frac{r}{\phi}$ before renting outside capital at interest rate ϕr. Thus there are four scenarios that determine equilibrium production and the relationship between production in the two sectors. In the first, marginal production is sold to the outside market at the lower effective price $\frac{p_T}{\tau}$ and marginal capital is borrowed at the higher rental rate ϕr. In the second, marginal demand for tradable goods is still from the outside market, but marginal capital is rented locally. In the third scenario, the opposite is true: marginal demand is local while marginal capital comes from the outside market. Finally, the fourth scenario is de factor autarky: no tradable goods are sold to the outside market, nor is capital borrowed from the outside.

Regardless of the which scenario the economy is in, profit maximization in the agricultural sector yields $L_A^* = \frac{\theta}{w^*}$; combined with the time budget constraint, we get $L_T = 1 - L_A = 1 - \frac{\theta}{w^*}$. Thus it is easy to see the first effect that transitory shocks have on nonfarm employment: in the absence of other forces, a high realization of θ will raise demand for labor in the agricultural sector, crowding out employment and production in the tradable sector.

Outside demand, outside capital

In the benchmark case, the tradable firm sells its production both locally and to the outside market, as well as using capital sourced locally and from the outside. As prices for its output are higher locally, it fulfills all local demand before selling remaining output to the external market. As the rental rate of capital is lower internally, the firm rents the full local capital stock $\overline{K}$ before renting additional capital from the outside market. The relevant prices that determine the scale of production are the price faced at the margin, $\frac{p_T}{\tau}$, and the rental rate of capital faced at the margin, ϕr.

Given CRS production in the tradable sector, zero marginal profits fix the only equilibrium wage for which there will be positive sales to the outside market:

$$w^* = (1-\alpha)(\frac{p_T A}{\tau})^{\frac{1}{1-\alpha}}(\frac{\alpha}{r\phi})^{\frac{\alpha}{1-\alpha}}.$$

The time budget constraints produces a solution for employment in the tradable sector:

$$L_T = 1 - L_A = 1 - \frac{\theta}{w^*}. \tag{1.1}$$

It should be noted that transport costs in both the goods and capital markets depress the wage, leading to higher employment in agriculture and lower in the tradable sector.

How does employment in the tradable sector respond to agricultural productivity shocks? Differentiating Equation 1.1 by θ yields:

$$\frac{\partial L_T}{\partial \theta} = \frac{-1}{1-\alpha}\left(\frac{\tau}{p_T A}\right)^{\frac{1}{1-\alpha}}\left(\frac{r\phi}{\alpha}\right)^{\frac{\alpha}{1-\alpha}} < 0.$$

Thus in our benchmark case, a high realization of θ unequivocally crowds out employment in the tradable sector. Both goods and capital transport costs amplify the crowd-out. The case of frictionless markets, with $\tau = \phi = 1$, is a particular example of this scenario, which leads to the following proposition:

Employment in the tradable sector is strictly decreasing in θ under either of the following conditions:

- Frictionless markets, defined as iceberg costs of goods (τ) and of capital (ϕ) both equal to 1.
- The firm sells positive output to the outside market and borrows positive capital from outside lenders.

Outside demand, local capital

In this scenario, the tradable sector still sells a positive amount of output to the outside market after meeting all local demand; however, it now finds itself credit constrained. At the local capital rental price, it demands more capital than the local capital stock ($K_T^*(\frac{r}{\phi}) > \overline{K}$), but it is not profitable for the firm to borrow capital from the outside market ($K_T^*(\phi r) = 0$). [5] The firm thus chooses L_T subject to $K_T = \overline{K}$. This maximization yields:

$$L_T^* = \overline{K}\left(\frac{(1-\alpha)Ap_T}{w^*\tau}\right)^{\frac{1}{\alpha}},$$

[5] In fact, another scenario is possible where the firm uses only local capital but does not find itself credit constrained. This is when the level of capital demanded at the local rental rate of $\frac{r}{\phi}$ is less than the local capital stock $\overline{K}$. This scenario becomes the same as in Section 1.2.2, but where the rental rate ϕr is replaced with $\frac{r}{\phi}$.

where w^* satisfies the following equality:

$$1 = \frac{\theta}{w^*} + \overline{K}\big(\frac{(1-\alpha)Ap_T}{w^*\tau}\big)^{\frac{1}{\alpha}}.$$

It is easy to see that even with capital constraints, tradable employment and production is crowded out by high realizations of agricultural productivity, which pushes up the wage. Because the tradable sector is dependent on local capital, however, its scale of production is increasing in $\overline{K}$, which is itself increasing in past realizations of θ. We thus have under this scenario that current agricultural productivity crowds out tradable production, while past agricultural productivity crowds in tradable production.[6]

Local demand, outside capital

In this scenario, the tradable sector meets local demand but finds it unprofitable to export to the outside market. It does, however, borrow capital from the outside market. In this equilibrium, it minimizes costs subject to meeting local demand C_T. This leads to the following employment in the tradable sector:

$$L_T^* = \overline{K}\big(\frac{(1-\alpha)Ap_T}{w^*\tau}\big)^{\frac{1}{\alpha}},$$

where w^* solves the following equality, coming from the fixed supply of labor:

$$1 = \frac{\theta}{w^*} + \overline{K}\big(\frac{(1-\alpha)Ap_T}{w\tau}\big)^{\frac{1}{\alpha}}.$$

Current agricultural productivity has countervailing effects on employment in the tradable sector. On the one hand, it lowers tradable employment by increasing returns to labor in agriculture and the equilibrium wage; on the other hand, it increases income and thus local demand for tradable goods. The total effect is ambiguous. Past agricultural

[6]In fact, the capital stock has countervailing effects on tradable employment. Higher local capital stock increases returns to labor in the tradable sector, increasing employment. However, it also increases the wage, lowering tradable employment. By differentiating the household time budget constraint with respect to the capital stock, it is easy to show that the former effect is strictly greater.

productivity has an unequivocal positive effect on tradable production by increasing the capital stock and thus current income. This scenario is characterized by a demand constraint on tradable production, leading to an increase in tradable production and labor.

Local demand, local capital

In the final scenario, the tradable sector neither sells to the outside market nor does it borrow capital from the outside market. There are actually three possibilities here. In the first, the firm demands less capital than is available at the local rental price ϕr and produces less than is demanded at the local price $\frac{p_T}{\tau}$. This case is equivalent to Section 1.2.2 but where the relevant prices are ϕr and $\frac{p_T}{\tau}$. We know from Section 1.2.2 that in this case, tradable employment and production are decreasing in agricultural productivity. In the second possibility, the firm is not constrained by local demand but is constrained by local capital $\overline{K}$ and finds it unprofitable to sell to the outside market; this is analogous to Section 1.2.2 but with marginal revenue τp_T. In the third possibility, the firm is constrained by local demand but not by the local supply of capital; this is analogous to Section 1.2.2 but with marginal price of capital $\frac{r}{\phi}$.

Finally, the firm may be constrained by both demand and capital, without finding it worthwhile to produce for the outside market or borrow additional capital. In this case, the firm simply employs enough labor to meet local production:

$$L_T^* = (\frac{C_T}{A\overline{K}^{\alpha}})^{\frac{1}{1-\alpha}}.$$

The wage is then set to satisfy the time budget constraint. Employment in the tradable sector is increasing in agricultural productivity via demand for tradable goods. Past agricultural productivity, however, produces competing effects through its impact on the capital stock. A greater capital stock increases capital earnings and thus demand for tradable goods. However, it also crowds out demand for labor, as labor in the tradable sector will only be sufficient to meet local demand in this scenario.

1.2.3 Discussion

In the absence of market frictions that limit the mobility of inputs to or outputs from the non-agricultural sector, positive agricultural shocks crowd-out production and employment by increasing competition for common inputs. Even in the presence of frictions in input and output markets, crowd-out may still occur if the marginal source of either inputs or demand comes from outside the local area and is thus independent of local agricultural productivity shocks, such that an increase in agricultural labor demand reduces the use of labor in the tradable sector.

However, for any firms that on the margin depend on local inputs or demand, such shocks can crowd-in rather than crowd-out non-farm economic activity. In the model presented above, costs associated with the movement of inputs and outputs play analogous roles. This is due to the dual effects of agricultural productivity shocks: they affect demand via earnings and capital supply via savings.

In order to test between these channels, it is necessary to identify the determinants of these frictions. Firms will be more likely to depend on local marginal demand in the presence of high transportation costs. These costs may result from characteristics of the industry, such as a high weight-to-value ratio, or of the location, such as low quality road infrastructure. Costs associated with outside borrowing, on the other hand, will depend on the availability of financial services and the credit intensity of an industry. These location and establishment characteristics will provide the basis for the empirical tests discussed in the following sections.

For the interpretation of the results presented in this paper, it is necessary to consider how transitory output shocks differ from the long-term growth in agricultural productivity more often considered in the literature. There are two primary differences. The first is that in terms of permanent income, a one or even five year period of good rain is much smaller than a permanent increase in agricultural productivity of the same annual magnitude. Second, weather-induced output shocks should not change expectations of long-run income. How do these differences matter in the framework discussed above? A permanent shift in agricultural productivity would lead to a permanent shift in local demand, leading to a level expansion in the size of the tradable sector so long as marginal demand is local. A temporary shock, on the other hand, will in the absence of market frictions lead to a much smaller increase in the long-run size of the nontradable sector, as demand in the long-run will depend only on the portion of the shock that was saved and continues to provide income to households many years after the shock. In a standard dynamic savings model, current consumption should not increase much as expectations over lifetime earnings will have changed only slightly. Thus, short-term shocks should result in larger changes to savings and smaller changes to consumption than equivalent changes in long-term agricultural productivity.

1.3 Data and construction of variables

1.3.1 Data

In order to examine the relationship between agricultural production and business activity, it was necessary to link datasets containing information on firms, location characteristics, agricultural production, land use, irrigation status and weather.

The Indian Economic Census is a comprehensive enumeration of all firms not engaged in crop production, both formal and informal. We use firm-level data from the 3rd, 4th and 5th rounds, undertaken respectively in 1990, 1998 and 2005.[7] The Economic Census contains a small number of characteristics about each firm, including the number of employees and some of their characteristics, the firm's source of power, details about the firm's registration, and the industrial code of the primary product.[8]

The Indian Population Census provides village and town demographic data in 1991 and

---

[7]These data are publicly available from the Ministry of Statistics and Programme Implementation (MoSPI), but are not organized as a panel. With the assistance of location keys from MoSPI, we constructed panels at the village, town, district and subdistrict levels, then linking them to population census identifiers.

[8]It is worth noting that this includes privately owned establishments, state-owned establishments and government establishments like schools and health centers.

2001, as well as local public goods (roads, electricity, schools and hospitals), distances from villages to major towns, and land area. We obtained geographic coordinates for population census locations from ML Infomap and matched them to weather data (described below) and to bounding polygons of administrative units. All population and economic census data was then aggregated to 1991 Population Census subdistrict level for rural areas and the district level for both urban and rural areas. For all results coming from the Economic Census, we measure employment growth as change in employment from 1990-98 and 1998-2005.

Agricultural data comes from a variety of sources. District-level data on area, production and prices by crop, 1990-2005, comes from the Center for the Monitoring of the Indian Economy and from the Indian Ministry of Agriculture. Subdistrict-level cropping patterns and irrigation data come from the Indian Agricultural Census of 1995, 2000, and 2005.[9] The Population Census also provides area under cultivation, by irrigation status, in 1991 and 2001.

Weather data for 1971-2005 comes from the Indian Meteorological Department's National Climate Centre (Rajeevan and Bhate, 2008; Srivastava et al., 2009). This dataset uses daily recordings from 6076 (rainfall) and 395 (temperature) weather stations to produce a 0.5° x 0.5° grid. We then interpolate weather values to match the geographic units used in our analysis. For subdistricts and districts, we obtain weather values using an average of village values weighted by agricultural land.

Industry measures come from a variety of sources, summarized by Table 1.7. Following Moretti (2010), we classify as tradable all industrial sectors that fall under the manufacturing header (Section D) of the 2004 National Industrial Classification of India. We use a cross section of the Indian Annual Survey of Industries (ASI) from 1991-1992 to generate two industry-level measures, both scaled to be between 0 and 1.[10] First, we calculate capital intensity as the ratio of gross fixed capital to labor costs. Second, we calculate reliance on

[9]Data was scraped in the spring of 2012 from dacnet.nic.in.

[10]This was the earliest year available to us that fell in the period for which we have outcome data (1990-2005).

external finance as the ratio of open loans to gross fixed capital. This second measure is in the spirit of Rajan and Zingales (1998), whose measure is highly correlated with our own but limited to just 19 industries that we are able to match with the Indian National Industrial Classification (NIC). For tradability, we match commodity-level data from the United States Census Commodity Flow Survey (CFS) of 2007 to NIC codes, providing a measure for each industry of the average shipping distance of its major commodity, a direct measure of the transportability of goods given high quality of infrastructure, as found in the United States. Our location-level measure of transportation costs estimates the cost of shipping goods on the Indian highway network. Using both the data and methodology of Lall et al. (2004), we estimate transportation costs from the district centroid to the nearest city with a population of at least 500,000, according to the 2001 Population Census. We consider this a much stronger measure of the relevant transportation costs in a given area than widely used alternatives, particularly the quality of rural roads, since the transport costs that will determine whether demand for tradable goods in urban areas is met by local or outside firms is the cost of moving goods in and out of the district, not from the local town to the agricultural areas.[11]

The Input-Output Tables produced by MoSPI provide industry measures of input intensity: the share of an industry's non-labor input costs that come from a particular sector of the economy. For example, we construct a measure of agricultural input share, defined as the share of input costs derived from non-timber agricultural inputs. We use the input flow matrix from the 2006-7 update to the 2004-5 Input-Output Tables.

1.3.2 Construction of variables

Table 1.1 summarizes the construction of the main variables used in this paper. We construct panels (district and subdistrict) with two periods as defined by the three economic censuses. Period 0 corresponds to 1990 - 1998, while period 1 corresponds to 1998 - 2005. For

[11] We thank Henry Jewell and Hyoung Gun Wang, of the World Bank Urban Unit, for generously generating and sharing these data.

growth regressions, baseline employment is given by 1990 employment in period 0 and 1998 employment in period 2. Agricultural income is summed over the five years preceding the end of the period: 1993-1997 for period 0 and 2000-2004 for period 1.

Following Cole et al. (2012) and consistent with various studies on Indian agriculture (Fishman, 2011; Guiteras, 2009), the primary climactic variable of interest is rainfall during the summer monsoon, defined as total precipitation over the months of June, July, August and September. This period roughly corresponds to the kharif (summer) growing season, although rabi (winter) crops also depend heavily on summer rain (citation). We demean and normalize rainfall values to produce a measure of rainfall that is exogenous to local characteristics. Although agricultural productivity is usually increasing in rainfall, there is the possiblity for too much rain. Like other recent papers (Cole et al., 2012; Hidalgo et al., 2010), we account for this non-montonic relationship between rainfall and agricultural output by defining our rain variable as the absolute value of the difference (in standard deviations) between observed rain and optimal rain for agricultural income, which our calculations suggest is 1.25 standard deviations above mean. As with agricultural income, we sum annual rainfall over the five year period preceding the Economic Census.

Working over a 15 year time period poses challenges towards the creation of a panel, as administrative units can change both name and area. In order to form a panel of districts, it was necessary to account for district splits and other redistributions of land between districts. We use Kumar and Somanathan (2009), as well as data provided by the Ministry of Statistics and Programme Implementation, to construct a panel of consistent districts over time. This involved agglomerating some districts into larger "super districts" when territorial transfers made districts inconsistent over time. From a total of 593 districts at the time of the 2001 Population Census, we construct a panel of 414 consistent districts and "super districts". Of these, we are able to match 388 across all datasets. Dropping small states and union territories reduces this number to 353. Finally, we restrict our sample to those districts in which agriculture is a major industry, keeping only districts that have more than 50% of the population living in rural areas. Missing data and trimming outliers results

Table 1.1: *Construction of Primary Variables*

Variable	Source	Description
Employment	Economic Census	Sum of employment of all nonfarm economic establishments in the Economic Census for each round (1990, 1998, 2005), including private firms, state-owned firms and government establishments (e.g. public schools).
Firm Count	Economic Census	Count of number of nonfarm economic establishments in the Economic Census for each round (1990, 1998, 2005), as defined above.
Rain	Indian Meteorological Department & ML Infomap	Demeaned and normalized measure of summer monsoon rain, defined as rain during the months of June, July, August and September. We match all villages to latitude and longitude using ML Infomap GIS data. We interpolate rainfall from grid points to the full village dataset of the 2001 Population Census. Subdistrict and district rainfall values are then computed as mean village rainfall, weighted by land under cultivation as given in the 2001 Population Census.
Agricultural income	Ministry of Agriculture & CMIE Indian Harvest	Sum of price $\times$ production for major crops at the district level, as reported by the Indian Ministry of Agriculture and compiled by the authors and CMIE. Missing prices are assigned using average state price for that year where available and average national price otherwise.
Infrastructure	Population Census	The village and town tables of the Indian Population Census (1991, 2001) provide measures of infrastructure (road, electricity, etc.).
Irrigation share	Agricultural & Population Censuses	The proportion of irrigated to non-irrigated land comes primarily from the Agricultural Census of 1995. Where missing it is provided first by the Agricultural Census of 2000 and if still missing from the 1991 Population Census. This proportion is then multiplied by land under cultivation as reported by the Indian Ministry of Agriculture and CMIE to get annual district-level land by irrigation status.

in 526 district-period observations in the two period panel.

1.4 Empirical strategy

]

This section describes the empirical strategy used to estimate the linkages between agricultural output and nonfarm economic activity. As discussed above, weather provides exogenous variation in agricultural output that will be used to estimate the causal impact of agricultural income on nonfarm economic activity.

Let $i = 1, ..., N$ index districts and $\tau = 1, ..., T_0(1990), ..., T_1(1998), ..., T(2005)$ index years. Time periods are indexed using t, which consist of the years between observations of variables; for example, $t = 1, 2$ correspond to the Economic Census periods 1990-1998 and 1998-2005, respectively. The outcome variable of interest is represented by $Y_{i,\tau}$, which in our case is log employment growth. The endogenous variable, agricultural income, is represented by $X_{i,\tau}$. $\tilde{X}_{i,\tau}$ is the $k \times 1$ vector of other explanatory variables and $Z_{i,\tau}$ is the $p \times 1$ vector of excluded instruments. For simplicity, we sometimes use the $(p + k) \times 1$ vector $W_{i,\tau} = (Z'_{i,\tau}, \tilde{X}'_{i,\tau})'$.

The structural equations for our model are:

$$Y_{i,\tau} = \beta X_{i,\tau} + \delta' \tilde{X}_{i,\tau} + \epsilon_{i,\tau} \tag{1.2}$$

and

$$X_{i,\tau} = \gamma' Z_{i,\tau} + \lambda' \tilde{X}_{i,\tau} + \nu_{i,\tau}, \tag{1.3}$$

which produce the reduced form equation

$$Y_{i,\tau} = \beta(\gamma' Z_{i,\tau}) + (\delta + \beta\lambda)' \tilde{X}_{i,\tau} + u_{i,\tau} \tag{1.4}$$

where $u_{i,\tau} = \epsilon_{i,\tau} + \beta \nu_{i,\tau}$.

To go from the model above to the data that we will use, it is necessary to sum up from years (τ) to periods (t) that correspond to the Economic Census; for example, $Y_{i,t=1} = \sum_{\tau=T_0}^{T_1} Y_{i,\tau}$ and $Y_{i,t=2} = \sum_{\tau=T_1}^{T} Y_{i,\tau}$.

1.5 Results

1.5.1 Agricultural income

As the identification strategy of this paper relies on the impact of weather on agricultural income, it is first worthwhile to investigate this relationship. Table 1.2 estimates the impact of rainfall on total crop income at the district level. Column 1 gives the impact of rain on annual district agricultural income, measured in standard deviations from the mean. Column 2 shows that this relationship is not monotonic. Column 3 uses the preferred specification for the rest of the paper, following (Cole et al., 2012), which is rain measured in standard deviations from the district optimum, which we estimate as 1.25 SD above mean. We estimate that an additional standard deviation of rain towards the optimum results in approximately 637 million rupees, or approximately $14.5 million.

1.5.2 Reduced form

Table 1.3 uses the same functional form to estimate the reduced form relationship between rain and nonfarm employment growth in rural areas, now controlling for baseline employment as well. There is no significant relationship between rural employment growth and rainfall. Column 1 regresses log employment growth on the sum of rainfall in the five years preceding the economic census. Table 1.4 repeats the exercise, dividing employment into nontradables and tradables, with tradable industries defined as all those in Section D (Manufacturing) of the Indian National Industrial Classification. Again, we see no significant relationship between rainfall and employment growth. In contrast, Table 1.5 reveals that urban employment in strongly increasing in rainfall. There does not appear to be an immediate effect; instead, employment growth appears to be increasing as the time between the agricultural income shock and the period of observation occurs. Table 1.6 divides this

Table 1.2: *Effect of rainfall shocks on district agricultural income (annual, 1999-2004)*

	1	2	3
Rain	5607.739	5933.648	
	(1115.639)***	(1116.941)***	
Rain^2		-3261.392	
		(857.973)***	
Rain AV125			6365.839
			(1248.261)***
District FE	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes
N	1524	1524	1524
r2	0.99	0.99	0.99

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. Dependent variable is annual district agricultural income, as reported by the Planning Commission. Rain is defined as standard deviations from mean rainfall, with mean and standard deviations defined at the district level using the panel 1971-2005. Rain AV125 is defined as the absolute deviations from 1.25 standard deviations above mean, multiplied by -1 in order to make agricultural income increasing in the variable. All regressions have district and state-year fixed effects, with robust standard errors clustered at the district level.

growth into nontradable and tradable industries; here we find significantly higher growth in manufacturing industries.

1.5.3 Channels

The results presented in Section 1.5.2 make clear that employment in urban manufacturing is strongly increasing in rainfall; one standard deviation increase in rainfall in one of the years preceding the economic census produces approximately 3.4 log points of additional employment growth. As discussed in Section 1.2, tradable employment would only respond to local income in the presence of some frictions in local input or output markets; we therefore take this result to demonstrate the presence of such frictions. These could take a number of forms. In the presence of immobile labor, we would expect to see crowd-out of tradable production, as increased returns to labor in agriculture increase wages that tradable goods producers must pay. This is the opposite of what we find. Instead, this result is more consistent with distance-related costs in either capital or output.

Table 1.3: *Effect of rainfall on rural employment growth*

	1	2	3	4	5	6	7
Rain (sum)	0.004						
	(0.010)						
Rain_{t-1}		0.016					0.016
		(0.016)					(0.013)
Rain_{t-2}			0.002				0.000
			(0.030)				(0.028)
Rain_{t-3}				0.005			0.007
				(0.029)			(0.027)
Rain_{t-4}					-0.031		-0.037
					(0.019)		(0.021)*
Rain_{t-5}						0.036	0.043
						(0.034)	(0.034)
N	526	526	526	526	526	526	526
r2	0.38	0.38	0.38	0.38	0.39	0.39	0.39

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. Dependent variable is rural district nonfarm employment growth, in logs. All regressions with state-year fixed effects and robust standard errors, clustered at the state level. Rain is defined as -1 × the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall. Rain (sum) is summed over the five year period before the Economic Census. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. Rain_{t-n} is rainfall n years before the end of the Economic Census period. Control variables for all regressions are period land, urban and rural population, and share of employment in banking in 1990.

Table 1.4: *Effect of rainfall on rural manufacturing employment growth*

	NT	NT	T	T
Rain (sum)	0.002		0.007	
	(0.008)		(0.015)	
Rain_{t-1}		0.020		0.016
		(0.016)		(0.018)
Rain_{t-2}		0.001		0.004
		(0.029)		(0.043)
Rain_{t-3}		0.012		0.007
		(0.023)		(0.052)
Rain_{t-4}		-0.046		-0.026
		(0.026)*		(0.019)
Rain_{t-5}		0.039		0.041
		(0.037)		(0.039)
N	526	526	526	526
r2	0.44	0.45	0.31	0.31

$^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$. Dependent variable is rural district employment growth, in logs. Tradable employment (columns 3 and 4) is defined as employment in all establishments in the Economic Census with NIC codes falling under Section D (Manufacturing). Non-tradable employment (columns 1 and 2) is the sum of employment in all economic establishments not classified as manufacturing enterprises. All regressions with state-year fixed effects and robust standard errors, clustered at the state level. Rain is defined as -1 × the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall. Rain (sum) is summed over the five year period before the Economic Census. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. Rain_{t-n} is rainfall n years before the end of the Economic Census period. Control variables for all regressions are period, land, urban and rural population, and share of employment in banking in 1990.

The most plausible story not explored in the model presented in Section 1.2 is that growth is driven by industries that rely on large quantities of agricultural inputs. It is natural to imagine that brewers and millers will thrive during agricultural booms, simply because of cheaper or more abundant inputs. We examine this by constructing a measure of agricultural input intensity: the share of non-labor input costs derived from agricultural goods, as described in Table 1.7. We construct a dummy variable for any industry that spends more than 20% of its non-labor input costs on agricultural products. Table 1.8 shows that in no specification do industries that are intensive in agricultural inputs respond

Table 1.5: *Effect of rainfall on urban employment growth*

	1	2	3	4	5	6	7
Rain (sum)	0.017						
	(0.006)***						
Rain_{t-1}		-0.003					-0.003
		(0.018)					(0.018)
Rain_{t-2}			0.039				0.032
			(0.024)				(0.024)
Rain_{t-3}				0.046			0.040
				(0.014)***			(0.012)***
Rain_{t-4}					-0.007		-0.011
					(0.021)		(0.020)
Rain_{t-5}						0.034	0.038
						(0.011)***	(0.012)***
N	526	526	526	526	526	526	526
r2	0.40	0.40	0.40	0.40	0.40	0.40	0.41

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Dependent variable is urban district nonfarm employment growth, in logs. All regressions with state-year fixed effects and robust standard errors, clustered at the state level. Rain is defined as -1 × the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall. Rain (sum) is summed over the five year period before the Economic Census. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. Rain_{t-n} is rainfall n years before the end of the Economic Census period. Control variables for all regressions are period, land, urban and rural population, and share of employment in banking in 1990.

Table 1.6: *Effect of rainfall on urban manufacturing employment growth*

	NT	NT	T	T
Rain (sum)	0.010		0.034	
	(0.005)*		(0.013)**	
Rain_{t-1}		0.002		-0.015
		(0.017)		(0.033)
Rain_{t-2}		0.019		0.079
		(0.023)		(0.028)**
Rain_{t-3}		0.028		0.075
		(0.017)		(0.029)**
Rain_{t-4}		-0.011		-0.020
		(0.017)		(0.026)
Rain_{t-5}		0.022		0.070
		(0.016)		(0.024)***
N	526	526	526	526
r2	0.39	0.39	0.38	0.40

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. Dependent variable is urban district employment growth, in logs. Tradable employment (columns 3 and 4) is defined as employment in all establishments in the Economic Census with NIC codes falling under Section D (Manufacturing). Non-tradable employment (columns 1 and 2) is the sum of employment in all economic establishments not classified as manufacturing enterprises. All regressions with state-year fixed effects and robust standard errors, clustered at the state level. Rain is defined as -1 $\times$ the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall. Rain (sum) is summed over the five year period before the end of the Economic Census period. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. Rain_{t-n} is rainfall n years before the end of the Economic Census period. Control variables for all regressions are period, land, urban and rural population, and share of employment in banking in 1990.

significantly more to rainfall shocks than other industries.[12]

Another potential explanation is that industries themselves may depend directly on water, in which case the identifying assumptions discuss in Section 1.4 are violated. We consider this unlikely. Only one industry in the input-output tables spends more than 1% of its non-labor input costs on water: water-based transport. Even manufactures of non-alcoholic beverages spend just over 1/10 of 1% of its non-labor inputs costs on water.

If the local availability of investment capital is an increasing function of local savings,

[12]We do find less statistically significant coefficients on Rain than in the main reduced form specification, but this is unsurprising given that we are not able to match all industries in the Economic Census to the industries in the input-output tables.

Table 1.7: *Construction of Location and Industry Characteristics*

Variable	Source	Description
Location classifications		
Bank90	Economic Census (1990)	Share of workers in location employed in the banking sector.
Market access	World Bank	Transportation costs from district centroid to nearest city of greater than 500,000 population in 2001 Population Census, as calculated using the Indian road network in Lall et al. (2004).
Industry classifications		
Tradable & Non-tradable	National Industrial Classification (2004)	Following Moretti (2010), we define tradable industries as all those falling under the manufacturing header (Section D)
Capital Intensity	Annual Survey of Industries (1991-2)	Total productive capital stock / total annual labor costs
External finance dependency	Annual Survey of Industries (1991-2)	Open loans / productive capital stock. Our best approximation of the measure used in Rajan and Zingales (1998).
Input Spending Shares	Input-Output Table (MoSPI)	Using the 2006-7 update to the 2004-5 Input Flow Matrix, we assign each $industry_i$-$industry_j$ pair an input share based on the percentage of non-labor and non-within-industry expenditure that $industry_i$ spends on inputs from $industry_j$. We then aggregate these for each industry based on the input of interest. For example, agricultural share is the by-industry share of input costs that go to agricultural products.
Tradability (CFS)	Commodity Flow Survey	The US Census Commodity Flow Survey gives average distance traveled per shipment by industry, which are then matched to NIC codes. Higher values mean more tradadable goods.

Table 1.8: *Impact of rainfall on employment growth by agricultural input intensity*

	Rural	Rural	Urban	Urban
Rain (Sum)	0.005	0.018	0.018	0.020
	(0.012)	(0.016)	(0.007)**	(0.007)**
High	-0.189	-0.356	-0.474	-0.505
	(0.078)**	(0.194)*	(0.068)***	(0.096)***
Rain × High		-0.025		-0.005
		(0.022)		(0.010)
N	1062	1062	1058	1058
r2	0.92	0.92	0.96	0.96

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. Dependent variable is urban district nonfarm employment growth, in logs. All regressions with state-year fixed effects and robust standard errors, clustered at the state level. Rain is defined as -1 × the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall, summed over the five year period before the end of the Economic Census period. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. High is a binary variable representing industries with agricultural input share of weakly greater than 0.2, determined using input-output tables. Control variables for all regressions are period, land, urban and rural population, and share of employment in banking in 1990. The coefficient on Rain therefore captures the effect of rainfall on employment growth in industries with agricultural input shares less than 0.2.

then tradable firms meeting external demand would expand in response to a positive shock to agricultural income. Table 1.9 provides some suggestive evidence for the existence of such a localization of capital markets, examining the share of firms reporting using local finance, as opposed to formal finance. We show that rainfall shocks increase the share of firms reporting use of local finance.

What about demand? There is certainly strong reason to believe that increased demand due to higher incomes from agricultural windfalls leads to employment growth: most of the employment response of the nonfarm economy is in the nontradable sector. Table 1.10 explores this channel further by categorizing districts as above or below median in road infrastructure, an important determinant of transport costs and thus tradability. We find no evidence for an increase in urban manufacturing employment in those districts that have higher transportation costs; the point estimate on the difference is actually negative, the opposite of what we would expect under scenario where urban manufacturing firms meet local demand.

Table 1.9: *Effect of rainfall on share of firms reporting informal financing*

	Employment	Firm Count
Rain (sum)	0.005	0.005
	(0.002)**	(0.003)
Pop urban share	-0.085	-0.113
	(0.046)*	(0.056)*
Bank emp share	0.002	-0.003
	(0.012)	(0.010)
N	263	263
r2	0.43	0.60

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. Dependent variable is change in share of establishments in the Economic Census reporting informal finance, 1998-2005. Firms receiving informal finance are those not reporting bank or government finance as the primary source of finance. All regressions with robust standard errors and state-year fixed effects, clustered at the state level. Rain is defined as -1 $\times$ the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall, summed over the five year period before the end of the Economic Census period. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. Control variables for all regressions are period, land, urban and rural population, and share of employment in banking in 1990.

1.6 Conclusion and next steps

How do frictions in markets in developing countries affect the spatial distribution of economic activity? Does agricultural income continue to play an important role in determining the economic opportunities of urban firms, or do global capital and goods markets make local income irrelevant? We provide new evidence on these classic questions in development economics. We find that positive shocks to rural agricultural income induce employment growth in urban areas, with growth concentrated in the manufacturing sector. In testing between demand and capital channels that could explain such manufacturing growth, our evidence suggests that agricultural surplus increases growth in manufacturing firms by providing capital, suggesting both that such firms are capital constrained and that local informal capital markets can successfully intermediate between local rural surplus and urban investment opportunities.

Given these findings, it is necessary for economists to move beyond their focus on the relationship between agriculture and the rural non-farm economy. It should be of little

Table 1.10: *Effect of rainfall on manufacturing growth by market access*

	Low Trans Costs	High Trans Costs	Pooled
Rain (sum)	0.044	0.025	0.046
	(0.013)***	(0.023)	(0.012)***
Rain * High TC			-0.026
			(0.016)
High TC			-0.215
			(0.128)
N	245	277	522
r2	0.47	0.38	0.39

$^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$. Dependent variable is urban district employment growth in manufacturing establishments, in logs. All regressions with robust standard errors and state-year fixed effects, clustered at the state level. Rain is defined as -1 $\times$ the absolute value of standard deviations of rainfall minus 1.25, considered the optimal amount of rainfall, summed over the five year period before the end of the Economic Census period. Period 0 thus corresponds to the five year window 1993-1997 and period 1 to 2000-2004. High TC is a variable representing above median transportation costs from the district to the nearest city of 500,000 people (2001). Control variables for all regressions are period, land, urban and rural population, and share of employment in banking in 1990.

surprise to development economists that the performance of urban economies is linked to that of nearby rural areas. Migration, consumption and capital flows are but three examples of the linkages between urban and rural economies.

These findings have implications for many debates in development policy. First, policies that increase agricultural production may also contribute to industrialization, although we must be cautious when applying findings from shocks to long term changes in productivity. Second, other policies that increase rural income, such as the Mahatma Gandhi National Rural Employment Guarantee Scheme (NREGS), may also increase industrialization. It should be noted that the increase in manufacturing employment that we observe occurs entirely in urban areas. This suggests that although manufacturing firms in rural areas may be credit constrained, returns in urban areas appear sufficiently high to overcome the proximity that rural firms have to the source of the income shock. Differences between returns to rural and urban manufacturing are certainly a topic for further inquiry.

Finally, our findings suggest that informal networks are able to intermediate capital between agricultural earnings and investment opportunities in manufacturing. Further research is needed to understand the channels by which agricultural surplus can fund urban manufacturing growth and the reasons that capital is able move from rural to urban areas within a district but is sufficiently localized to produce local growth, rather than being invested nationally. Although informal markets appear to successfully facilitate manufacturing growth, it is necessary to understand the efficiency with which they are able to fund entrepreneurship in order to judge how well they can substitute for formal financial institutions.

Chapter 2

Natural Resource Wealth and Local Politics[1]

2.1 Introduction

At least as far back as Thomas Malthus, economists have been concerned with the role of natural resources in economic development. Sachs and Warner (1995) launched a modern literature into the issue, demonstrating that natural resource wealth was correlated across countries with underdevelopment. That natural resource dependence often coexists with poor governance is widely known; however, we are still far from understanding the channels by which mineral wealth could undermine institutional quality, and the conditions under which this occurs. As pointed out by Isham et al. (2005), while Chad and the Democratic Republic of the Congo may suffer from clear curses of natural resources, the United States and Norway provide equally strong examples of countries with long and consistent growth records whose institutions were not undermined by vast natural resource wealth.

Natural resources do not appear at random, and they tend to be fixed over time, making inference about their effects difficult. Places rich in natural resources may lack other natural advantages; for example, they may be mountainous or remote. Counterfactual stories for places rich in natural resources are inherently challenging to defend. Rather than attempt

[1]Co-authored with Paul Novosad.

the very difficult task of constructing a counterfactual political environment for a resource rich location, this paper compares natural resource rich locations with each other over time. We use price shocks to provide plausibly exogenous variation to the value of subsurface natural resources in a given location. The context is India, a nation which as a whole is not highly dependent on point-source natural resources[2], but has many regions which produce a wide range of natural resources, including coal, iron, gold and various other minerals.

This provides an ideal context for examining the impact of mineral wealth on political outcomes. First, by using subnational and subregional variation, we are able to compare outcomes in regions that share basic political institutions. Second, by interacting global prices changes with the presence of mineral deposits, we generate plausibly exogenous variation in the value of subsoil mineral resources. India contains considerable internal variation in mineral deposits but overall production is small relative to world totals (Indian Bureau of Mines, 2011), eliminating the possibility of reverse causality from local political outcomes to global mineral prices. Third, Indian law defines the national and state governments to be claimants of all mineral taxes and royalties. We are therefore able to rule out one of the primary mechanisms by which natural resources are thought to undermine political outcomes: significant increases in local government revenue.

The starting point for this paper is the observation that changes in natural resource wealth influence political outcomes through many, often highly tangled and overlapping, channels. We propose and test three mechanisms by which mineral wealth may influence political outcomes. First, the selection effect posits that greater mineral wealth will change the composition of the candidate pool in local elections. Second, the election effect suggests that conditional on the candidate pool, mineral wealth changes the likelihood of electoral victory of certain types of candidates. Third, in the moral hazard effect politicians change their behavior in office in response to changes in the value of subsoil resources. We find strongest evidence for the election effect: winning candidates are much more likely than other candidates to have criminal cases against them. Our incumbency results also suggest

[2] We exclude agriculture from this study.

that the probability of victory for certain types of candidates is increasing in mineral wealth. At this point we cannot identify the underlying mechanism for these results. As will be discussed later in this paper, potential candidate and party responses to mineral wealth prevent us from attributing such results to changes in voter behavior alone.

This paper provides some of the first microeconomic causal evidence of a direct effect of natural resource wealth on local political outcomes. When a region experiences a boom in the value of subsurface natural resources, politician criminality increases significantly. In particular, we find that winning candidates are significantly more likely to be implicated in criminal cases. Margins of victory increase and incumbents are more entrenched: the vote share of both local incumbents and the state ruling coalition is increasing in the global prices of local mineral deposits. Thanks to the exceptional nature of Indian laws regarding mineral royalties, we are able to provide what we believe to be the first evidence of political consequences of natural resource wealth in the absence of significant changes to local revenue.

Despite competitive elections, criminality is common among politicians in democracies around the world. This is particularly true in developing countries, with potentially large negative consequences for welfare (Chemin, 2012). Postulated reasons for such a prevalence of criminals range from the advantage of criminals in intimidating voters (Aidt et al., 2011) to a lack of political competition (Banerjee et al., 2012a). Our findings, while not conclusive, point away from a lack of choice: winners are more likely to be criminal than the rest of the field of candidates. In subsequent sections we discuss some of the reasons why criminals may have electoral advantages in the presence of high mineral wealth.[3]

The rest of the paper is organized as follows. Section 2.2 reviews the existing literature examining the linkages between natural resources and political outcomes. Section 2.3 gives background information on mining in India and Indian political organization. Section 2.4 describes the sources of data and the construction of variables. Section 2.5 presents the

[3]Such electoral advantages may include greater appeal to voters, but not necessarily. For example, criminal politicians may exert greater effort to win elections if they are then better able to capture rents once in office.

conceptual framework that guides our empirical investigation. Section 2.6 explains our empirical strategy. Section 2.7 presents and discusses the results. Section 2.8 concludes.

2.2 Literature

There is a large if inconclusive literature seeking to understand the effects of natural resource wealth on political outcomes.This literature may be divided into two broad categories: conflict and democratization.[4]

The last ten years have seen numerous papers utilizing natural experiments to provide variation in income generally and natural resource wealth specifically, primarily at the country level, to understand the determinants of conflict. Miguel et al. (2004) use weather shocks to find that economic growth in Africa strongly reduces the likelihood of civil conflict. Bruckner and Ciccone (2010) find similar results using fluctuations in the prices of export commodities in sub-Saharan Africa, perhaps the closest existing paper to ours methodologically. Other papers find the opposite effect. Angrist and Kugler (2008) use within country variation in coca suitability to study the effect of a large increase in the demand for coca, finding an increase in conflict as groups fought over the rents in the coca trade. Still other studies find that resource wealth, once other variables are properly controlled for, has no impact on conflict or the overthrow of governments (Bazzi and Blattman, 2011). Cotet and Tsui (2013) use new oil discoveries to provide exogenous variation in resource wealth, also finding no significant effect on conflict. Dube and Vargas (2012) attempt to rationalize these disparate findings, arguing that resource production may increase conflict over government revenues but also decrease conflict through increasing the opportunity cost of time. They find evidence for this by comparing the differential effects of oil and coffee price shocks in Colombia; positive shocks to oil, a capital-intensive commodity, result in increases in conflict, while positive shocks to coffee, a labor intensive commodity, have the opposite effect.

[4] These are neither perfectly distinct nor comprehensive. Democratization often occurs as the result of armed conflict, and institutional quality involves many components beyond democratic depth.

The literature on democratization and institutional quality has likewise focused on natural resources and political outcomes at the country level, with different studies finding positive, negative and insignificant associations. Brunnschweiler and Bulte (2008) find that after properly controlling for other factors, natural resource endowments are positively associated with institutional quality. Likewise, Bruckner et al. (2012) find that increases in the international price of oil are actually associated with positive changes in the democracy scores of and democratic transitions in oil exporters. Brollo et al. (2013) find that increases in government revenues both attract lower quality politicians and induce politicians in office to be more corrupt. Bhattacharyya and Hodler (2010) find that natural resources only increase corruption in places that have been nondemocratic for most of the years since 1956. Burke and Leigh (2010) find evidence for heterogeneous effects depending on the type of resource; negative agricultural shocks (droughts) cause increases in democratization, while mineral price shocks have little to no effect. They take their findings to be supportive of the theory of Acemoglu and Robinson (2001) that democratization occurs when pressure is brought on governments by non-elites; agricultural production shocks create mass pressure for democratization specifically, while price drops in point-source commodities would not, although they would put fiscal pressure on governments highly dependent on natural resource royalties for government revenue. Our results contribute to this literature on democratization, not on the extensive margin but on the consolidation of democracy and determinants of institutional quality. An increase in the criminality of those holding political power is likely to weaken rule of law, as recent evidence from India seems to suggest (Chemin, 2012).

Ross (2001) proposes three channels by which point-source resource abundance may harm the quality of democratic institutions: a "rentier effect" in which resource wealth allows governments to avoid accountability; a "repression effect" in which resource wealth funds costly political repression; and a "modernization effect" in which resource wealth causes growth without the social and cultural change required for effective democratic governance. This paper seeks to contribute to our understanding of this "rentier effect", which is the result of multiple channels beyond a simple increase to government revenue. Changes to natural resource wealth are changes to the rents available to politicians, which Fisman et al. (2013) show to be significant in India. The availability of rents influences both who enters politics and their behavior in office, what Brollo et al. (2013) refer to as the selection and moral hazard effects of changes in government resources.

Yet political rents are not simply a function of government revenue. Certainly point-source natural resource extraction is often an important source of government revenue. But it is also an industry whose characteristics facilitate political rent seeking (Isham et al., 2005). First, firms are generally large, making political rent extraction easier. Second, mineral extraction involves high up front fixed costs followed by marginal costs that are often well below market prices, creating large profits for politicians to tax. Third, mining is a government-input intensive industry: governments often have rights to the land on which mining takes place, leases, permits and licenses (e.g. environmental clearances) are allocated at the discretion of government agencies, and mineral extraction is intensive in government-provided infrastructural inputs such as roads, water and electricity. All of these characteristics make mineral extraction an attractive target for rent-seeking politicians, particularly in developing countries where there are weak institutional constraints on elected officials.

It is generally difficult to disentangle the political impact of mineral wealth's revenue effect from other rent seeking discussed above, as mining almost always makes important contributions to government revenue, particularly at the national level. Brollo et al. (2013) make some progress on this front by taking advantage of exogenous variation in government revenues unrelated to natural resource extraction. Our work is complementary to theirs, isolating instead the non-revenue impact of mineral wealth. Indian mineral policy directs all taxes and royalties on mineral extraction to the national and state governments, with no provision for redistribution of such revenue back to the source locations. This grants us the unique opportunity to examine the other means by which natural resources can influence political outcomes.

Recent work in economics and political science has attempted to explain the prevalence of criminal politicians, with a particular focus on India. Chemin (2012) deepens the puzzle by finding that representation by criminal politicians is associated with a significant decrease in the living standards of minorities, which he suggests is the result of an increase in crime and corruption. Aidt et al. (2011) find that political parties choose more criminal candidates when literacy levels are low and electoral outcomes are uncertain, which they interpret as evidence for the ability of criminal politicians to intimidate voters.[5] Banerjee et al. (2012a) conduct a voter information experiment in Uttar Pradesh, finding that improved information about candidates lowers support for criminal candidates and that support for criminal candidates does not covary with educational or ethnic status. They conclude that such results are not consistent with a theory of voter preferences for criminality, but rather imperfect information or a lack of alternatives. Our findings suggest a somewhat different story. We find significantly larger effects of mineral wealth on the criminality of winners than on the criminality of the candidate pool as a whole. This suggests that criminal politicians are more likely to win election in response to an increase in mineral wealth. While we may only speculate as to why this is, our findings are not consistent with the theory that voters have no choice over criminality. As we discuss in more detail below, there are various reasons why voters may prefer criminal politicians during periods of high mineral prices.

2.3 Mining and politics in India

2.3.1 The mineral industry in India

Although modern India is not normally associated with great mineral wealth, it does in fact contain a large and varied natural resource sector. In 2010, the mining sector employed 521,000 workers and produced 2.5% of Indian GDP (Indian Bureau of Mines, 2011). Since

[5] See Vaishnav (2012) for further discussion of why parties may prefer to field criminal politicians under certain circumstances.

natural resources are not found in all locations, the share of output of the natural resource sector in the areas of its occurrence are much larger. Over sixty different major minerals are mined in approximately 3000 mines (Tata Environmental Research Institute, 2002).[6]

Historically, Indian mines were predominantly state owned until significant privatization in the 1990s. In 2010, 2229 of 2999 mines were private, representing 36% of total production value (Indian Bureau of Mines, 2011). Significant minerals deposits are found in nearly every region and state of India. The major exception is the region known as the Deccan Traps, a large flood basalt (lava flow) that covers a vast region in western-central India.

The mining sector is jointly regulated by the federal and state governments. Notably, royalties and taxes paid by mining corporations are paid directly to state and federal governments. Importantly for this study, there is no requirement for fiscal proceeds from mining to be spent in communities affected by mines, nor any evidence that these communities have any entitlement to such funds.

2.3.2 Political context

India is a federal parliamentary democracy. Representatives for the lower house of parliament, the Lok Sabha, are elected in first past the post, single elector constituencies. Members of the state legislative assemblies (MLAs), the focus of this paper, are elected in the same fashion. State governments are generally ruled by a chief minister, the leader of the ruling party or coalition in the state legislative assembly. State government ministers are drawn from the ranks of MLAs in the ruling party or coalition. The responsibility of MLAs is primarily to vote in the state legislative assembly, although they do have other more executive powers in their local areas (e.g. personal development funds) as well as considerable informal authority. Importantly, politicians also have some control over bureaucratic assignments, enabling them to exert influence over policymaking and implementation formally assigned to the bureaucracy (Iyer and Mani, 2012).

[6]Major minerals such as iron ore are jointly regulated by the national and state governments, while minor minerals such as granite are regulated entirely by state governments.

Electoral boundaries are set by a Delimitation Commission. Although such a commission is supposed to redraw all boundaries every ten years following the completion of the decennial population census, boundaries were frozen between 1976 and 2007. Thus, in the period covered by our analysis, only one redistricting took place; elections taking place during or after 2008 follow the new constituency boundaries.

2.4 Data

2.4.1 Sources

Data on electoral outcomes comes from the Election Commission of India (ECI).[7] This dataset contains candidate-level information on every candidate competing in state legislative assembly elections, including name, gender, party and votes received. Constituency level data on total electors and turnout is also available. We constructed our own dataset of state assembly governing coalitions going back to 1976, which we use to measure ruling coalition performance. Where possible we used official election commission and other state government documents to construct coalitions, augmenting them with newspaper accounts where necessary.

Data on politician characteristics comes from sworn affidavits submitted by candidates to the ECI. The Indian Supreme Court declared in rulings in 2002 and 2003, that all candidates must submit criminal, financial and educational information in order to be eligible for office. Politicians may be penalized for inaccuracies with fines, imprisonment and disqualification. By 2004, such requirements were in place in all states. The Association for Democratic Reform, a Delhi-based non-governmental organization, has collected, digitized and made publicly available these data.[8] Financial information contained in these affidavits includes

[7]While these data are publicly available in pdf format on the ECI website, we thank Francesca Jensenius for generously sharing her cleaned data files. See Jensenius (2013) for a more detailed description of Indian electoral data.

[8]http://adrindia.org/about-adr/who-we-are

Map of all major mineral deposits in India, as listed by the Mineral Atlas of India (Geological Survey of India, 2001). This includes deposits not used in this paper, for reasons discussed in Section 2.4.4. Nearly all states have major mineral deposits. The major exceptions are in the Indo-Gangetic Plain (Punjab, Uttar Pradesh) and in the northeast of India (Mizoram, Tripura).

Figure 2.1: *Map of deposit locations*

own assets and liabilities, as well as those of spouses and family members.[9] Politicians are also required to report any open criminal cases in which they are a defendant. Election laws in India bar convicted criminals of contesting elections; thus, pending criminal cases are the best available measure of politician criminality.[10]

Data on the location, type and size of mineral deposits come from the Mineral Atlas of India (Geological Survey of India, 2001), whose appendix contains the following data for major mineral deposit in India: centroid latitude and longitude, mineral type, and size class. Figure 2.1 shows a map of mineral deposit locations. Commodity prices come from the United States Geological Survey (Kelly and Matos, 2013), to our knowledge the most complete set of historical commodity prices available. All prices are for average prices in the United States for a given year. Where available, we use the price for the ore as it is listed in the Indian deposit data. Where not, we match deposits to the price of the processed output of the mineral deposit (e.g. aluminum for bauxite). Finally, the geographic data necessary for matching electoral constituencies to mineral deposits comes from the ML Infomap Pollmap dataset, which contains digitized GIS data based on maps published by the Election Commission of India.

The unit of observation for this paper is the constituency-year. Whenever data is available at the candidate level (e.g. candidate wealth), we collapse individual candidate level data to constituency-year averages.

[9] Although private financial information is difficult to verify, other researchers have also found this data reliable enough to work with. For example, Fisman et al. (2013) use affidavit data to estimate the private returns to holding political office.

[10] One potential concern is that spurious criminal charges may be made against certain candidates by others in an attempt to reduce competition. While this is possible, candidates are only required to report cases that a judge has decided warrants judicial proceeding upon presentation of evidence by police and prosecutors.

2.4.2 Definition of variables

Dependent variables

For each outcome derived from candidate affidavits, we create two variables, one for the winner and one for the average of the candidate pool (winner included). This allows us to differentiate between effects driven by candidate entry and those driven by voter choice, as discussed in Section 2.5. In this way we create variables for candidate assets, criminality and education. Candidate assets are defined as the sum of personal and family assets. We construct three measures of the criminality of political candidates. First, the binary variable *Criminal* indicates whether any criminal cases are being brought against a candidate at the the time of the election. Second, *Criminal Count* is an integer containing the number of criminal cases being brought against a candidate. Finally, the Association for Democratic Reform provides data on serious criminal cases such as murder and rape. We thus create the binary variable *Serious Criminal* to indicate whether a candidate is charged with any serious criminal cases. Finally, for education we create the variable *Graduate* to indicate whether a candidate has graduated from secondary school.

We likewise aggregate data from the Election Commission of India to the constituency-year level. We construct two measures of political competition: number of candidates contesting the election (top coded at 30) and effective number of parties (ENOP, an inverse Herfindahl measure of inequality of votes widely used in political science since its introduction by Laakso and Taagepera (1979)). We then define a series of variables measuring ex post political competitiveness. *Turnout*, the share of eligible voters casting votes, is taken directly from ECI reports. *Margin* is defined as the vote share of the winner minus the vote share of the runner up. *Local Inc Margin* is defined as the vote share of the party incumbent in the constituency minus the largest vote share of the non-incumbent party candidate. *State Inc Margin* is the same measure but for the ruling coalition: the top placing incumbent state coalition vote getter minus the top placing candidate from a party not in the ruling coalition. Positive values of *Local Inc Margin* and *State Inc Margin* indicate that the incumbent (party in the constituency or state coalition) is the winner. We thus define

Local Inc Win and *State Inc Win* as binary variables indicating that *Local Inc Margin* and *State Inc Margin* are greater than zero, respectively.

Price shocks

As discussed by Van der Ploeg (2011), natural resource dependence is likely to be endogenous to the quality of political institutions, a challenge to researchers who wish to identify the causal impact of mineral wealth on political outcomes. We address this concern by developing a measure that combines the location of mineral deposits with movements in the global price of the commodity. Our approach is closest to that of Bruckner and Ciccone (2010), who use price movements of international commodities to estimate the impact of economic shocks on civil conflict in sub-Saharan Africa. They construct the commodity price index for a given country in a given year as the weighted average of the price indexes of the various export commodities, with weights assigned by export share. We interact mineral price indexes with the presence of mineral deposits. From a list of 45 minerals for which we have both deposit and price data, we discard economically unimportant minerals, defined as those for which the Indian Bureau of Mines does not publish production statistics or those whose average output per deposit is valued at less than 1 million INR in 2009, the most recent year for which we have reliable disaggregated data.[11] We thus end up with 1325 deposits of 27 distinct minerals spread across 25 states in India. It is worth noting that we use the presence of deposits instead of the presence of mines, as the existence of deposits cannot be the outcome of political processes, while mines require not only deposits but also political inputs such as infrastructure, clearances and capital. As discussed by Brunnschweiler and Bulte (2008), among others, common measures of resource abundance, such as share of GDP from primary commodities, are in fact correlated with institutional factors and are better described as measures of resource dependence. Our use of mineral deposits avoids this endogeneity.

[11] We also discard fuel mineral (coal, lignite, oil and natural gas) producing regions, which account for 70% of the mineral production in India and are highly spatially concentrated (Indian Bureau of Mines, 2011). Their inclusion makes less tenable our assertion that price movements are uncorrelated with location characteristics.

Map of all state legislative constituencies in India (1976-2007 delimitation). Constituencies are shaded according to mean mineral price shock, as defined in Section 2.4.2. Red constituencies received above median price shocks in example year 2005, while blue constituencies received below average price shocks. White constituencies are those excluded from our sample, either because they had no major mineral deposits within 25 km or because they were within 25 km of a coal deposit. Definition of sample is discussed in more detail in Section 2.4.4.

Figure 2.2: *Map of price shocks (2005)*

We then assign deposits to state legislative constituencies by calculating the distance between deposit and constituency boundary. We consider a deposit to be linked to a constituency if the distance between deposit centroid and constituency boundary is less than 25 km. There are multiple reasons to believe deposits that do not fall within constituency boundaries will still affect political outcomes. First, deposits may be large geographically, but our data gives only the coordinate of the deposit centroid; thus a nearby constituency may actually contain mines from a deposit outside of its boundaries. Second, mines outside of a constituency may still be political linked to a constituency; for example, if a mine requires that a road be built through a neighboring constituency, it is likely to seek the support of the local MLA. Figure 2.2 is a map of constituencies, shaded according to their price shocks in 2005.

We then generate a constituency-year price shock using international price data from the United States Geological Survey. As legislative assembly terms are generally five years long, we calculate the price index for commodity c in year t as the average price in years $t-5$ through $t-1$, scaling by the average price in years $t-15$ through $t-6$. Mathematically:

$$PriceShock_{c,t} = \frac{\sum_{\tau=t-5}^{t-1} price_{c,\tau}}{\sum_{\tau=t-16}^{t-6} price_{c,\tau}}$$

Figure 2.3 presents mineral-wise price shocks for sample year 2005. We use a five year numerator and a ten year denominator in order to create a reliable price index that is not overly driven by large but transitory swings in the price of the commodity. In order to assign a single price shock to a constituency that could have multiple deposits, we follow Bruckner and Ciccone (2010) in averaging the price indexes, weighting by the number of

Figure 2.3: *Mineral price shocks 2000-2005*

Bar graph of price shocks by mineral in 2005. All 25 minerals used in this paper are displayed.

deposits of each mineral type assigned to the constituency.

2.4.3 Control variables

We use a range of constituency-level variables to control for constituency characteristics that may condition the political environment. We calculate the number of deposits and number of large deposits within 25km of a constituency as described in the preceding section. We then use our geospatial merge of constituencies and census locations (villages and towns) to generate two classes of controls. The first set of controls seeks to control for basic constituency characteristics that may affect the degree of political competition and types of politicians that compete for office. These are log population, share of population that lives in rural areas (villages), share of the population that is employed outside of basic crop production and average firm size. The first two come from the 2001 Population Census. The latter two, which we use as proxies for the economic development in a constituency, come from the 2005 Economic Census.[12] The second set of controls are proxies for the size and competence of government in the constituency: share of villages with electricity, primary schools per capita and government share of nonfarm employment. The first two of these variables are drawn from the 2001 Population Census, while the third comes from the 2005 Economic Census.

2.4.4 Construction of sample

Before the redistricting of 2007, there were 4090 constituencies in the states of India.[13]. We eliminates states that lack major mineral deposits, leaving 3733 constituencies in 24 states.

[12]See Asher and Novosad (2013) for a detailed description of this dataset.

[13]Union territories are governed differently and generally do not have legislative assemblies

Much of the empirical work in this paper, however, utilizes variation in mineral prices to provide exogenous variation in the value of mineral wealth. This analysis is conducted with the 1324 remaining constituencies that are matched to mineral deposits. For the analysis restricted to large deposits, this number shrinks to 1053 unique constituencies.[14]

2.5 Conceptual framework

The theoretical approach of this paper begins with the career concerns model of Brollo et al. (2013), itself based on earlier work by Persson and Tabellini (2000). They advance our understanding of the political consequences of increases in government revenue by distinguishing between two effects, what they call selection and moral hazard. The selection effect comes from the attraction of low quality candidates to an increase in political rents, while the moral hazard effect is an increase in corruption due to an increase in the incumbent's probability of reelection. We argue that there are, in fact, three channels by which a natural resource boom may affect the characteristics and behavior of politicians. The selection effect is actually comprised of two, potentially competing, effects. The true selection effect comes from different potential candidates deciding to run. The election effect, in contrast, comes from the collective decision of the voters of whom to elect to office. Finally, once in office, there is a potential change of officeholder behavior in response to a natural resource boom.

Theoretically, the impact of a mining boom on both the selection and election effects is unclear. Brollo et al. (2013) assume that political rents are more valuable to lower quality politicians, thus predicting (and finding) that an increase in government revenue would lead to a decrease in politician quality. This is not, however, obvious. It is possible that rents are more valuable to criminal politicians, who are experts in illicit gain, but it is also possible that political rents are most valuable to well educated politicians who can successfully control a larger bureaucracy or set of projects.

On the election effect, the literature provides little guidance as to how voter preferences

[14]Large deposits are defined as type I and II deposits according to Geological Survey of India (2001), the size of which varies by mineral.

over candidate characteristics might change in response to changes in the value of mineral wealth. Robinson et al. (2006) explicitly model the effect of natural resource price changes on development via politician behavior, but by assuming voter policy preferences are independent of mineral prices, they do not consider how preferences over politicians may respond to price changes. Focusing on criminality, voters may prefer less criminal politicians to manage a resource boom because they are less likely to divert needed investments for their own gain; on the hand, voters may prefer a more criminal politician with a higher ability to get fast results out of India's notoriously slow and corrupt bureaucracy, thus increasing the local economic benefits of a price increase. This is complicated by yet another determinant of electoral outcomes: campaign spending. Sadly, it is beyond the scope of this paper to investigate the response of campaign spending to resource booms. Although the affidavit data does include campaign spending by candidate, we consider it to be of very low quality. Kapur and Vaishnav (2011) discuss how extremely low limits on campaign spending results in a very high level of "black" (unreported) money in Indian elections. Official, self-reported campaign spending is thus of little value to researchers. Finally, parties themselves may choose to field different candidates in response to a mining boom.

2.6 Empirical strategy

In this section we explain the two approaches that we use to analyze the effect of natural resource wealth on political outcomes.

2.6.1 Cross-section OLS

In line with much of the literature on natural resources, we begin with the OLS comparison of resource-rich regions to non-resource rich regions, controlling for observable characteristics. This cross-sectional comparison shows how regions with natural resource wealth compare with regions without natural resource wealth.

We use geological deposits as the explanatory variable rather than quantity of natural resources exported or mining sector jobs because the latter may be endogenously determined by other factors. For example, endemic corruption in a region could mitigate the rise of a secondary sector, causing natural resource exports to dominate a local economy even if they do not cause political corruption. Geological deposits are a superior measure as they are an unchanging characteristic of a region.[15]

The empirical specification is:

$$Y_i = \beta_0 + \beta_1 * Deposit_i + \beta_2 * X_i' + \gamma_{s,t} + \epsilon_i, \tag{2.1}$$

where $Deposit_i$ is a dummy variable indicating existence of a mineral deposit, or a deposit above a certain size threshold, X_i is a vector of controls and $\gamma_{s,t}$ represents state-year fixed effects. The sample is the entire set of Indian legislative constituencies for which we have data, as discussed in Section 2.4.4. In all specifications we present heteroskedasticity-robust standard errors, clustered at the state level.

2.6.2 Price shocks

The problem with the OLS specification is that places with large mineral deposits may differ in unobservable ways from places without deposits. If towns were founded with the objective of exploiting natural resource deposits, they may lack other natural advantages, such as accessibility to major urban centers or pre-existing public goods. Mineral deposits are often found in rural, underdeveloped areas where political variables such as competition and corruption may differ significantly from more developed parts of the country. Of equal concern, mineral deposits can only be discovered in places where firms have made investments in exploration; such willingness to invest is likely to be endogenous to location characteristics.

To address these problems, we limit the sample to locations with natural resource deposits, and examine how political outcomes respond to the effect of recent commodity

[15] There is some possibility that discovery of deposits is a function of political conditions. This problem does not persist when we move to the price shock specifications below.

price shocks on political outcomes.[16] We then run a panel regression of outcomes on trailing 5-year price shocks. We estimate the following equation at the constituency-year level:

$$Y_{i,t} = \beta_0 + \beta_1 * pshock_{i,t} + \zeta * X_{i,t-1} + \gamma_{s,t} + \epsilon_{i,t}, \tag{2.2}$$

where $pshock_{i,t}$ is the size-weighted price change of geological deposits in constituency i in year t, X is a vector of constituency controls, and $\gamma_{s,t}$ represents state-year fixed effects. The coefficient β_1 identifies the effect of a commodity price on the political outcome. As we are using interacted state-year fixed effects, our estimates are the result of variation in commodity price changes within a given state in a given election. In all regressions we present heteroskedasticity-robust standard errors, clustered at the state level.

2.7 Results

2.7.1 Summary statistics

Table 2.7.1 gives summary statistics for our sample. Of the 4090 total state legislative constituencies in India, there are on average approximately two deposits within 25km of a constituency and approximately one large deposit per constituency. In the sample for which we have affidavit data, nearly 22% of candidates have criminal cases pending against them; this number exceeds 31% for winning candidates. Winning candidates are more likely to have serious criminal cases against them (14% to 10%). Winning candidates also, on average, have assets worth more than twice as much as the candidate pool as a whole. Our electoral outcome sample is much larger, as it begins in 1980 rather than 2004 for the affidavit data. Political competition is mixed: the mean margin of victory is nearly 14%, but turnout is over 64% and nearly 9 candidates contest the mean election. In this sample, local incumbent parties retain control of their constituencies in 46.7% of elections, while the incumbent ruling coalition wins 48.7% of seats in the following election.

[16] At a global scale, India is not a large producer of any of these commodities, so we are not concerned about endogeneity of prices.

Table 2.1: *Summary statistics*

Variable	Mean	(Std. Dev.)	N
Deposit count	2.227	(5.624)	4090
Large deposit count	1.039	(2.743)	4090
Pr(candidate criminal)	0.218	(0.245)	6294
Pr(winner criminal)	0.311	(0.463)	6294
Pr(candidate serious criminal)	0.098	(0.171)	6294
Pr(winner serious criminal)	0.143	(0.35)	6294
Criminal cases count (mean)	0.537	(0.996)	6294
Criminal cases count (winner)	0.932	(2.45)	6294
Candidate assets (million INR)	10.781	(31.023)	6294
Winner assets (million INR)	22.897	(78.829)	6294
Number of candidates	8.851	(6.01)	30962
Margin of victory	0.138	(0.126)	27810
Turnout	0.642	(0.143)	24666
Vote share incumbent party - vote share top opponent	-0.016	(0.195)	14634
Local incumbent win	0.467	(0.499)	14634
Vote share previous coalition - vote share top opponent	0.003	(0.218)	13981
State incumbent win	0.487	(0.5)	13981

2.7.2 Cross-sectional analysis

We first seek to understand how areas rich in natural resources vary from those that do not have natural resources. Table 2.2 shows that the candidate pool in areas with mineral deposits are approximately 15 log points less wealthy than those in areas without mineral deposits. There are not, however, significant differences in candidate criminality or education between constituencies with and without mineral deposits. These findings hold for winning candidates as well (Table 2.3): they are 11 log points less wealthy than winning candidates in other constituencies, but are not significantly different in other features that we observe.

Electoral results demonstrate additional differences between areas having and lacking mineral deposits. Table 2.4 presents results for the number of candidates running, effective number of parties (ENOP, an inequality measure of vote shares across candidates), margin of victory and voter turnout. We find that margins of victory are somewhat larger in places with mineral deposits, but we also find that ENOP is weakly higher in constituencies with mineral deposits. Table 2.5 presents findings on the performance of local and state incumbents. We find no no significant differences in incumbent performance between constituencies with and without mineral deposits.

Table 2.2: *Cross sectional relationship between mean candidate characteristics and existence of mineral deposits*

	Criminal	Serious Criminal	Criminal Count	Assets	Graduate
Deposit	-0.004	0.007	-0.018	-0.152	-0.003
	(0.009)	(0.006)	(0.036)	(0.038)***	(0.014)
Log population	-0.001	-0.000	-0.008	-0.070	-0.002
	(0.005)	(0.003)	(0.023)	(0.029)**	(0.006)
Rural pop share	0.003	-0.002	-0.006	0.016	-0.001
	(0.010)	(0.006)	(0.049)	(0.044)	(0.011)
Employment share	0.003	-0.065	0.480	1.452	0.046
	(0.046)	(0.034)*	(0.540)	(0.489)***	(0.147)
Firm size	-0.005	0.001	0.014	0.059	-0.012
	(0.005)	(0.003)	(0.040)	(0.053)	(0.009)
Rural electrification	-0.003	0.001	0.007	0.404	0.099
	(0.023)	(0.021)	(0.076)	(0.110)***	(0.037)**
Primary schools per capita	-8.284	-2.750	-43.325	-81.313	-17.385
	(6.120)	(3.800)	(26.072)	(40.783)*	(6.398)**
Government employment share	-0.067	-0.021	-0.125	-0.578	0.164
	(0.034)*	(0.029)	(0.169)	(0.219)**	(0.073)**
N	4983	4983	4983	4980	4943
r2	0.21	0.13	0.12	0.34	0.14

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. The independent variable of interest, *Deposit*, is a binary variable indicating whether a constituency is within 25km of a mineral deposit. The columns, in order, estimate the effect of price shocks on the share of candidates with any criminal cases pending, the share of candidates with serious criminal cases pending, the mean number of criminal cases pending against candidates, the natural log of mean candidate assets and the share of politicians who are secondary school graduates. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

Table 2.3: *Cross sectional relationship between winning candidate characteristics and existence of mineral deposits*

	Criminal	Serious Criminal	Criminal Count	Assets	Graduate
Deposit	-0.002	0.008	-0.048	-0.121	-0.014
	(0.013)	(0.012)	(0.064)	(0.041)***	(0.018)
Log population	0.001	0.009	0.058	-0.056	-0.001
	(0.008)	(0.007)	(0.043)	(0.030)*	(0.012)
Rural pop share	0.022	0.024	0.057	-0.037	-0.026
	(0.021)	(0.016)	(0.112)	(0.064)	(0.017)
Employment share	0.077	-0.031	0.767	1.599	0.125
	(0.142)	(0.089)	(1.161)	(0.682)**	(0.222)
Firm size	-0.014	0.007	0.018	0.056	-0.046
	(0.014)	(0.008)	(0.093)	(0.048)	(0.015)***
Rural electrification	-0.000	-0.026	0.114	0.529	0.206
	(0.031)	(0.023)	(0.113)	(0.113)***	(0.062)***
Primary schools per capita	-17.132	-5.639	-62.005	-63.330	-16.785
	(11.948)	(5.899)	(40.730)	(42.901)	(10.850)
Government employment share	-0.050	-0.059	-0.220	-0.678	0.178
	(0.117)	(0.071)	(0.437)	(0.288)**	(0.082)**
N	4983	4983	4983	4911	4943
r2	0.11	0.07	0.07	0.27	0.05

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The independent variable of interest, *Deposit*, is a binary variable indicating whether a constituency is within 25km of a mineral deposit. The columns, in order, estimate the effect of price shocks on the probability that the winning candidate has any criminal cases pending, the probability that the winning candidate has serious criminal cases pending, the total number of criminal cases pending against the winning candidate, the natural log of winning candidate assets and the probability that the winning candidate is a secondary school graduate. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

Table 2.4: *Cross sectional relationship between electoral outcomes and existence of mineral deposits*

	Candidates	ENOP	Margin	Turnout
Deposit	-0.143	0.092	0.007	-0.010
	(0.240)	(0.052)*	(0.004)*	(0.005)*
Log population	-0.423	0.009	-0.005	0.012
	(0.088)***	(0.015)	(0.002)**	(0.004)**
Rural pop share	-0.549	0.090	-0.005	0.018
	(0.286)*	(0.026)***	(0.003)	(0.006)***
Employment share	7.385	-0.457	-0.036	-0.014
	(1.745)***	(0.206)**	(0.028)	(0.038)
Firm size	0.153	-0.002	0.004	-0.008
	(0.108)	(0.010)	(0.001)***	(0.003)**
Rural electrification	1.960	-0.282	0.012	0.022
	(0.836)**	(0.095)***	(0.012)	(0.015)
Primary schools per capita	-822.864	-17.657	5.657	10.214
	(237.305)***	(20.803)	(2.485)**	(6.269)
Government employment share	0.073	-0.235	0.046	-0.153
	(1.867)	(0.227)	(0.023)*	(0.044)***
N	20035	22098	22490	20034
r2	0.59	0.38	0.15	0.68

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The independent variable of interest, *Deposit*, is a binary variable indicating whether a constituency is within 25km of a mineral deposit. The columns, in order, estimate the effect of price shocks on the total number of candidates running, the effective number of parties, the margin of victory and voter turnout. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

2.7.3 Response to price shocks

In this section we present our findings of the effect in exogenous changes in the global price level of locally produced minerals, as described in Section 2.6.2. We present our findings with a minimum of interpretation; we discuss possible interpretations of these results in Section 2.7.4. In the discussion below, point estimates represent the effect of an increase of 1 in the price shock variable, which is equivalent to a constant, sustained 100% increase in the price level over the baseline price, or a 200% linear increase in the price level over the 5 years prior to the election.

Table 2.6 presents the effect of price shocks on mean candidate criminality across a range of specifications. We find that candidate criminality is weakly increasing in mineral prices, with results stable across a range of specifications. In Table 2.7 we present the results of the same specifications, with the dependent variable being winner criminality instead of mean candidate criminality. Columns 1 through 3 present the results using our full sample of deposits. We find that winners are approximately 10 percentage points more like to have criminal cases against them in

Table 2.5: *Cross sectional relationship between incumbent performance and existence of mineral deposits*

	Local Inc Margin	Local Inc Win	State Inc Margin	State Inc Win
Deposit	0.002	0.003	0.001	-0.023
	(0.007)	(0.015)	(0.005)	(0.013)
Log population	-0.005	-0.010	0.000	0.001
	(0.005)	(0.010)	(0.004)	(0.011)
Rural pop share	-0.012	-0.019	0.000	0.007
	(0.006)*	(0.014)	(0.005)	(0.011)
Employment share	-0.050	0.019	-0.100	-0.163
	(0.068)	(0.174)	(0.040)**	(0.094)*
Firm size	0.006	0.008	0.002	0.001
	(0.003)**	(0.006)	(0.003)	(0.005)
Rural electrification	0.009	0.055	-0.012	-0.048
	(0.027)	(0.057)	(0.019)	(0.031)
Primary schools per capita	11.647	22.390	2.991	1.526
	(7.760)	(16.904)	(7.867)	(21.316)
Government employment share	0.000	0.040	-0.070	-0.150
	(0.034)	(0.071)	(0.039)*	(0.084)*
N	11802	11802	11619	11619
r2	0.16	0.13	0.36	0.28

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The independent variable of interest, *Deposit*, is a binary variable indicating whether a constituency is within 25km of a mineral deposit. The columns, in order, estimate the effect of price shocks on the margin of victory for the constituency incumbent party, probability of reelection of the constituency incumbent party, the margin of victory of the state ruling coalition and the probability of victory of the state ruling coalition. Margins of victory are defined as the vote share of the incumbent candidate minus the vote share of the highest placing non incumbent candidate. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

constituencies that have experienced a price shock of 1. This result is stable to the introduction of demographic controls (column 2) and controls proxying for the size and effectiveness of government. Columns 4 through 6 present the same specifications but restricted to large deposits (dropping type 3 deposits, the smallest classification given by Geological Survey of India (2001)). We find similar results: winning politicians are approximately 8.5 percentage points more likely to be criminal in the presence of a positive price shock in the specification with full controls. All results are significant at the 10% confidence level, with the specifications with full controls yielding coefficients on the price shock variable significant at the 1% level.

For the remaining tables, we restrict ourselves to the specification with full controls and the price shocks based on all deposits. Table 2.8 presents the effects of price shocks on mean candidate characteristics, and Table 2.9 presents the same results for winning candidates only. For the most part we do not find that price shocks predict significantly more criminal, wealthy or educated politicians;

Table 2.6: *Effect of price shocks on mean candidate criminality*

	1	2	3	4	5	6
Price shock	0.026	0.026	0.028			
	(0.021)	(0.021)	(0.020)			
Price shock (large deposits)				0.023	0.025	0.025
				(0.019)	(0.019)	(0.019)
Deposit count	-0.000	-0.000	-0.000			
	(0.001)	(0.001)	(0.001)			
Large deposit count				-0.001	-0.000	-0.000
				(0.001)	(0.001)	(0.001)
Log population		-0.015	-0.017		-0.018	-0.017
		(0.009)*	(0.009)*		(0.010)*	(0.010)*
Rural pop share		0.003	0.004		0.014	0.014
		(0.018)	(0.017)		(0.020)	(0.019)
Employment share		0.087	0.044		0.121	0.056
		(0.106)	(0.118)		(0.144)	(0.166)
Firm size		-0.007	-0.007		-0.008	-0.006
		(0.008)	(0.008)		(0.008)	(0.010)
Rural electrification			-0.035			-0.026
			(0.052)			(0.064)
Primary schools per capita			-24.293			-25.889
			(12.400)*			(16.305)
Government employment share			-0.064			-0.099
			(0.073)			(0.084)
N	1812	1755	1755	1453	1408	1408
r2	0.24	0.24	0.24	0.25	0.24	0.25

$^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$. All regressions include constituency-year fixed effects. The sample for columns 1 through 3 is all constituencies within 25km of mineral deposits, while columns 4 through 6 exclude constituencies that are within 25km of only the smallest category of deposits (type 3). Columns 2 and 5 introduce demographic constituency-level controls. Columns 3 and 6 use the full set of demographic and political controls. Standard errors are heteroskedasticity-robust, clustered at the state level.

Table 2.7: *Effect of price shocks on winning candidate criminality*

	1	2	3	4	5	6
Price shock	0.094	0.099	0.101			
	(0.037)**	(0.036)**	(0.035)***			
Price shock (large deposits)				0.079	0.085	0.085
				(0.029)**	(0.028)***	(0.028)***
Deposit count	-0.001	-0.001	-0.000			
	(0.002)	(0.002)	(0.002)			
Large deposit count				-0.002	-0.001	-0.000
				(0.002)	(0.002)	(0.002)
Log population		-0.021	-0.023		-0.032	-0.030
		(0.019)	(0.019)		(0.021)	(0.021)
Rural pop share		-0.003	-0.002		0.030	0.029
		(0.033)	(0.032)		(0.038)	(0.038)
Employment share		0.195	0.109		0.250	0.159
		(0.272)	(0.276)		(0.349)	(0.382)
Firm size		-0.029	-0.029		-0.034	-0.033
		(0.019)	(0.023)		(0.023)	(0.028)
Rural electrification			-0.027			-0.002
			(0.098)			(0.134)
Primary schools per capita			-50.581			-52.274
			(21.805)**			(26.166)*
Government employment share			-0.119			-0.119
			(0.187)			(0.217)
N	1812	1755	1755	1453	1408	1408
r2	0.12	0.12	0.12	0.11	0.11	0.12

$^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$. All regressions include constituency-year fixed effects. The sample for columns 1 through 3 is all constituencies within 25km of mineral deposits, while columns 4 through 6 exclude constituencies that are within 25km of only the smallest category of deposits (type 3). Columns 2 and 5 introduce demographic constituency-level controls. Columns 3 and 6 use the full set of demographic and political controls. Standard errors are heteroskedasticity-robust, clustered at the state level.

however, we do find some evidence that candidates have more criminal cases against them when mineral prices are high. In contrast, winning candidates are significantly more criminal, across our different measures of criminality, when mineral prices are high. For a price shock of 1, candidates are approximately 10 percentage points more likely to have any criminal cases against them, 6.5 percentage points more likely to have serious criminal cases against them and have, on average, .34 additional criminal cases. We also find that assets and education levels are higher among such candidates, but these results are not statistically significant.

We now turn to electoral outcomes. Table 2.10 presents findings on the result of price shocks on number of candidates running for office, ENOP, margin of victory and turnout. We find that price shocks significantly increase margins of victory, but have no significant effect on other measures of candidate entry or voter turnout. In order to better understand this increased margin of victory, Table 2.11 presents our results on incumbent performance. We find that a price shock of 1 is associated with a 3.4 percentage point increase in the vote share of the local (constituency) incumbent party, resulting in an 8.4 percentage point increase in the likelihood of reelection. We also find that a price shock of 1 is associated with a 2.6 percentage point increase in the vote share of the top performing candidate from the state incumbent ruling coalition. The state incumbent is also approximately 4 percentage points more likely to win a constituency, although this result is not statistically significant.

Finally, we test for the effect of price shocks on the same politicians over time. We use a panel of 334 winning candidates who run for reelection in constituencies with mineral deposits. We find that incumbent criminality does not respond significantly to mineral price shocks. We do find evidence that assets respond negatively to positive price shocks, a result that will be discussed in the following section. Of course, given the small sample size, these results should be interpreted cautiously.

2.7.4 Discussion

Following the discussion in Section 2.5, we seek to shed light on three potential effects of mineral price shocks on political outcomes, each of which has multiple potential channels. The first is the selection effect, in which lower quality candidates compete when rents from resource extraction are high. The second is the election effect, in which criminal candidates are more successful in elections following a mining boom. The third is the moral hazard effect: elected officials change their behavior in response to high rents or other changes in incentives for those already in political office. The following sections examine each of these effects in turn.

Table 2.8: *Effect of price shocks on mean candidate characteristics*

	Criminal	Serious Criminal	Criminal Count	Assets	Graduate
Price shock	0.028	0.014	0.166	0.088	-0.026
	(0.020)	(0.013)	(0.080)*	(0.074)	(0.035)
Deposit count	-0.000	0.000	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)
Log population	-0.017	-0.002	-0.053	-0.059	0.014
	(0.009)*	(0.005)	(0.051)	(0.049)	(0.006)**
Rural pop share	0.004	-0.000	0.010	0.081	0.003
	(0.017)	(0.013)	(0.083)	(0.102)	(0.013)
Employment share	0.044	-0.010	0.705	0.658	0.174
	(0.118)	(0.106)	(0.966)	(0.821)	(0.215)
Firm size	-0.007	-0.008	-0.011	0.128	-0.028
	(0.008)	(0.006)	(0.053)	(0.058)**	(0.022)
Rural electrification	-0.035	0.021	0.038	0.596	0.105
	(0.052)	(0.038)	(0.149)	(0.211)***	(0.068)
Primary schools per capita	-24.293	-10.698	-65.404	-126.116	-10.206
	(12.400)*	(8.929)	(37.574)*	(96.512)	(20.557)
Government employment share	-0.064	0.012	-0.028	-0.340	0.198
	(0.073)	(0.054)	(0.225)	(0.347)	(0.088)**
N	1755	1755	1755	1754	1730
r2	0.24	0.12	0.14	0.35	0.12

$^{*} p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. The sample for all columns is all constituencies within 25km of mineral deposits. The columns, in order, estimate the effect of price shocks on the share of candidates with any criminal cases pending, the share of candidates with serious criminal cases pending, the mean number of criminal cases pending against candidates, the natural log of mean candidate assets and the share of politicians who are secondary school graduates. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

Table 2.9: *Effect of price shocks on winning candidate characteristics*

	Criminal	Serious Criminal	Criminal Count	Assets	Graduate
Price shock	0.101	0.065	0.335	0.155	0.023
	(0.035)***	(0.023)**	(0.142)**	(0.122)	(0.034)
Deposit count	-0.000	0.000	0.004	0.001	0.002
	(0.002)	(0.001)	(0.005)	(0.003)	(0.001)
Log population	-0.023	-0.009	-0.028	-0.017	0.012
	(0.019)	(0.013)	(0.067)	(0.063)	(0.011)
Rural pop share	-0.002	-0.001	0.002	0.029	-0.041
	(0.032)	(0.019)	(0.180)	(0.140)	(0.032)
Employment share	0.109	-0.008	1.074	0.669	0.568
	(0.276)	(0.197)	(2.020)	(1.228)	(0.403)
Firm size	-0.029	-0.019	-0.035	0.092	-0.093
	(0.023)	(0.012)	(0.115)	(0.063)	(0.028)***
Rural electrification	-0.027	0.043	-0.022	0.722	0.123
	(0.098)	(0.064)	(0.431)	(0.227)***	(0.097)
Primary schools per capita	-50.581	-27.309	-115.807	-157.899	-11.115
	(21.805)**	(13.830)*	(71.941)	(108.585)	(28.568)
Government employment share	-0.119	0.010	-0.267	-0.805	0.106
	(0.187)	(0.141)	(0.605)	(0.464)*	(0.164)
N	1755	1755	1755	1717	1730
r2	0.12	0.08	0.08	0.26	0.06

$^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$. The sample for all columns is all constituencies within 25km of mineral deposits. The columns, in order, estimate the effect of price shocks on the probability that the winning candidate has any criminal cases pending, the probability that the winning candidate has serious criminal cases pending, the total number of criminal cases pending against the winning candidate, the natural log of winning candidate assets and the probability that the winning candidate is a secondary school graduate. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

Table 2.10: *Effect of price shocks on electoral outcomes*

	Candidates	ENOP	Margin	Turnout
Price shock	-0.019	0.006	0.017	0.006
	(0.304)	(0.029)	(0.005)***	(0.006)
Deposit count	-0.003	0.006	-0.000	-0.000
	(0.009)	(0.002)**	(0.000)	(0.000)
Log population	-0.298	0.030	-0.002	0.004
	(0.177)	(0.028)	(0.004)	(0.005)
Rural pop share	-0.462	0.042	0.005	0.014
	(0.248)*	(0.044)	(0.005)	(0.005)**
Employment share	2.300	-0.468	-0.095	0.085
	(1.944)	(0.363)	(0.048)*	(0.058)
Firm size	0.835	0.023	0.007	-0.026
	(0.313)**	(0.031)	(0.004)*	(0.007)***
Rural electrification	2.424	-0.121	-0.014	0.059
	(0.886)**	(0.104)	(0.014)	(0.019)***
Primary schools per capita	-343.153	22.384	9.847	3.021
	(235.055)	(23.307)	(3.232)***	(7.943)
Government employment share	-5.490	-0.752	0.029	-0.184
	(1.869)***	(0.329)**	(0.038)	(0.058)***
N	7421	8315	8326	7421
r2	0.51	0.40	0.20	0.66

$^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$. The sample for all columns is all constituencies within 25km of mineral deposits, 1980-2012. The columns, in order, estimate the effect of price shocks on the total number of candidates running, the effective number of parties, the margin of victory and voter turnout. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

Selection effect

Do lower quality politicians compete for office when high mineral prices create higher rents? Brollo et al. (2013) find evidence that average candidate quality falls in response to an increase in government revenue. Our results are less clear: we find consistent but insignificant positive increases in mean candidate criminality associated with high mineral prices. We do find that candidates have, on average, more criminal cases against them. We also find no significant change in candidate likelihood of being a high school graduate. We also find no evidence that additional candidates compete for election when price shocks are high. Our results on a selection effect are, therefore, inconclusive.

Taking seriously for a moment the observed increase in mean criminal cases associated with positive price shocks, there are multiple ways to interpret the results. In the model of Brollo et al. (2013), lower quality candidates are attracted to an increase in available rents. In their paper they can only proxy for candidate quality with education; however, it is not obvious why lower education

Table 2.11: *Effect of price shocks on incumbent performance*

	Local Inc Margin	Local Inc Win	State Inc Margin	State Inc Win
Price shock	0.034	0.084	0.026	0.042
	(0.013)**	(0.041)*	(0.012)**	(0.026)
Deposit count	-0.001	-0.002	-0.000	-0.001
	(0.000)	(0.001)	(0.000)	(0.001)
Log population	-0.013	-0.030	-0.000	-0.008
	(0.010)	(0.019)	(0.006)	(0.011)
Rural pop share	-0.003	0.004	0.004	0.004
	(0.008)	(0.029)	(0.007)	(0.018)
Employment share	-0.152	-0.185	-0.010	-0.005
	(0.102)	(0.260)	(0.050)	(0.127)
Firm size	0.013	0.018	-0.002	-0.009
	(0.007)*	(0.016)	(0.008)	(0.014)
Rural electrification	0.018	0.072	0.006	-0.032
	(0.034)	(0.060)	(0.027)	(0.052)
Primary schools per capita	17.134	32.981	-1.980	-1.566
	(9.275)*	(22.589)	(6.633)	(16.299)
Government employment share	-0.011	0.096	0.021	0.040
	(0.071)	(0.152)	(0.049)	(0.100)
N	4312	4312	4635	4635
r2	0.21	0.15	0.38	0.26

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. The sample for all columns is all constituencies within 25km of mineral deposits, 1980-2012. The columns, in order, estimate the effect of price shocks on the margin of victory for the constituency incumbent party, probability of reelection of the constituency incumbent party, the margin of victory of the state ruling coalition and the probability of victory of the state ruling coalition. Margins of victory are defined as the vote share of the incumbent candidate minus the vote share of the highest placing non incumbent candidate. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

candidates are better able to extract rents than those with more education. In contrast, there are multiple channels by which more criminal politicians may benefit more from an increase in potential rents. First, with criminal cases already pending, they likely have less to lose from corruption than candidates not facing criminal conviction. Second, rent extraction is a criminal undertaking; past criminals are presumably better able to carry it out than those without criminal experience. We believe that this is particularly true for this particular context, where rent extraction comes not from government revenue, but from mining companies themselves. The drawback of our measure of criminality, however, is that it measures criminal cases and not criminal convictions. Criminal cases may be brought erroneously by politicians in power seeking to limit competition. Thus, what appears to be a selection effect may in fact be a change in behavior on the part of the ruling party in response to a price shock.

Election effect

We now consider whether, conditional on the candidate pool, voters elect lower quality candidates when mineral prices have increased. We do find strong evidence that in the presence of positive price shocks, winners are more likely to be criminal, have serious criminal cases against them and have more total cases against them. What explains such results?

One possibility is that voters actually prefer more criminal politicians during a mining boom. While at first glance, criminality may seem like a liability to a candidate or a party, there are multiple demand side reasons why voters may prefer criminal politicians under certain circumstances. In fact, criminal candidates are twice as likely to win election in India as their non-criminal counterparts (Vaishnav and Sircar, 2012). The simplest reason is that criminal politicians may have other skills that voters find desirable and that correlate with criminality. Criminality could signal that a candidate is willing to work outside the law when necessary. While lawlessness on the part of elected officials may hamper development in general (Chemin, 2012), a mining boom may be a time when voters actually prefer a politician to be less restricted by the law. First, if criminal politicians are better able to convince India's notoriously slow bureaucracy to make public investments such as roads, voters will prefer them at times when returns to such investments are high. Second, if mining companies engage in more extralegal activities during periods of high mineral prices, voters may prefer politicians who are willing and able to fight back extra-legally.

A second possibility is that criminal politicians may expend more effort to win election when rents from minerals are high. This effort could come in the form of campaign expenditure, patronage, voter intimidation or voter fraud. We intend to test this hypothesis in the future using election-specific crime data. A closely related hypothesis is that parties may favor criminal politicians when mineral prices are high; such politicians may pay more into state party coffers when potential rents are high.

Voters also appear to favor incumbent politicians in the presence of positive price shocks. We find that both local and state incumbents receive a greater vote share when mineral prices are high. This is consistent with the hypothesis of imperfect information in which voters attribute to incumbent politicians some of the economic gains that are beyond their control (Cole et al., 2012). Incumbent politicians may now be able to commit more crimes for a given probability of reelection. We do not find support for this hypothesis in our incumbent politician panel, but the small sample size prevents us from putting too much weight on these results.

Moral hazard

We now consider how incumbent politicians' behavior responds to changes in mineral wealth, the "moral hazard" effect. We have already found that margins of victory and incumbency advantages increase in response to an increase in mineral prices. One might expect incumbents to engage in worse behavior, given the increased probability of winning the next election. We do not find evidence in support of this channel. Table 2.12 presents the impact of price shocks on the criminality and assets of winning politicians who reran in the following election. We find no evidence for an association between price shocks and incumbent criminality. There are three potential explanations for this. The first is that our small sample size does not allow us to pick up such effects. The second is that incumbent politicians are able to prevent the government from bringing charges against them. The third is that there is no such effect in this context.

Our results do show, however, that incumbent assets are strongly decreasing in price shocks. At first glance, this result is surprising, as most theories of the political resource curse hinge on an increase in rents available to politicians. However, we propose three theories for explaining this finding, particularly in light of our other results. First, changes in mineral prices increase the long-term value of political office in two important ways: increased incumbency advantages mean politicians are more likely to be in power in the future, and persistence of price changes increase the expected value of rents in the future. Thus, politicians solving an intertemporal optimization are likely to choose to extract more tomorrow if rent seeking today lowers the probability of reelection. Second, given the common perception of corruption in the mining industry, mineral booms are likely to increase scrutiny of politician behavior, particularly incumbents. India, for all of its democratic shortcomings, enjoys a free and tenacious press, the more so after the passage of the Right to Information Act of 2005 (RTI). Third, if rents come in the form of illegal transfers from or preferential contracts with mining companies, politicians will seek to keep such transactions private, one part of which is not having the money flow into their official bank accounts. The rules that provide us with our data – the requirement that all politicians submit financial and criminal affidavits – provide a strong incentive to conceal such transactions, a corruption Hawthorne effect of sorts.

Table 2.12: *Effect of price shocks on incumbent assets and criminality in following election*

	Assets	Criminal	Serious Criminal	Criminal Count
Price shock	-0.199	-0.021	0.018	0.082
	(0.084)**	(0.090)	(0.056)	(0.357)
Baseline assets	0.569			
	(0.034)***			
Baseline criminal		0.266		
		(0.095)**		
Baseline serious criminal			0.316	
			(0.042)***	
Baseline criminal case count				0.418
				(0.104)***
Deposit count	0.002	0.001	-0.001	0.009
	(0.007)	(0.006)	(0.005)	(0.016)
Log population	0.036	0.007	-0.019	0.188
	(0.063)	(0.053)	(0.031)	(0.270)
Rural pop share	0.067	0.120	0.075	-0.167
	(0.115)	(0.057)*	(0.043)	(0.348)
Employment share	0.048	0.380	-0.361	-1.058
	(0.695)	(0.559)	(0.186)*	(1.377)
Firm size	-0.179	0.013	0.027	0.188
	(0.075)**	(0.045)	(0.022)	(0.167)
Rural electrification	0.219	0.291	0.026	1.295
	(0.230)	(0.104)**	(0.151)	(0.942)
Primary schools per capita	237.908	57.558	-19.168	-365.978
	(137.070)	(41.705)	(25.521)	(205.625)
Government employment share	0.890	0.539	-0.054	-0.122
	(0.719)	(0.346)	(0.149)	(1.260)
N	334	334	334	334
r2	0.70	0.26	0.17	0.25

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. Included in the sample for all columns are all incumbent politicians who run for reelection and we have linked in our data. Outcome variables are, respectively, log candidate assets, candidate criminality, candidate serious criminality, and the total number of criminal cases pending. All regressions include constituency-year fixed effects. Standard errors are heteroskedasticity-robust, clustered at the state level.

2.8 Conclusion

In this paper we present some of the first microeconomic evidence on the effect of point-source natural resource wealth on local political outcomes. We interact global price movements with the locations of mineral deposits in India in order to generate exogenous variation in the value of subsoil mineral wealth. India is an ideal country for such a study. Its free democratic elections and candidate disclosure requirements make available the necessary data. India contains a large number of mineral deposits spread around the country with a high degree of spatial variation. We construct a detailed constituency-year level dataset of politician characteristics and electoral outcomes to test for the impact of such shocks and provide insight into the mechanisms by which such effects occur.

We find that positive prices are associated with significant increases in the criminality of winning candidates for state legislative assembly seats. Margins of victory increase in the presence of high mineral prices. Incumbency advantages are significantly higher as well: the vote share of both local incumbents and the state ruling coalition is increasing in the global prices of local mineral deposits.

India is an open democracy with a free press and competitive elections. Nevertheless, we find strong evidence for a political resource curse. In fact, the evidence suggests that criminality increases more among winning candidates than the pool of candidates as whole. This may be due to greater effort on the part of criminal politicians or due to voter preferences for more criminal politicians during mining booms. The next step in this research is to examine the specifics of politician criminality. For example, what crimes are increasing in mineral prices? This will allow us to disentangle questions of politician effort to gain political office versus voter preferences. Regardless, mineral wealth appears to undermine the quality of democratic representation, particularly problematic in light of recent findings that more criminal politicians are associated with significant decreases in constituency welfare (Chemin, 2012).

Chapter 3

The Employment Effects of Rural Road Construction[1]

3.1 Introduction

Universal access to paved roads, much like clean water and consistent electricity, remains an unreached goal in many developing countries, particularly in rural areas. Fifty-four years after independence, 33% of Indian villages did not have a paved approach road in 2001 (Population Census). The absence of such infrastructure raises trade costs and reduces access to both outside markets and government services. The high costs of infrastructure investments mean that both economic and political considerations tend to guide their placement, posing challenges for the estimation of their impact. In this paper we exploit the allocation rules of a large-scale rural road construction program in India to estimate the impact of feeder roads on rural nonfarm economic activity.

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India. The government developed specific guidelines to prioritize large, unconnected habitations: those with populations above 1000 were to receive highest priority, followed by those with populations above 500. Lower priority was given to smaller localities and those with existing "fair-weather" roads. At the start, about 170,000 habitations were eligible for the program, a number that has grown

[1]Co-authored with Paul Novosad.

as the guidelines have been expanded to include smaller habitations. By March 2011, over 420,000 km of roads had been sanctioned to connect nearly 110,000 habitations at a cost of 1.19 trillion INR ($27 billion) (Ministry of Rural Development, 2012).

These rules provide us with three distinct ways to estimate the impact of a new road on rural economic activity. First, we provide OLS estimates of the relationship between PMGSY road construction and employment growth, based on the timing of road construction in villages that eventually received roads. However, endogenous timing of road construction creates unknown bias in these estimates. We address this in two ways. Program rules create discontinuities in the likelihood of receiving a road at populations of 500 and 1000, allowing us to use a fuzzy regression discontinuity approach to estimate the impact. Finally, we take advantage of the fact that planning and implementation was carried out at the district level. As a result, the probability of receiving a road early in the program was a function not only of village size, but also of its relative size within its district. We are thus able to instrument for road treatment with the within-district population rank, controlling flexibly for population.

We construct a new dataset that combines data on road construction with village characteristics and economic outcomes. We match Population Census data (1991, 2001) to Economic Census data (1998, 2005) to measure the economic consequences of road construction during the first five years of the PMGSY (2000-04). The Economic Census is a complete enumeration of nonfarm economic establishments in India, covering over 4000 towns and 500,000 villages. It contains, among other variables, data on employment and industrial sector for each establishment, allowing us to estimate the effect of rural roads on employment growth, firm size and formality.

We interpret the construction of a village feeder road under the PMGSY as a large reduction in the costs of moving goods, capital and people to and from a village. Theoretically, it is unclear what impact such a change in factor mobility should have on village economic activity. For example, out-migration may increase as the costs of travel decrease, or fall as economic opportunity expands in the village. Nonfarm employment may grow as firms serving outside demand now face sufficiently low transportation costs to be competitive in outside markets; alternatively, villages may further specialize in agriculture as trade for outside goods becomes more feasible. Increasing returns to scale, such as those due to fixed costs, would predict that certain firms may only exist when transport costs are low enough to enable them to access markets beyond the confines of a village. Thus road construction may not result in the specialization predicted by simple trade models but rather diversification of the village economy as it becomes integrated into a larger trade network.

We find that the construction of a road results in a significant increase in nonfarm village employment. While our results vary depending on the specification, all estimates predict over 100 percent growth in nonfarm employment upon receiving a road. As we assign treatment based on road completion, we interpret these not as temporary increases due to ongoing construction but medium-term changes in the level of village employment. In future work, we will test whether these results represent a change in the level or growth rate of village employment, a distinction recently explored in research on the economic impact of railroads in China (Banerjee et al., 2012b).

In addition to aggregate nonfarm employment, we find evidence that roads facilitate significant structural transformation and diversification of the local economy. Average village firm size increases, as does the number of industries present in the village. These findings are consistent with evidence from Nepal suggesting that villages closer to cities serve much larger markets, with greater specialization of household economic activities but greater diversification at the village level (Fafchamps and Shilpi, 2005). Further, we provide evidence that growth is significantly higher in villages that also have a supply of electricity, suggesting significant complementarities between transportation and power infrastructure.

We also provide some of the first well-identified estimates of the cost effectiveness of rural road construction. We find that, assuming no positive or negative employment spillovers to non-treated villages, $1,000,000 in rural road construction generates between 500 and 700 nonfarm jobs, or approximately $1400 to $2000 per job. Given that India's per capita GDP in 2005 was $732, our findings suggest very high returns to rural road construction, a conclusion supported by other attempts to estimate the returns to infrastructure investment (Fan and Hazell, 2001). These results should be interpreted with caution, as we do not observe concurrent changes to agricultural employment, and thus are unable to estimate changes in total village employment.

The rest of the paper proceeds as follows: Section 3.2 summarizes the most relevant literature on transport costs and rural roads. Section 3.3 provides a description of the PMGSY rural road construction program. Section 3.4 describes our empirical strategies. Section 3.5 explains the data used. Section 3.6 presents results and discussion. Section 3.7 concludes.

3.2 Literature

Recent years have seen a renewed interest in the importance of transportation costs in facilitating growth and development, particularly in the trade literature. Limao and Venables (2001) use quotes from a shipping company and other sources of data to estimate the impact of inter-country transportation costs on trade flows, concluding that much of the low trade volume in sub-Saharan Africa is due to high transportation costs that result from poor quality infrastructure. Djankov et al. (2010) estimate that every additional day required to ship a container between two countries is associated with a reduction in bilateral trade of more than 1%, in addition to causing a distortion in trade away from time-sensitive exports. Intra-national transport costs exacerbate the challenge of realizing gains from trade; Atkin and Donaldson (2012) estimate that internal trade costs in Ethiopia and Nigeria are 7-15 times larger than in the United States, greatly reducing the benefits of globalization.

Infrastructure has long been one of the priorities of economic development policy and research: fully 15% of World Bank spending between 1995 and 2005 was dedicated to infrastructure projects, with 42% of that amount spent in China and India alone (The World Bank, 2007). Recent research has utilized novel identification strategies to investigate the link between the expansion of infrastructure and local economic performance. Banerjee et al. (2012b) examine the impact of railroads in China, finding that while railroads caused a level increase in income, nearby locations grew no faster than farther locations during a nearly 20 year period of rapid economic growth. Storeygard (2012) interacts global oil price shocks with distances to the nearest port to investigate the impact of transport costs on urban economic activity in Africa, finding a significant inverse relationship between transport costs and urban economic output, as proxied by nighttime luminosity. Donaldson (2012) develops a multi-region, multi-commodity trade model to assess the impact of railroad construction in colonial India, estimating that the expansion of the railroad network into a region increased real income by approximately 16% and greatly reduced trade costs. Michaels (2008) finds that the construction of the US Interstate Highway System generates sectoral and wage growth consistent with standard trade theory.

Another strand of research estimates the impact of changes in market access rather than average transportation costs generally. Redding and Sturm (2008) find that cities close to the border between West and East Germany experienced population loss relative to cities further from the border whose market access was less impeded by the partition of Germany following World War II. Hornbeck and Donaldson (2012) seek to unite the market access literature with the infrastructure literature by estimating the impact of railroads on agricultural land values using a market access approach, finding that the railroad network had by 1890 more than tripled the total value of agricultural land in the United States.

Of course, there are many reasons why rural roads may have very different economic effects when compared to core infrastructure projects such as interstate highways and long-distance railroads. Casaburi et al. (2012) use a fuzzy regression discontinuity design to examine the effect of village feeder roads on agricultural markets, finding that such roads significantly lower market prices of local agricultural goods. Gollin and Rogerson (2010) build a multi-sector model and calibrate it using data from Uganda to understand the relationship between agricultural productivity, transport costs and economic activity. They predict that high transport costs produce inefficiently high specialization in agriculture, and that investments in road infrastructure would lead to significant reallocation of labor to the nonfarm economy. Khandker et al. (2009) use propensity score matching to evaluate the impact of a rural road program in Bangladesh, estimating that receiving a road lowers village poverty by 5-6% while increasing household consumption by 8-10%, although subsequent work suggests that some of these gains may not persist over time (Khandker and Koolwal, 2011). Most closely related to this paper, Banerjee et al. (2012c) study the impact of the PMGSY on a broad range of outcomes in a sample of 267 villages in Uttar Pradesh. They find that road construction results in greater access to government services, lower consumer prices, higher agricultural prices, increased employment outside of agriculture and less daily migration, with no effect on longer-term migration. While the majority of evidence points towards large economic gains from the construction of rural roads, some studies have suggested that low incomes and population densities in rural areas may not generate sufficient demand for transportation services on rural roads (Raballand et al., 2011).

Methodologically, our use of within-district population rank is similar to Andrabi et al. (2013), who use local population rank to instrument for the placement of a girls' secondary school. Banerjee et al. (2012c) also use within-district population rank to estimate the effects of the PMGSY although, as discussed above, their focus is on the response of households, rather than firms, to the construction of a new road.

3.3 Context and background

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India. The focus was on the provision of new feeder roads to localities that did not have access, although in practice many projects under the scheme upgraded pre-existing roads. Originally, the stated goal was to provide all habitations with populations greater than 1000 with connectivity by

2003 and all habitations with population greater than 500 with connectivity by 2007. These thresholds were to be lower in desert and tribal areas, as well as hilly states and districts affected by left-wing extremism.[2]

Although funded and overseen by the federal Ministry of Rural Development, responsibility for road construction was delegated to state governments. District Rural Road Plans were drafted for every district in India. Funding comes by a combination of taxes on diesel fuel (0.75 INR per liter), central government support and loans from the Asian Development Bank and World Bank. By March 2011, over 420,000 km of roads had been sanctioned to connect nearly 110,000 habitations at a cost of 1.19 trillion INR ($27 billion) (Ministry of Rural Development, 2012).[3] The mandate of the program has recently been expanded to include all habitations with populations above 100.

3.4 Empirical Strategy

Our goal is to estimate the effect of the construction of a new rural road on changes in village-level economic activity. We start by estimating the OLS relationship between road construction and village employment growth using the following estimating equation:

$$Y_{v,s,t} = \beta_0 + \beta_1 * newroad_{v,s} + \zeta X_{v,s,t-1} + \eta_s + \epsilon_{v,t-1}, \tag{3.1}$$

where $Y_{v,s,t}$ is log employment in village v in state s at time t, and *newroad* indicates that a new feeder road was constructed in village v at time t, $X_{v,s,t-1}$ is a vector of village controls measured at baseline, and η_s is a state fixed effect.

The problem with this approach is that roads are not allocated at random. Roads are expensive infrastructural investments likely to be demanded by nearly all unconnected villages. Governments may target such investments to locations that are particularly needy, or have high potential for growth, or are politically connected or favored by powerful politicians. Controlling for baseline village characteristics or regional fixed effects will reduce this bias, but it remains a concern at the heart of the literature on infrastructure.

We undertake three strategies to mitigate bias from selection in the placement of roads: (i) using

[2]Habitations are defined as clusters of population whose location does not change over time. They are distinct from, but form parts of, revenue villages used by the Economic and Population Censuses. See National Rural Roads Development Agency (2005) for more details.

[3]We use an exchange rate of 44.06 INR per USD, the average for 2005

Table 3.1: *Tabulation of Villages Receiving PMGSY Roads by Year*

	Year Completed										
Year Sanctioned	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**Total**
2000	6	166	609	1080	990	604	315	419	111	61	4361
2001	0	0	21	720	1387	1372	704	316	161	126	4807
2003	0	0	0	2	403	1251	1153	682	364	266	4121
2004	0	0	0	0	0	300	774	791	861	474	3200
2005	0	0	0	0	0	0	178	1108	1541	1839	4666
2006	0	0	0	0	0	0	1	569	1047	2263	3880
2007	0	0	0	0	0	0	0	0	134	1150	1284
2008	0	0	0	0	0	0	0	0	0	149	149
Total	6	166	630	1802	2780	3527	3125	3885	4219	6328	26468

OLS to exploit variation in the timing of PMGSY road construction; (ii) using regression discontinuity to exploit population thresholds that determine road priority; (iii) using an instrumental variable approach to exploit within-district population ranking of villages, which is an additional determinant of road prioritization.

3.4.1 Road timing (OLS)

The naive OLS is biased because the types of villages that receive new rural roads differ on many unobserved characteristics from the types of villages that do not receive new roads. By limiting our sample to villages that eventually did receive PMGSY roads, and exploiting the timing of road construction, we can eliminate any confounders that differ between villages that did and did not receive new roads.

Table 3.1 shows the number of rural roads built under the PMGSY in each year. Of the 75,399 roads in our sample that were built by 2009, 25,354 were built by 2005, the year in which we measure employment outcomes.[4] We then estimate Equation 3.1 on this limited sample, where the treatment variable indicates that a road was built before 2005.

Under the assumption that the order in which roads were constructed under the PMGSY is uncorrelated with other factors that affect growth, this would be an unbiased estimate of the effect of rural roads on village outcomes. However, this assumption is tenuous: order of construction is likely to be influenced by both political and economic factors that could bias OLS estimates either upwards or downwards.

[4]The 2005 Economic Census was conducted from late 2005 to early 2006.

3.4.2 Population priority thresholds (RD)

Timing of road construction is endogenous, which means the OLS estimates of the effects of roads are likely to be biased. In order to overcome this endogeneity, we exploit the population thresholds intended to guide the allocation of roads under the PMGSY. State implementing officials were instructed to target habitations in the following order: (i) habitations with population greater than 1000; (ii) habitations with populations greater than 500; and lastly, (iii) habitations with populations greater than 250.

Even if selection into PMGSY treatment is biased by political or economic factors, these factors are not likely to change discontinuously at these population thresholds. If these rules were followed to any degree by state officials, the likelihood of PMGSY treatment will discontinuously increase at these population thresholds, making it possible to estimate the effect of the program using a fuzzy regression discontinuity design.

Under the assumption of continuity at the treatment threshold, the fuzzy RD estimator (Imbens and Lemieux, 2008) estimates the local average treatment effect (LATE) of receiving a new road, for a village with population equal to the threshold:

$$\tau = \frac{\lim_{pop \to T^+} \mathbb{E}[Y_v | pop_v = T] - \lim_{pop \to T^-} \mathbb{E}[Y_v | pop_v = T]}{\lim_{pop \to T^+} \mathbb{E}[newroad_v | pop_v = T] - \lim_{pop \to T^-} \mathbb{E}[newroad_v | pop_v = T]}, \tag{3.2}$$

where pop_i is habitation population, T is the threshold population, and $newroad_i$ is an indicator variable for whether village v received a new road in the sample period.

A given population threshold increases the probability of receiving a road by a different amount in different states. For example, states with a large number of large, unconnected villages, are more likely to have large first stages at the high threshold of 1,000. Analysis of PMGSY documentation and discussions with public officials have led us to focus on the population threshold of 500, in five major states which cover a large share of the geographic and economic diversity of India: Gujarat, Chhattisgarh, Assam, Maharashtra and Tamil Nadu.

We estimate the reduced form fuzzy RD using equation 3.3:

$$Y_{v,s,t} = \beta_0 + \beta_1(pop_{v,s,t-1} > T) + \beta_2 pop_{v,s,t-1} + \beta_3 pop_{v,s,t-1} * (pop_{v,s,t-1} > T) + \zeta X_{v,s,t-1} + \eta_s + \epsilon_{v,s,t}, \tag{3.3}$$

where $Y_{v,s,t}$ is log village employment at time t, T is the population threshold of 500, $pop_{v,s,t}$ is habitation population at time t, $X_{v,s,t-1}$ is a vector of village controls measured at baseline, and η_s is a state fixed effect. Village controls and state fixed effects are not necessary for identification but improve the efficiency of the estimation. The local average treatment effect of a road, identified in a village at the population threshold T, is $\beta_1 + \beta_3 * T$. For ease of exposition, we subtract the threshold value 500 from the population variable, such that $T = 0$, and β_1 fully describes the treatment effect.

The fuzzy regression discontinuity approach accurately identifies the treatment effect of rural roads, under the assumption that crossing the population threshold affects the probability of receiving a road, and nothing else of significance. There are two potential threats to this identification strategy. First, if other village characteristics vary discontinuously at the threshold in a way that we are unable to control for (e.g. if participation in other government programs uses the same thresholds), then our estimates will be biased. Second, if the running variable (habitation population) can be manipulated, randomness of assignment at the threshold is violated. We discuss this possibility when we describe results from the regression discontinuity strategy below.

3.4.3 Population rank (IV)

In addition to the population threshold rules, district-level planning and implementation of the PMGSY meant that priorization was determined not only by population but also by relative population ranking within a district: a village would receive higher prioritization than an equivalent village if it had fewer larger eligible villages in its district. Holding population constant, a village in a district with many larger unconnected villages is less likely to receive a new road under the PMGSY. Under the assumption that, after controlling flexibly for total population, the population rank of a village within a district does not affect a village's growth prospects except through the likelihood of receiving a road through PMGSY, instrumental variable estimation provides an unbiased estimate of the effect of a rural road (Angrist and Lavy, 1999).

Our empirical specification is:

$$Y_{v,s} = \beta_0 + \beta_1 * newroad_{v,s} + \beta_2 f(pop_{v,s}) + \zeta X_{v,s} + \eta_s + \epsilon_{v,s} \tag{3.4}$$

where $Y_{v,s}$ is the outcome of interest in village v in state s, $newroad_{v,s}$ is an indicator for whether the village received a road under the PMGSY, $f(pop_{v,s})$ is a function of village population, $X_{v,s}$ is a vector of village controls and η_s is a state fixed effect. We estimate Equation 3.4 using $RANK_{v,s}$, the within-district population rank of village v, as an instrument for $newroad_{v,s}$. In alternate specifications, in order to reduce noise, we instrument for $newroad_{v,s}$ with a dummy variable indicating that

$RANK_{v,s} < 75$.[5]

The estimation provides an unbiased estimate of the effect of a new rural road on employment growth, so long as the exclusion restriction is not violated: $RANK_{v,d,t-1}$ must affect growth only through the increased likelihood of obtaining a new road under the PMGSY. In Section 3.6 we discuss robustness checks to ensure satisfaction of the exclusion restriction.

3.5 Data

3.5.1 PMGSY

Data on the PMGSY is generated through the Online Management and Monitoring System (OMMS), the software used in program tracking and implementation. These data are not a survey - they are the administrative records of the actual program. Data include but are not limited to road sanctioning and completion dates, cost and time overruns, contractor names, and quality monitoring reports.

PMGSY data are reported at either the habitation or the road level. There is a many-to-many correspondence between habitations and roads: roads serve multiple habitations, and habitations may be connected to multiple roads. Habitations are subsets of census villages, which tend to comprise between one and three habitations; approximately 200,000 villages consist of only a single habitation.

3.5.2 Economic and population census

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 4th and 5th Economic Censuses respectively in 1998 and 2005.[6] The Economic Census is a complete enumeration of all economic establishments except those engaged in crop production and plantation; there is no minimum firm size, and both formal and informal establishments are included.

The Economic Census records information on the town or village of each establishment, whether ownership is public or private, the number and demographic characteristics of employees, the sources of electricity and finance, and the caste group of the owner. The main product of the firm is also coded using the 4-digit National Industrial Classification (NIC), which corresponds roughly to a 4-digit ISIC code. More detailed information on income or capital is not included. The main

[5] Our results are robust to different cutoffs for this dummy variable and are available upon request.

[6] The 6th Economic Census is ongoing at the beginning of 2013.

strengths of the data are its comprehensiveness, and rich detail on spatial location and industrial classification of firms.

We obtained location directories for the Economic Censuses, and then used a series of fuzzy matching algorithms to match villages and towns by name to the population censuses of 1991 and 2001.[7] We were able to match approximately 93% of villages between 1998 and 2005. We also use data from the Population Census of India in 1991 and 2001, which includes village population and other demographic data, as well as information on local public infrastructure (roads, electricity, schools and hospitals).

We matched PMGSY data to economic and population census data at the village level, using population census codes where they were reported in the PMGSY, and a Hindi-language fuzzy matching algorithm to match village names across the two datasets. We successfully matched over 85% of habitations listed in the PMGSY to their corresponding population census villages.

Table 3.2 shows village-level summary statistics for the entire sample of villages used in our analysis.

Table 3.2: *Summary statistics*

Variable	Mean	(Std. Dev.)	N
New road	0.049	(0.216)	181232
Employment (1998)	68.079	(100.469)	181232
Employment (2005)	84.257	(119.934)	181232
Ln employment growth	0.213	(0.876)	181232
Firm count (1998)	33.719	(45.697)	181232
Firm count (2005)	45.368	(60.037)	181232
Ln firm count growth	0.289	(0.842)	181232
2001 Population	1422.329	(1021.174)	181232
Pop growth 1991-2001	1.202	(0.263)	181232
Irrigation share	0.431	(0.365)	174649
Ln land area	5.363	(1.085)	174649
Distance from town	20.832	(19.453)	180783
Diversity (1998)	8.727	(6.503)	181232

[7]The Economic Census of 1998 was conducted with the house listing for the 1991 population census, while the 2005 Economic Census used codes from the 2001 population census.

Table 3.3: *OLS: Employment growth on roads*

	(1)	(2)	(3)	(4)
New road before 2005	0.113	0.079	0.058	0.036
	(0.019)***	(0.016)***	(0.015)***	(0.017)**
Baseline log employment	-0.275	-0.328	-0.477	-0.496
	(0.009)***	(0.008)***	(0.013)***	(0.014)***
Population	0.000	0.000	0.000	0.000
	(0.000)***	(0.000)**	(0.000)**	(0.000)***
Share of land irrigated			0.099	0.078
			(0.024)***	(0.026)***
Log(land area)			0.141	0.126
			(0.008)***	(0.008)***
Distance from town			-0.002	-0.002
			(0.000)***	(0.000)***
Baseline number of industries			0.024	0.026
			(0.002)***	(0.002)***
Constant	1.115	1.734	1.305	1.377
	(0.030)***	(0.063)***	(0.078)***	(0.078)***
N	48216	48216	46720	34888
r2	0.13	0.17	0.21	0.22

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents OLS estimates of the relationship between log employment growth (1998-2005) and treatment, as defined as having received a completed PMGSY road by 2005. The sample is all locations that received a PMGSY road before 2012. Column 1 presents the estimate only controlling for 1998 (log) employment and village population. Column 2 introduces state fixed effects. Column 3 introduces standard village level controls of share of land irrigated, log land area, distance from nearest town and number of non-farm industries present in 1998. Column 4 limits to villages in which the largest habitation had fewer than 1500 people. Standard errors are clustered at the district level.

3.6 Results

3.6.1 OLS

Table 3.3 presents OLS estimates of the relationship between log employment growth (1998-2005) and treatment, defined as having received a completed PMGSY road by 2005. The sample is all locations that received a PMGSY road before 2012. Column 1 presents the estimate only controlling for 1998 (log) employment and village population. Column 2 introduces state fixed effects. Column 3 introduces standard village level controls of share of land irrigated, log land area, distance from nearest town and number of non-farm industries present in 1998. Column 4 limits to villages in which the largest habitation had fewer than 1500 people. Standard errors are clustered at the district level.

The table shows that villages that received a new PMGSY road by 2005 on average had non-farm employment growth that was 3-12 log points higher than villages that did not receive a new road. The point estimates fall as more controls are included, suggesting that selection of villages for roads is non-random. In particular, the falling coefficient as state fixed effects are included suggests that higher growth states were more likely to implement the program early - this is consistent with reports that state administrative capacity played an important role in early implementation of PMGSY.

3.6.2 Regression discontinuity

Table 3.4 presents regression discontinuity estimates of the effect of PMGSY prioritization on a village's likelihood of receiving a road. Habitations with a population greater than 500 should have received higher priority than habitations with population under 500. The dependent variable in these regressions is an indicator variable that is set to 1 if a village received a road by 2005. The running variable is the population of the largest habitation in the village. The treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to 500. The value 500 has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect.

Columns 1-3 show estimates from the local linear specification (Equation 3.3). Column 1 is the baseline specification without controls. Column 2 adds state fixed effects, and column 3 adds village level controls from the 1991 population census, which are log land area, share of land that is irrigated, distance to the nearest town, and the number of different industries in the village in 1998. Column 4 adds a 4th degree polynomial in habitation population. The estimations show a robust and highly significant effect of the population threshold on the probability of a village receiving a new road. A village is on average 3% more likely to receive a new road if the population of the largest habitation is just above 500 than if the population is just below 500. Figure 3.1 depicts graphically the increase in probability of receiving a new road when the largest habitation is just above the threshold.

The bandwidth used in all specifications is 250, so the sample for the estimation are villages with a largest habitation in the range of 250-750, but results are robust to alternate bandwidth choices. All standard errors are clustered at the district level to account for spatial correlation. Controls and fixed effects are not necessary for identification, but their inclusion increases the efficiency of the estimator.

Table 3.5 presents instrumental variable regression discontinuity estimates of the effect of receiving a new road under the PMGSY on log non-farm employment growth from 1998-2005. As above, the running variable is the population of the largest habitation in the village, and a dummy variable indicating population greater than 500 instruments for a village's receiving a PMGSY road

Table 3.4: *First stage: RD estimates of effect of population threshold on probability of new road*

	(1)	(2)	(3)	(4)
1(Pop $\geq$ 500)	0.034	0.031	0.030	0.034
	(0.010)***	(0.010)***	(0.010)***	(0.016)**
Population	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Population * 1(Pop $\geq$ 500)	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.001)
Share of land irrigated			-0.004	-0.004
			(0.007)	(0.007)
Log(land area)			0.001	0.001
			(0.004)	(0.004)
Distance from town			-0.000	-0.000
			(0.000)	(0.000)
Baseline number of industries			0.001	0.001
			(0.000)*	(0.000)*
Constant	0.016	0.012	0.010	0.007
	(0.006)***	(0.006)**	(0.018)	(0.018)
N	28747	28747	26440	26440
r2	0.01	0.02	0.02	0.02

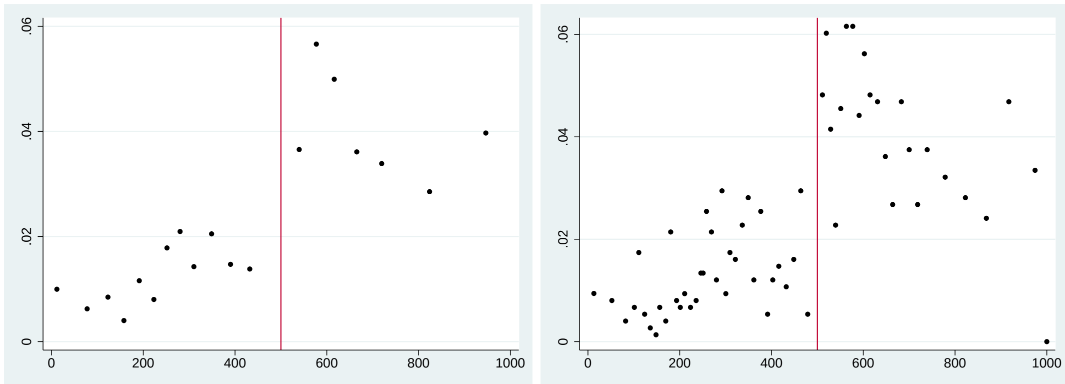
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table presents regression discontinuity estimates from Equation 3.3 of the effect of PMGSY prioritization on a village's likelihood of receiving a road. The dependent variable is an indicator variable that is set to 1 if a village received a road by 2005. The running variable is the population of the largest habitation in the village, and the treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to 500. Columns 1-3 show estimates from the local linear specification, and column 4 includes a quartic in population. Column 2 adds state fixed effects, and column 3 adds village level controls from the 1991 population census, which are log land area, share of land that is irrigated, distance to the nearest town, and the number of different industries in the village in 1998. The value 500 has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect. Standard errors are clustered at the district level.

by 2005. Columns 1-3 show local linear specifications. Column 2 adds state fixed effects, column 3 adds village-level controls, and column 4 adds a quartic in population. The estimates show a very large and statistically significant effect of a new feeder road on rural non-farm employment. A new feeder road more than doubles the growth of non-farm employment. Effect sizes range from 120 to 240 log points, reflecting the high variation in growth rates across villages.[8] Controls, clustering and bandwidth are the same as in Table 3.4. All regressions include an additional control for non-farm employment in 1998.

A standard regression discontinuity validity test is that the density of the running variable is

[8]We omit villages with non-farm employment less than 10 or greater than 200 in 1998. The median village has employment of 180, so this effect is not driven by growth in villages with very small levels of employment.

Figure 3.1: *First stage: population threshold and new rural roads*



The figure shows the share of habitations that received a road, by population. Each point represents approximately 1000 habitations in the top panel, and 300 habitations in the bottom panel. The PMGSY instructed states to target roads to habitations with population greater than 500, the value indicated by the solid vertical line.

continuous across the treatment threshold. The standard test recommended by McCrary (2008) cannot be applied, because the population threshold affects both the probability of receiving a new road, and selection into the RD sample.[9] The histogram of PMGSY-listed populations of habitations that received roads by 2011 (the RD sample) is displayed in Figure 3.2, with clear discontinuous increases in density at the population thresholds of 250 and 500. Figure 3.3 shows the histogram of village population in the 2001 population census, where there is no evident sign of a discontinuity at either of the PMGSY thresholds.

3.6.3 IV

In this section we present results from the IV specification described in Section 3.4.3. We first verify that within-district village population rank appears to be a valid instrument and then discuss results of this estimation strategy.

We define our sample to be villages listed as not having paved roads at the time of the 2001 Population Census. This is for two reasons. First, villages unconnected by paved all-weather roads

[9]The sample for the RD is all villages in which a PMGSY project was sanctioned by 2012. Habitations with populations greater than 500 are more likely to be prioritized and receive a road by 2005, which drives our empirical result. However, habitations with population greater than 500 are also more likely to receive a road by 2011, which drives selection into the sample. An alternate approach which would not face this issue would be to use the entire set of Indian villages as a control group - but with fewer than 40,000 out of 600,000+ villages connected, the increased probability of receiving a road under PMGSY is too small to detect.

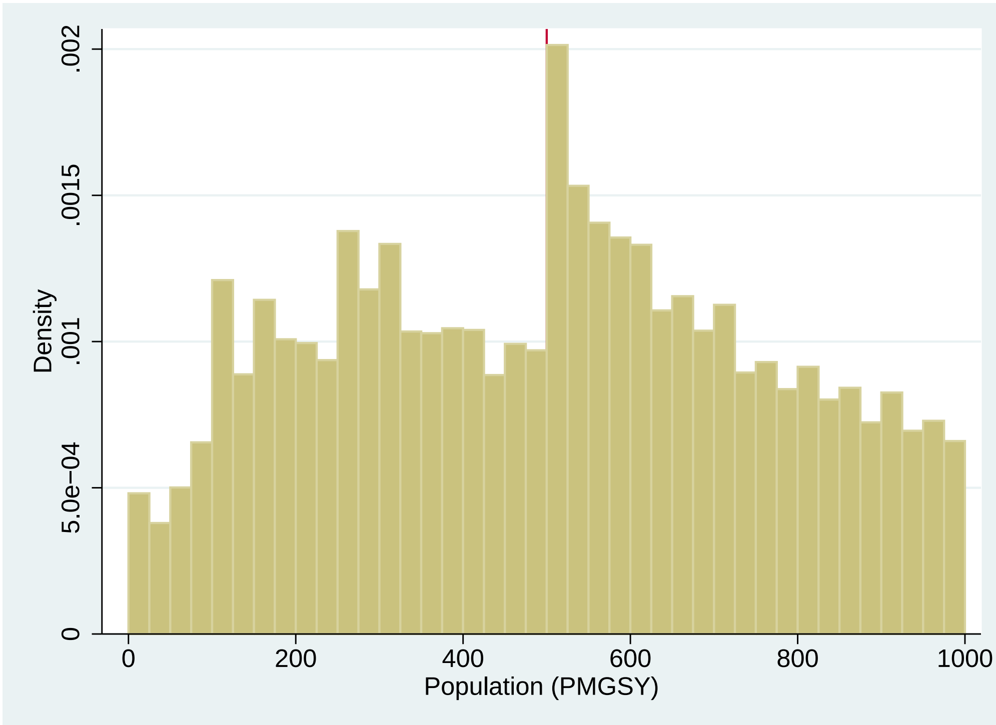
Table 3.5: *RD estimates of effect of new road on employment growth*

	(1)	(2)	(3)	(4)
r05	2.355	1.893	1.282	1.351
	(0.528)***	(0.530)***	(0.502)**	(0.678)**
Population	-0.000	-0.000	-0.000	-0.000
	(0.000)**	(0.000)*	(0.000)***	(0.000)***
Population * 1(Pop $\geq$ 500)	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Share of land irrigated			0.193	0.192
			(0.021)***	(0.022)***
Log(land area)			0.216	0.215
			(0.008)***	(0.009)***
Distance from town			-0.003	-0.003
			(0.000)***	(0.000)***
Baseline number of industries			0.013	0.013
			(0.001)***	(0.002)***
Constant	1.224	1.520	0.643	0.644
	(0.022)***	(0.046)***	(0.064)***	(0.077)***
N	19288	19288	18972	16196
r2	0.00	0.05	0.17	0.17

$^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$. The table presents instrumental variable regression discontinuity estimates of the effect of receiving a new road under the PMGSY on log non-farm employment growth from 1998-2005. The dependent variable is village level non-farm employment growth from 1998-2005. The running variable is the population of the largest habitation in the village, and the treatment indicator of population greater than 500 instruments for a village receiving a PMGSY road by 2005. Columns 1-3 show estimates from the local linear specification, and column 4 includes a quartic in population. Column 2 adds state fixed effects, and column 3 adds village level controls from the 1991 population census, which are log land area, share of land that is irrigated, distance to the nearest town, and the number of different industries in the village in 1998. The value 500 has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect. All columns include a control for log non-farm employment in 1998. Standard errors are clustered at the district level.

to the road network are listed as the highest priority under the PMGSY. Second, we can be confident that roads to such villages are in fact new roads and not upgrades of existing roads, in order to be able to more cleanly interpret the results. The sample is further restricted to villages in districts in which the PMGSY built roads to more than 5% of villages by 2005. Finally, because few villages with very high within-district population ranks (very small villages) receive roads under the PMGSY, we drop villages with population ranks above 300. These sample definitions yield a sample of 12073 eligible villages in our rank IV sample, out of which 1910 received PMGSY roads by 2005.

Figure 3.2: *Histogram of PMGSY-project population (SP)*

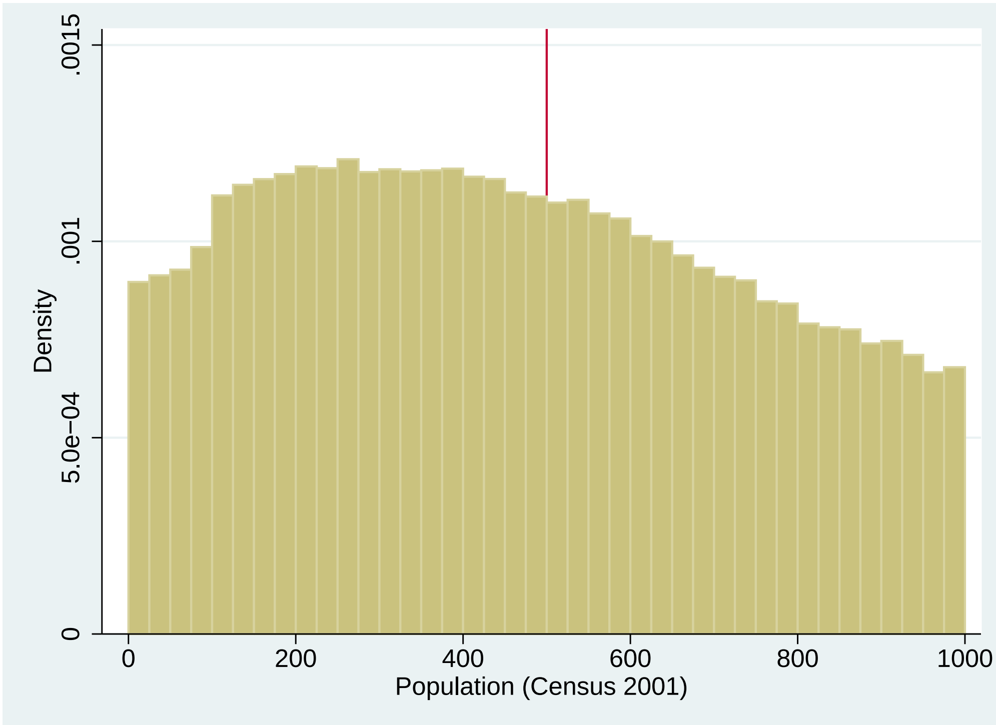


The figure displays a histogram of habitation population as recorded in the PMGSY project data. The density bulges at 250 and 500 reflect the project's prioritization of villages with population greater than 500, and then greater than 250.

First stage and reduced form

We define our rank instrument two ways. The first is a simple field rank that assigns a village a rank of the number of villages in the sample with populations greater than the village, plus one. The second creates a binary variable out of this rank that takes on the value 1 when a village rank is less than 75. Table 3.6 presents the results of the first stage, regressing an indicator for the completion of a PMGSY road by 2005 on these two rank instruments. For simplicity of presentation, the rank variable has been divided by 100. Columns 1 and 2 present the effect of rank on the probability that a village receives a road by 2005 with quadratic and quartic polynomial population controls, respectively. A reduction of rank of 100 is associated with a 4 to 5 percentage point increase in the probability of receiving a road, controlling for population. Columns 3 and 4 present the effect of our binary instrument on the likelihood of receiving a PMGSY road by 2005: being in the top 75 villages

Figure 3.3: *Histogram of population (population census 2001)*



The figure displays a histogram of village population as recorded in the 2001 population census. The PMGSY eligibility thresholds of 250 and 500 do not show obvious changes in density.

within one's district is, after controlling for population and other village characteristics, associated with an approximately 5.5 percentage point increase in the probability of receiving a PMGSY road. The results are little changed by adding additional population controls, which is reassuring for the validity of the instruments.

Table 3.7 presents the reduced form results. The dependent variable in all four columns is log nonfarm employment growth. Columns 1 and 2 show that an increase in rank of 100 (decrease in prioritization for PMGSY road) is associated with a reduction of approximately 7 log points of employment growth, with negligible change when going from quadratic to quartic population controls. Likewise, columns 3 and 4 show that having a rank of less than 75 is associated with an increase of approximately 6 log points of employment growth.

One concern is that the rank instrument may be correlated with prioritization not only in the PMGSY but in other government programs that are also carried out at the district level. We

Table 3.6: *First stage effect of rank on probability of receiving road*

	1	2	3	4
Pop rank (no road)	-0.047	-0.043		
	(0.007)***	(0.007)***		
Top pop (no road)			0.056	0.055
			(0.008)***	(0.008)***
Ln baseline employment	-0.000	-0.000	0.001	0.000
	(0.005)	(0.005)	(0.005)	(0.005)
Village pop	0.000	0.002	0.000	0.002
	(0.000)***	(0.001)***	(0.000)***	(0.001)***
Village pop ^{2}	-0.001	-0.016	-0.001	-0.017
	(0.000)***	(0.006)***	(0.000)***	(0.006)***
Village pop ^{3}		0.053		0.058
		(0.023)**		(0.023)**
Village pop ^{4}		-0.006		-0.007
		(0.003)**		(0.003)**
Pop growth 1991-2001	-0.065	-0.066	-0.066	-0.067
	(0.013)***	(0.013)***	(0.013)***	(0.013)***
Irrigation share	0.018	0.017	0.015	0.013
	(0.012)	(0.012)	(0.012)	(0.012)
Ln land area	0.009	0.008	0.009	0.008
	(0.004)**	(0.004)*	(0.004)**	(0.004)*
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Diversity (1998)	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	12073	12073	12073	12073
r2	0.11	0.11	0.11	0.11

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents first stage regressions using within-district population rank as a predictor of receipt of a PMGSY road by 2005. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 3.7: *Reduced form effect of rank on employment growth*

	1	2	3	4
Pop rank (no road)	-0.073	-0.069		
	(0.016)***	(0.016)***		
Top pop (no road)			0.060	0.059
			(0.018)***	(0.018)***
Ln baseline employment	-0.610	-0.610	-0.608	-0.608
	(0.011)***	(0.011)***	(0.011)***	(0.011)***
Village pop	0.001	0.003	0.001	0.003
	(0.000)***	(0.001)**	(0.000)***	(0.001)***
Village pop ^{2}	-0.001	-0.024	-0.001	-0.026
	(0.000)**	(0.013)*	(0.000)***	(0.013)**
Village pop ^{3}		0.087		0.094
		(0.050)*		(0.049)*
Village pop ^{4}		-0.011		-0.012
		(0.007)		(0.007)*
Pop growth 1991-2001	-0.122	-0.123	-0.123	-0.124
	(0.034)***	(0.034)***	(0.034)***	(0.034)***
Irrigation share	0.107	0.105	0.104	0.102
	(0.026)***	(0.026)***	(0.026)***	(0.026)***
Ln land area	-0.040	-0.041	-0.041	-0.042
	(0.010)***	(0.010)***	(0.010)***	(0.010)***
Distance from town	-0.002	-0.002	-0.002	-0.002
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Diversity (1998)	0.011	0.011	0.011	0.011
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
N	12073	12073	12073	12073
r2	0.33	0.33	0.33	0.33

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents reduced form regressions using within-district population rank as a predictor of log employment growth between 1998 and 2005. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

thus define a placebo sample of villages with equivalent rankings as our rank IV sample but in districts that did not construct PMGSY roads in more than 1% of villages. Table 3.8 presents the results of the same reduced form specification discussed above but in this placebo sample. We find insignificant effects of both the continuous and binary population rank instruments on village nonfarm employment growth. We take this as suggestive, if not conclusive, evidence that the village population rank affects employment growth only through its impact on the likelihood of receiving a PMGSY road.

Rank IV estimates

Table 3.10 presents the results of the IV estimation based on Equation 3.4. As in the preceding tables, the first two columns use $RANK$ as the excluded instrument for the construction of a road, while columns 3 and 4 use the binary low rank variable. We find that a new road leads to highly significant positive increases in nonfarm employment growth. Regardless of specification, a new road is associated with an increase of employment growth of over 100 log points. Table 3.9 seeks to unpackage these results, instrumenting for new road with only the binary rank variable. Column 1 estimates the effect of a new road on the number of economic establishments (in logs); in contrast to the highly significant positive results for employment growth, the estimated effect on firm count is smaller and less significant. Consistent with this result, column 2 finds that a new road is associated with a large and significant increase in firm size: the mean firm employs more than 1 additional employee than in villages that do not receive roads. Column 3 estimates the treatment effect on the percent change in economic diversity of the village. We follow Hidalgo and Hausmann (2009) in defining the diversity of a village economy as the number of industries present in the village.[10] The outcome variable is defined as the proportional change in the economic diversity of a village, top-coded at 4 (400% increase) to reduce the effect of outliers. We find that a new road is associated with a positive significant increase of approximately 0.65 industries per village. Column 4 estimates the effect of a new road on the village share of employment in informal firms, defined as those not registered with the Indian government. We find a large increase in informality, although it is unclear how best to interpret this result. One hypothesis is that high transportation costs keep firms inefficiently small and informal by denying them access to larger markets. This theory would

[10] The 1998 Economic Census reports 3 digit NIC-1987 codes for each establishment, while the 2005 Economic Census uses 4-digit NIC-2004 codes. We use correspondence tables published by the Ministry of Statistics and Programme Implementation to generate a list of 217 self-contained industries, which we use as the basis of all industry-level analysis in this paper.

Table 3.8: *Reduced form effect of rank on employment growth (placebo sample)*

	1	2	3	4
Pop rank (no road)	0.019	0.019		
	(0.033)	(0.033)		
Top pop (no road)			0.023	0.022
			(0.045)	(0.045)
Ln baseline employment	-0.585	-0.585	-0.588	-0.587
	(0.024)***	(0.024)***	(0.024)***	(0.024)***
Village pop	0.001	0.002	0.001	0.002
	(0.000)***	(0.003)	(0.000)***	(0.003)
Village pop ^{2}	-0.001	-0.013	-0.001	-0.012
	(0.000)***	(0.026)	(0.000)***	(0.026)
Village pop ^{3}		0.041		0.038
		(0.099)		(0.099)
Village pop ^{4}		-0.005		-0.005
		(0.014)		(0.014)
Pop growth 1991-2001	-0.059	-0.058	-0.058	-0.057
	(0.067)	(0.067)	(0.067)	(0.067)
Irrigation share	-0.130	-0.131	-0.133	-0.133
	(0.055)**	(0.055)**	(0.055)**	(0.055)**
Ln land area	-0.023	-0.023	-0.023	-0.023
	(0.020)	(0.020)	(0.020)	(0.020)
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Diversity (1998)	0.000	0.000	0.000	0.000
	(0.005)	(0.005)	(0.005)	(0.005)
N	2871	2871	2871	2871
r2	0.35	0.35	0.35	0.35

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents reduced form regressions using within-district population rank as a predictor of log employment growth between 1998 and 2005 for a placebo sample of villages in districts not receiving many PMGSY roads. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in less than 1% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

predict increased formalization in response to the construction of a new road, but if the employment response to falling transport costs is faster than the process of formalization, we should expect to see a short-term rise informality. Future work with the Economic Census of 2012 should enable us to test this hypothesis.

Table 3.11 considers heterogeneity of our estimated treatment effect by village characteristics. Theoretically, we expect that firms with better access productive inputs to grow most in response to a fall in transportation costs. Column 1 presents the full sample estimate of the effect of a road on total employment, while columns 2 through 7 present subsample results. We find suggestive evidence that electricity appears to be an important complementary input to a village feeder road: while we estimate a very small (if noisy) estimate for villages without electrical supply, we find significant, statistically larger effects in villages with electricity supply. This result should be interpreted with some caution, given that our identification strategy provides exogenous variation in road placement, not in the interaction of road construction and availability of electricity. Next, we estimate effects for villages above and below the median distance to the nearest town, which for our sample is 20 km. We estimate a larger treatment effect for villages far from towns, although both estimates are within one standard deviation of our full-sample estimate and are not statistically different from each other. Likewise, we find no significant difference in our estimates when dividing our sample on human capital, as measured by literacy.

Given the very high costs of infrastructure projects such as roads, it is of great interest to policymakers to understand the economic impact of such projects in terms of money spent rather than per project. The OMMS described in Section 3.5 contains data on total spending per road, which we use to construct a village road expenditure variable that takes on the value 0 if a road is not built in the village by 2005 and otherwise equals the sum of all PMGSY road spending in habitations contained in that village.[11] As our objective is to estimate the cost of job creation, we define our outcome variable to be the level change in employment between 1998 and 2005, rather than log growth as in the preceding tables. Table 3.12 presents the results of this estimation. Our estimates range, depending on the instrument used, between approximately 500 and 700 nonfarm jobs per million dollars spent on road construction, which results in a per job cost of between approximately 1430 and 2000 USD. Expressed differently, the per job cost of the PMGSY is roughly two to three

[11] We generate total road spending village in million USD, using an exchange rate of 44.06 INR per USD, the average for 2005. In the case of roads that connect multiple habitations, spending is allocated equally between the habitations.

Table 3.9: *IV effect of road on firm characteristics*

	Firm Count	Firm Size	Diversity	Informality
New road	0.510	1.423	0.654	0.313
	(0.316)	(0.422)***	(0.307)**	(0.110)***
Ln baseline employment				
Village pop	0.002	-0.002	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)**
Village pop ^{2}	-0.012	0.016	0.009	0.008
	(0.013)	(0.016)	(0.013)	(0.004)*
Village pop ^{3}	0.046	-0.056	-0.030	-0.027
	(0.051)	(0.062)	(0.050)	(0.017)
Village pop ^{4}	-0.007	0.007	0.004	0.003
	(0.007)	(0.008)	(0.007)	(0.002)
Pop growth 1991-2001	-0.109	0.111	-0.043	0.008
	(0.039)***	(0.048)**	(0.039)	(0.013)
Irrigation share	0.095	-0.055	0.137	-0.003
	(0.025)***	(0.032)*	(0.025)***	(0.008)
Ln land area	-0.025	-0.066	-0.026	0.002
	(0.010)**	(0.012)***	(0.010)**	(0.003)
Distance from town	-0.001	0.000	-0.001	0.000
	(0.000)***	(0.001)	(0.000)*	(0.000)
Diversity (1998)	0.011	0.007	-0.081	0.002
	(0.002)***	(0.002)***	(0.002)***	(0.001)***
o.Diversity (1998)			0.000	
			(.)	
N	12073	11891	12073	12073
r2	0.32	.	0.18	0.01

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. This table presents instrumental variable regression results for which the endogenous regressor is a dummy variable indicating the construction of a PMGSY road by 2005, instrumented by a dummary variable indicating that within-district, within-sample population rank is less than 75. Dependent variables for columns 1 through 4 are, respectively: count of economic establishments, mean firm size (number of employees), proportional change in economic diversity and 2005 share of employment in informal firms. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 3.10: *IV effect of road on employment growth*

	1	2	3	4
New road	1.542	1.613	1.071	1.081
	(0.400)***	(0.446)***	(0.356)***	(0.364)***
Ln baseline employment	-0.610	-0.610	-0.609	-0.609
	(0.014)***	(0.014)***	(0.013)***	(0.013)***
Village pop	0.000	-0.000	0.000	0.001
	(0.000)	(0.002)	(0.000)**	(0.002)
Village pop ^{2}	0.000	0.002	-0.000	-0.007
	(0.000)	(0.017)	(0.000)	(0.015)
Village pop ^{3}		0.002		0.032
		(0.066)		(0.059)
Village pop ^{4}		-0.001		-0.005
		(0.009)		(0.008)
Pop growth 1991-2001	-0.021	-0.017	-0.052	-0.052
	(0.046)	(0.048)	(0.042)	(0.043)
Irrigation share	0.078	0.078	0.087	0.087
	(0.032)**	(0.032)**	(0.029)***	(0.029)***
Ln land area	-0.054	-0.054	-0.051	-0.051
	(0.013)***	(0.013)***	(0.012)***	(0.012)***
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)**	(0.001)**
Diversity (1998)	0.011	0.011	0.011	0.011
	(0.003)***	(0.003)***	(0.002)***	(0.002)***
N	12073	12073	12073	12073
r2	.	.	0.15	0.15

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table presents instrumental variable regression results for which the dependent variable is log employment growth (1998-2005) and the endogenous regressor is a dummy variable indicating the construction of a PMGSY road by 2005, instrumented by population rank. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 3.11: *IV effect of road on employment growth by village characteristics*

	All	No Electricity	Electricity	Far	Near	Low Literacy	High Literacy
New road	1.081	0.019	1.728	1.293	0.846	1.415	0.559
	(0.364)***	(0.699)	(0.511)***	(0.537)**	(0.536)	(0.493)***	(0.651)
Ln baseline employment	-0.609	-0.680	-0.608	-0.628	-0.596	-0.618	-0.616
	(0.013)***	(0.024)***	(0.019)***	(0.019)***	(0.017)***	(0.018)***	(0.019)***
2001 Population	0.001	-0.004	0.005	-0.003	0.005	-0.001	0.003
	(0.002)	(0.005)	(0.002)*	(0.003)	(0.002)**	(0.003)	(0.002)
2001 Population ^{2}	-0.007	0.040	-0.042	0.027	-0.047	0.008	-0.021
	(0.015)	(0.044)	(0.022)*	(0.023)	(0.020)**	(0.022)	(0.021)
2001 Population ^{3}	0.032	-0.169	0.175	-0.109	0.194	-0.020	0.075
	(0.059)	(0.159)	(0.087)**	(0.089)	(0.078)**	(0.085)	(0.081)
2001 Population ^{4}	-0.005	0.025	-0.025	0.016	-0.028	0.002	-0.010
	(0.008)	(0.021)	(0.012)**	(0.012)	(0.011)***	(0.012)	(0.011)
Pop growth 1991-2001	-0.052	-0.068	0.021	0.040	-0.157	-0.109	0.028
	(0.043)	(0.063)	(0.060)	(0.064)	(0.056)***	(0.058)*	(0.063)
Irrigation share	0.087	0.132	0.042	0.032	0.175	0.217	-0.075
	(0.029)***	(0.050)***	(0.051)	(0.040)	(0.043)***	(0.043)***	(0.039)*
Ln land area	-0.051	-0.045	-0.046	-0.064	-0.030	-0.046	-0.036
	(0.012)***	(0.018)**	(0.019)**	(0.016)***	(0.017)*	(0.016)***	(0.022)
Distance from town	-0.001	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001
	(0.001)**	(0.001)**	(0.001)	(0.002)	(0.001)*	(0.001)	(0.001)
Diversity (1998)	0.011	0.015	0.011	0.014	0.008	0.011	0.011
	(0.002)***	(0.005)***	(0.004)***	(0.003)***	(0.003)**	(0.004)***	(0.003)***
N	12073	3106	6453	6646	5427	6886	5187
r2	0.15	0.34	.	0.08	0.22	0.05	0.26

$^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$. This table presents instrumental variable estimates of the effect of the construction of a PMGSY road on log employment growth (1998-2005) for different subsamples in order to estimate heterogeneous treatment effects. All subsamples are defined based on 2001 Population Census characteristics. Column 1 presents full sample results. Columns 2 and 3 are for villages with and without electricity supply, respectively. Column 4 presents results for villages with below median distance from the nearest town (20km). Column 5 presents results for villages above median distance from a town. Column 6 presents results for villages with below median literacy, and Column 7 for villages with above-median literacy. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

times India's GDP per capita in 2005 ($731.70). How should we interpret this result? As the Economic Census was conducted in late 2005 and early 2006, it is safe to assume that nearly all of the roads listed as completed by 2005 were finished by the time of data collection; thus, these results should be understood not as temporary jobs related to road construction, but rather short- to medium-term effects. These estimates also assume that the net employment effect of PMGSY spending in locations not receiving roads is zero. This is a strong assumption – one we intend to investigate in future work – as there may be either crowd-in or crowd-out of economic activity in villages and towns that are now better connected to PMGSY villages. Finally, all estimates assume that control villages do not receive any PMGSY road spending, an unlikely result given the multiple years that road construction generally requires, the multiple habitations per PMGSY road project and the many roads that were completed in 2006 and 2007. For this last reason, we consider our estimates to be likely lower bounds on the true effect of road construction in this context.

3.7 Conclusion

Universal access to paved roads, easily taken for granted in many rich countries, is far from a reality for many of the world's poor. High transportation costs inhibit gains from the division of labor, economies of scale and comparative advantage. Nevertheless, little is known of the economic effects of road provision on rural economic activity. Theoretically, roads could facilitate migration to urban areas and the specialization of village economies in agriculture; alternatively, lower transportation costs could cause the emergence and expansion of rural economic activities, with potentially large consequences for economic development, urbanization and the spatial distribution of economic activity.

In this paper we estimate the economic impacts of a large-scale program that seeks to provide near universal access to paved "all-weather" roads in rural India. The program design provides two sources of exogenous variation to allow us to overcome the usual challenge of endogeneity of large infrastructure projects. First, the program calls for highest priority to be given to habitations above population thresholds, which may be 250, 500 or 1000 depending on the area. This creates a discontinuity in the probability of receiving a road at these cutoffs, allowing us to use a fuzzy regression discontinuity design to estimate the impact of these roads. A second identification strategy takes advantage of the fact that habitations are prioritized not only by population but also by the relative population rank within a district: a village of a certain size is more likely to receive a

Table 3.12: *IV effect of road spending on level change in employment*

	1	2	3	4
Road cost (million USD)	481.839	479.030	709.091	715.636
	(303.054)	(324.152)	(333.751)**	(341.696)**
Employment (EC98)	-0.356	-0.356	-0.356	-0.356
	(0.029)***	(0.029)***	(0.030)***	(0.030)***
Village pop	0.004	0.042	-0.002	0.000
	(0.012)	(0.145)	(0.013)	(0.148)
Village pop ^{2}	0.059	-0.300	0.076	0.001
	(0.034)*	(1.358)	(0.035)**	(1.382)
Village pop ^{3}		1.422		0.486
		(5.428)		(5.516)
Village pop ^{4}		-0.199		-0.092
		(0.778)		(0.789)
Pop growth 1991-2001	-1.652	-1.679	-0.373	-0.339
	(2.850)	(2.925)	(3.095)	(3.134)
Irrigation share	4.674	4.661	4.815	4.845
	(2.336)**	(2.341)**	(2.429)**	(2.439)**
Ln land area	-5.673	-5.673	-6.180	-6.180
	(1.123)***	(1.138)***	(1.206)***	(1.201)***
Distance from town	-0.093	-0.093	-0.097	-0.098
	(0.042)**	(0.042)**	(0.044)**	(0.044)**
Diversity (1998)	-0.018	-0.019	0.000	0.002
	(0.247)	(0.248)	(0.250)	(0.251)
N	11833	11833	11833	11833
r2	0.08	0.08	.	.

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. This table presents instrumental variable regression results for which the dependent variable is log employment growth (1998-2005) and the endogenous regressor is the amount of spending on a PMGSY road by 2005. For roads are not completed by 2005, this number is 0. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

road if there are fewer larger villages within its district than an equivalent village that has a lower population rank. We instrument for road construction using this rank, conditioning on population, to provide a second set of estimates of the impact of village feeder roads.

We find that the provision of a new, paved village approach road produces significantly faster nonfarm employment growth. Rather than a simple proportional expansion of existing economic activity, new industries form in villages that receive roads. Firm size also grows, which we take as supporting evidence of the theory that in the presence of high transportation costs, firms are inefficiently small. The responsiveness of nonfarm employment growth to lower transportation costs appears to be strongest in villages that also have access to the supply of electricity, suggesting significant complementarities between these different infrastructural investments. Finally, we provide some of the first well-identified estimates of the cost effectiveness of rural road construction: one job is created for every $1400 to $2000 in road construction costs, suggesting very high returns to such investments.

Future work will allow us to further disentangle the channels by which rural roads promote village nonfarm employment. The 2011 Population Census, not yet available at the village level, will provide the data necessary to examine the impact of roads on migration and agricultural employment, neither of which is covered in the Economic Census. We intend to use the 2012 Economic Census, the collection of which is still in progress, to differentiate between the short- to medium-run effects presented in this paper and sustained, longer-run changes to village economic activity.

References

Acemoglu, Daron and James A Robinson, "A Theory of Political Transitions," *The American Economic Review*, 2001, *91* (4), 938–963.

Aidt, Toke, Miriam A Golden, and Devesh Tiwari, "Incumbents and Criminals in the Indian National Legislature," 2011.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Students today, teachers tomorrow: Identifying constraints on the provision of education," *Journal of Public Economics*, April 2013, *100*, 1–14.

Angrist, Joshua D and Adriana D Kugler, "Rural Windfall or a New Resource Curse? Coca, Income and Civil Conflict in Colomiba," *The Review of Economics and Statistics*, 2008, *90* (2), 191–215.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 1999, *114* (2).

Asher, Sam and Paul Novosad, "Digging for Development: Natural resource wealth, mining booms and local economic development," 2013.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," 2012.

Banerjee, Abhijit, "Contracting Constraints , Credit Markets and Economic Development," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 3*, number September, University of Cambridge, 2003, pp. 1–44.

____, Donald P. Green, Jeffery McManus, and Rohini Pande, "Are Poor Voters Indifferent to Whether Elected Leaders are Criminal or Corrupt? A Vignette Experiment in Rural India," 2012.

____, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," 2012.

____, Santosh Kumar, and Rohini Pande, "Connectivity and Rural Development: Examining India's Rural Road Building Scheme," 2012.

Bazzi, Samuel and Christopher Blattman, "Economic Shocks and Conflict: The (Absence of?) Evidence from Commodity Prices," 2011.

Behrman, Jere R, "Labor Markets in Developing Countries," *Handbook of Labor Economics*, 1999, *3*, 2859–2939.

Besley, Timothy, "Savings, Credit and Insurance," *Handbook of Development Economics*, 1995, *III.*

Bhattacharyya, Sambit and Roland Hodler, "Natural Resources, Democracy and Corruption," *European Economic Review*, 2010, *54* (4).

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini, "The Political Resource Curse," *American Economic Review (forthcoming)*, 2013.

Bruckner, Markus and Antonio Ciccone, "International Commodity Prices, Growth and the Outbreak of Civil War in Sub-Saharan Africa," *The Economic Journal*, 2010, *120* (May), 519–534.

——, ——, **and Andrea Tesei**, "Oil Price Shocks, Income, and Democracy," *The Review of Economics and Statistics*, 2012, *94* (2).

Brunnschweiler, Christa N. and Erwin H. Bulte, "The resource curse revisited and revised: A tale of paradoxes and red herrings," *Journal of Environmental Economics and Management*, May 2008, *55* (3), 248–264.

Burgess, Robin and Rohini Pande, "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment," *The American Economic Review*, 2005, *95* (3), 780–795.

Burke, Paul J and Andrew Leigh, "Do Output Contractions Trigger Democratic Change?," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 124–157.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri, "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone," 2012.

Chemin, Matthieu, "Welfare Effects of Criminal Politicians: A Discontinuity-Based Approach," *Journal of Law and Economics*, 2012, *55* (3), 667–690.

Cole, Shawn, Andrew Healy, and Eric Werker, "Do voters demand responsive governments? Evidence from Indian disaster relief," *Journal of Development Economics*, March 2012, *97* (2), 167–181.

Cotet, Anca M and Kevin K Tsui, "Oil and Conflict: What does the cross country evidence really show?," *American Economic Journal: Macroeconomics*, 2013, *5* (1), 49–80.

Dercon, Stefan, "Growth and shocks: evidence from rural Ethiopia," *Journal of Development Economics*, August 2004, *74* (2), 309–329.

Djankov, Simeon, Caroline Freund, and Cong S. Pham, "Trading on time," *The Review of Economics and Statistics*, 2010, *92* (1), 166–173.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *October*, 2012, (October).

Dube, Oeindrila and Juan F Vargas, "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," 2012.

Fafchamps, Marcel and Forhad Shilpi, "Cities and Specialisation: Evidence from South Asia," *The Economic Journal*, 2005, *115* (503), 477–504.

Fan, Shenggen and Peter Hazell, "Returns to Public Investments in the Less-Favored Areas of India and China," *American Journal of Agricultural Economics*, 2001, *83* (5), 1217–1222.

Fishman, Ram Mukul, "Climate Change , Rainfall Variability , and Adaptation through Irrigation : Evidence from Indian Agriculture," 2011.

Fisman, Raymond, Florian Schulz, and Vikrant Vig, "Private Returns to Public Office," 2013.

Foster, Andrew D and Mark R Rosenzweig, "Agricultural Productivity Growth, Rural Economic Diversity and Economic Reforms, India 1970-2000," *Economic Development and Cultural Change*, 2004.

Geological Survey of India, *Mineral Atlas of India*, Kolkata: Geological Survey of India, 2001.

Gilje, Erik, "Does Local Access To Finance Matter?: Evidence from U.S . Oil and Natural Gas Shale Booms," 2012.

Gollin, Douglas, “Agricultural Productivity and Economic Growth,” in “Handbook of Agricultural Economics,” Vol. 4, Elsevier B.V., 2010, chapter 73, pp. 3825–3866.

____ **and Richard Rogerson**, “Agriculture, Roads and Economic Development in Uganda,” 2010.

____ **, Stephen Parente, and Richard Rogerson**, “The Role of Agriculture in Development,” *American Economic Review. Papers and Proceedings*, February 2002, *92* (2), 160–164.

Guiteras, Raymond, “The Impact of Climate Change on Indian Agriculture,” 2009.

Haggblade, Steven, Peter Hazell, and Thomas Reardon, “The Rural Non-farm Economy: Prospects for Growth and Poverty Reduction,” *World Development*, October 2010, *38* (10), 1429–1441.

Hidalgo, César and Ricardo Hausmann, “The building blocks of economic complexity.,” *Proceedings of the National Academy of Sciences of the United States of America*, June 2009, *106* (26), 10570–5.

Hidalgo, F Daniel, Suresh Naidu, Simeon Nichter, and Neal Richardson, “Economic Determinants of Land Invasions,” *The Review of Economics and Statistics*, 2010, *92* (3), 505–523.

Hornbeck, Richard and Dave Donaldson, “Railroads and American Economic Growth: A "Market Access" Approach,” 2012.

____ **and Pinar Keskin**, “Local Economic Spillovers from the Agricultural Sector: Evidence from Access to Ogallala Groundwater,” 2012.

Imbens, Guido W. and Thomas Lemieux, “Regression discontinuity designs: A guide to practice,” *Journal of Econometrics*, February 2008, *142* (2), 615–635.

Indian Bureau of Mines, *Indian Minerals Year Book 2010*, Nagpur, India: IBM Press, 2011.

Isham, Jonathan, Michael Woolcock, Lant Pritchett, and Gwen Busby, “The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth,” *The World Bank Economic Review*, August 2005, *19* (2), 141–174.

Iyer, Lakshmi and Anandi Mani, “Traveling Agents: Political Change and Bureaucratic Turnover in India,” *The Review of Economics and Statistics*, 2012, *94* (3), 723–739.

Jensen, Robert, “The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector,” *Quarterly Journal of Economics*, 2007, *122* (3).

Jensenius, Francesca, “Power, Performance and Bias: Evaluating the Electoral Quotas for Scheduled Castes in India,” 2013.

Kapur, Devesh and Milan Vaishnav, “Quid Pro Quo: Builders, Politicians, and Election Finance in India,” 2011.

Kelly, Thomas D. and Grecia R. Matos, “Historical Statistics for Mineral and Material Commodities in the United States,” Technical Report, US Geological Survey 2013.

Khandker, Shaidur R. and Gayatri B Koolwal, “Estimating the Long-term Impacts of Rural Roads A Dynamic Panel Approach,” 2011.

____ **, Zaid Bakht, and Gayatri B. Koolwal**, “The Poverty Impact of Rural Roads: Evidence from Bangladesh,” *Economic Development and Cultural Change*, 2009, *57* (4), 685–722.

Kumar, Hemanshu and Rohini Somanathan, “Mapping Indian Districts Across Census Years, 1971-2001,” 2009.

La Porta, Rafael, Florencio Lopez de Silanes, and Andrei Shleifer, “Government ownership of banks,” *The Journal of Finance*, 2002, *57* (1).

Laakso, Markku and Rein Taagepera, “‘Effective’ Number of Parties: A Measure with Application to West Europe,” *Comparative Political Studies*, 1979, *12* (3).

Lall, Somik V., Zmarak Shalizi, and Uwe Deichmann, "Agglomeration economies and productivity in Indian industry," *Journal of Development Economics*, April 2004, *73* (2), 643–673.

Lanjouw, Jean and Peter Lanjouw, "The rural non-farm sector: issues and evidence from developing countries," *Agricultural Economics*, 2001, *26*, 1–23.

Limao, Nuno and Anthony J Venables, "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, February 2008, *142* (2), 698–714.

Mian, Atif and Amir Sufi, "What Explains High Unemployment? The Aggregate Demand Channel," 2012.

Michaels, Guy, "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 2008, *90* (November), 683–701.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112* (4).

Ministry of Rural Development, "Working Group on Rural Roads in the 12th Five Year Plan," Technical Report 2012.

Moretti, E., "Local multipliers," in "American Economic Review. Papers and Proceedings" 2010.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, "Industrialization and the big push," *The Journal of Political Economy*, 1989, *97* (5), 1003–1026.

National Rural Roads Development Agency, "Pradhan Mantri Gram Sadak Yojana - Operations Manual," Technical Report, Ministry of Rural Development, Government of India 2005.

Nunn, N. and N. Qian, "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment," *The Quarterly Journal of Economics*, June 2011, *126* (2), 593–650.

Persson, T and G Tabellini, *Political Economics: Explaining Economic Policy*, Cambridge, MA: MIT Press, 2000.

Raballand, Gael, Rebecca Thornton, Dean Yang, Jessica Goldberg, Niall Keleher, and Annika Muller, "Are Rural Road Investments Alone Sufficient to Generate Transport Flows? Lessons from a Randomized Experiment in Rural Malawi and Policy Implications," 2011.

Rajan, Raghuram and Luigi Zingales, "Financial Dependence and Growth," *The American Economic Review*, February 1998, *88* (3), 559–586.

Rajeevan, M and Jyoti Bhate, "A High Resolution Daily Gridded Rainfall Data Set (1971-2005) for Mesoscale Meteorological Studies," Technical Report August, Indian Meteorological Department, Pune, Maharashtra, India 2008.

Redding, By Stephen J and Daniel M Sturm, "The Costs of Remoteness : Evidence from German Division and Reunification," *American Economic Review*, 2008, (1999), 1766–1797.

Robinson, James, Ragnar Torvik, and Thierry Verdier, "Political foundations of the resource curse," *Journal of Development Economics*, April 2006, *79* (2), 447–468.

Ross, Michael L., "Does Oil Hinder Democracy?," *World Politics*, 2001, *53* (3), 325–361.

Sachs, Jeffrey and A.M. Warner, "Natural Resource Abundance and Economic Growth," in G Meier and J Rauch, eds., *Leading Issues in Economic Development*, Vol. 3, New York: Oxford University Press, 1995.

Srivastava, A K, M Rajeevan, and S R Kshirsagar, "Development of a High Resolution Daily Gridded Temperature Data Set (1969-2005) for the Indian Region," *Atmospheric Science Letters*, 2009, *10*, 249–254.

Storeygard, Adam, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," 2012.

Tata Environmental Research Institute, "Overview of Mining and Mineral Industry in India," Technical Report 185 2002.

The World Bank, "A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005," Technical Report, The World Bank, Washington, DC 2007.

Vaishnav, Milan, "Social Divisions, Credibility and Criminality: Political Selection in India," 2012.

____ **and Neelanjan Sircar**, "Core or Swing ? The Role of Electoral Context in Shaping Pork Barrel," 2012.

Van der Ploeg, Frederick, "Natural resources: curse or blessing?," *Journal of Economic Literature*, 2011, *49* (2), 366–420.

Venables, Anthony J., "Spatial disparities in developing countries: cities, regions, and international trade," *Journal of Economic Geography*, January 2005, *5* (1), 3–21.

World Bank, "Agriculture for Development," Technical Report, World Bank 2008.