Essays on Development Economics

Juan Pablo Rud

Thesis submitted in the fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Economics

London School of Economics and Political Science

November 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I, Juan Pablo Rud, declare that the work presented in this thesis is my own. No part of this dissertation contains material previously submitted to examiners of any other university. Where information has been derived from other sources, I confirm that this has been duly indicated in the thesis.

Abstract

This thesis presents three papers that contribute to the measurement and understanding of the process of economic development. In particular, I deal with issues of significant importance in the current literature in development economics: the provision and regulatory institutions of infrastructure, firms and industries' behaviour and performance, and the process of human capital accumulation and its link to gender issues.

In Chapter 2 I investigate the effect of electricity provision on industrialization using a panel of Indian states from 1965-1984. To address the endogeneity of investment in electrification, I use the introduction of a new agricultural technology intensive in irrigation (the Green Revolution) as a natural experiment. As electric pumpsets are used to provide farmers with cheap irrigation water, I use the uneven availability of groundwater to predict divergence in the expansion of the electricity network and, ultimately, to quantify the effect of electrification on industrial outcomes. I present a series of tests to rule out alternative explanations that could link groundwater availability to industrialization directly or through other means than electrification. Overall, the uneven expansion of the electricity network explains between 10 and 15 percentage points of the difference in manufacturing output across states in India.

In Chapter 3 I explore how firms in India cope with the erratic and expensive provision of electricity. In a model that combines upstream regulation with downstream heterogeneous firms in a monopolistic competition framework, I investigate the role of the electricity regulator's preferences and the economic environment (i.e. regulation and openness) in determining the decision to adopt a captive generator of electricity and industries' aggregate productivity. I show that a firm's productivity, the electricity regulator's disregard for the well-being of industrial producers consuming electricity and greater industry protection from competition are associated with greater adoption of captive power. The mechanisms I propose are present for a representative repeated cross-section sample of Indian firms in the 1990s, with heterogeneous effects along dimensions such as location.

In Chapter 4 I investigate the effect of the Green Revolution on rural literacy and rural women's employment and literacy levels, using a panel of 254 districts for census years, before and after the introduction of the high yield variety (HYV) seeds. Even though the new technology has been shown to increase returns to education, aggregate effects on literacy are ambiguous a priori, if claims are correct that the process excluded most poor farmers and that mechanization replaced women labour and their effects are strong. I find robust evidence that the increase in adoption of the new seeds is associated with increases of around 2 percentage points in literacy levels. The effects are only present for treated cohorts. Additionally, I find no evidence of a Green-Revolution related increase in the gender gap: even though results indicate that the

percentages of working and literate women in rural India fall over time, a greater intensity in HYV is shown to mitigate this trend.

Acknowledgments

I am grateful to my supervisors Robin Burgess and Guy Michaels for their continuous support and guidance. I am also indebted to faculty and fellow PhD students at the Economic Organisation and Public Policy Programme (EOPP), STICERD and the Department of Economics at the London School of Economics for fruitful conversations, suggestions and ideas. In particular, Steve Redding, Oriana Bandiera, Andrea Prat, Fernando Aragon, Paolo Masella, Irma Clots-Figueres and Raja Kali have been extremely helpful at different stages of this project. Participants and discussants in the conferences and seminars where I presented have also contributed to this research.

I am also thankful for the financial support provided by the Department of Economics at LSE, the Anglo-Jewish Association and the Overseas Research Students Awards Scheme at different stages of my Doctoral degree and to support staff at the Department of Economics and STICERD.

My parents, Silvia and Eugenio, and my sister Flavia played an important part in this process and I am forever grateful to them for their unconditional support, as I am to my friends from Buenos Aires and to those I made in London. And to Emma, for getting my energies flowing in the late stages. Last, but not least, the development and completion of this project would not have been possible without the indefatigable love, companionship, patience and encouragement offered by Laura, to whom I dedicate this thesis.

Contents

Chapter 1
Introduction

The accumulation of physical and human capital remains a fundamental yet difficult task in the objective of reducing poverty and increasing standards of living in most of the developing world. Individuals in developing countries do face an economic environment full of constraints, such as the absence or imperfection of credit markets. However, in many cases, the decision not to invest in physical and human capital can be explained by low private returns. In particular, in this thesis I investigate two sources: the lack of adequate infrastructure, that hinders the birth and growth of firms, and the prevalence of traditional farming technologies with low returns to education. The economic structure of a large number of underdeveloped countries suggests the relevance of these sources of low private returns. For example, the regions of Sub-Saharan African and South Asia where–out of the combined population of 2 billion people–more than 30% live with less than a dollar a day, 70% live in rural areas and around 30% of GDP comes from the agricultural sector (World Bank (2005; 2007)). Both infrastructure and human capital indicators depict a bleak outlook even when compared to other developing regions, such as Latin America: only 24% of people in Africa and 43% in South Asia have access to electricity (vs. 89% in Latin America), around 35% in both areas have access to sanitation (vs. 74% in Latin America), 34% and 65% of the rural population in Africa and South Asia, respectively, live within 2 kilometres of paved roads (vs. 54% in Latin America) and fixed lines and mobile connections reach around 6% of the population on average (vs. 42% in Latin America). The picture looks very similar for human capital indicators: according to the Millennium Development Goals, in 2003 only 61% of children in the relevant age group in Sub-Saharan Africa and 87% in South Asia had completed primary school (vs. 96% in Latin America). Similarly, illiteracy among youth aged 15 to 24 in 2003 was above 30% in both regions, while in Latin America, it was below 10%.

Even though these indicators are conclusive indicators of underdevelopment, the understanding of the underlying drivers of economic failure and the prescription for successful policies are still a work in progress. This is due to the fact that the simultaneity of phenomena makes it very hard to disentangle causal mechanisms and the identification of causal effects. The objective of this thesis is to address some of these analytic challenges. In doing so, we can learn about issues of significant importance in the current literature in development economics, including the provision and regulatory institutions of infrastructure, firms and industries' behaviour and performance, and the process of human capital accumulation and its link to gender issues.

More specifically, the main issues I discuss in the following three chapters revolve around two main features that have characterised the economy of India in the last fifty years but could be used to understand the process of economic development more generally. The first one is the electricity sector, an important input in industrial and agricultural production that has been underperforming for a long time in India, as highlighted by the World Bank (2000) and by the US Department of Energy (2003) (among others) as one of the main constraints to development in India. For example, the Investment Climate Survey done by the Asian Development Bank and the World Bank in 2002 reports information for around 1800 firms and shows that, on average, firms face a power outage or a power surge every other day and that more than 60% of firms have turned to captive power generators. This phenomenon of self-generation of electricity is recurrent in other developing countries in South Asia and Africa. For example, similar surveys in Bangladesh, Sri Lanka and Kenya show that more than 70% of surveyed firms use a captive generator. In Pakistan, the rate is around 40%; in Tanzania, 55%; and in Uganda, 36%. In all of these countries, including India, the inadequacy of electricity is considered an important constraint for firms (Asian Development Bank (2002)).

The second recurrent phenomenon highlighted in this thesis is the Green Revolution, i.e. the introduction of a new agricultural technology in India in the mid 1960s consisting of seeds intensive in the use of irrigation and fertilizers. In the following chapters, I make use of two features of this technological change: the increase of electricity provision in rural areas to provide the required timely irrigation (chapters 2 and 3) and the technology's complementarity with human capital (chapter 4). The lessons learned from the transformations associated with the Green Revolution in India can shed light on some of the benefits and hazards of adopting similar techniques in the many developing countries where the agricultural sector employs more than 60% of the work force, its share of GDP is around 30% on average, and where illiteracy rates are in most cases above 30% (for the "agricultural-based countries" as defined by the World Development Report (2007)).

In Chapter 2, I investigate the effect of electricity provision on industrial development by looking at a panel of Indian states between 1965 and 1984. An interesting feature in that period is that each state in India has its own State Electricity Board (SEB)–politically dependent on the state government and independent of the central government–that is in charge of expanding the electricity network. This allows me to compare the industrial performance of different states and link it to their electrification efforts. But the econometric identification of this effect is subject to the endogeneity concerns of reverse causality and unobserved time varying effects that might produce spurious results. To deal with these issues I make use of the start of the Green Revolution in the mid 1960s, where the successful introduction of High Yield Variety (HYV) seeds was determined by geographical characteristics such as groundwater availability, allowing farmers to pump water so as to provide timely irrigation to the new seeds (see Foster and Rosenzweig (2008), for example). As electric pumpsets were employed to provide farmers with cheap irrigation water, I use the uneven availability of groundwater to predict divergence in the expansion of the electricity network and, ultimately, to quantify the effect of electrification on industrial outcomes. That means that I can link the time varying effects of a time invariant characteristic (groundwater availability) to the expansion of the electricity network to address the endogeneity concerns that would bias least squares estimates. Additionally, I present a series of tests to mitigate concerns about alternative explanations that could link groundwater availability to industrialization directly or through other means than electrification, such as an increase in urbanization, expenditure in manufacturing goods and credit availability. This chapter contributes to the growing literature on infrastructure, reinforcing the idea that geographic characteristics can be used to instrument for investment in infrastructure projects, as in Duflo and Pande (2007).

Chapter 3 explores how firms in India cope with the erratic and expensive provision of electricity. I develop a model that combines monopoly regulation of the provider of a public utility, as in Laffont and Tirole (1993), with downstream heterogeneous firms that use the infrastructure good as an input in a monopolistic competition framework. As in Melitz (2003) and Bernard et al. (2006), I assume firms differ exogenously in their productivity. In particular, I investigate the role of the electricity regulator's preferences (i.e. the importance attributed to industrial consumers in the objective function) in determining the quality adjusted price charged to industrial consumers and its effect on firms' decisions to adopt a captive power generator (presented as a cost reducing device). The existence of this technology changes the configuration of the industry, since the exposure to high price or low quality is different for firms that have adopted and those who have not. I show that higher electricity tariffs increase the rate of adoption by two means: the productivity level of the marginal adopting firm is lower and the productivity level of the marginal surviving firm is greater. I subsequently show that the economic environment (i.e. openness and regulation) also plays a part in determining the decision to adopt a captive generator. A more protected environment translates into higher final good prices, allowing less productive firms to recoup the fixed cost of the electricity generator. I test the mechanisms I propose in the model using a repeated cross-section representative sample of Indian firms for the years 1990, 1994 and 1997, combined with data at the state and industry level. I first analyse the role of a regulator's preferences in pricing decisions and check whether industries face higher electricity tariffs in states where the agricultural sector's need for electricity is greater (i.e. where the Green Revolution is more prevalent). The evidence suggests that states with greater intensity in HYV are associated with industries paying significantly more for their network electricity. I subsequently show that the probability of adopting a captive power generator is greater in states that charge industries more for their electricity and provide a more unreliable service (as measured by outages and network energy losses). In line with the model, I find evidence that the more protected the industry, the more likely the adoption of captive power is. Adopting firms are also on average bigger and more productive.

In Chapter 4, I investigate the relationship between the Green Revolution, rural literacy and gender status using a panel of 254 Indian districts for the census years of 1961 (before the introduction of the new technology), 1971 and 1981. There is a general consensus that the benefits of the Green Revolution were more likely to be reaped by more prosperous and more educated households, and that returns to education increased with the adoption of the new high yield varieties (HYV), as shown by Foster and Rosenzweig (1996). However, many authors (see Dhanagare (1989), for example) argue that only relatively well-off farmers and large landowners benefited from this new technology while the many poor and small farmers, tenants or agricultural labourers might have not experienced any benefits. The inequality associated with the technical change raises the concern as to whether the increased returns to education generated effects of any significance at the aggregate level. I am also interested in the concern expressed by many authors (see Sridhar (2004) for a review) that the Green Revolution has worsened the status of women, because mechanization has reduced the demand for women labour and their returns to education. To test the whether HYV adoption is associated with more rural literacy, I use a similar identification strategy to Duflo (2001) and compare the change in literacy for rural cohorts that were in primary schooling age during the Green Revolution (according to districts' HYV intensity) with respect to the same cohort in 1961. I find that literacy has increased in all districts on average, but a district at the mean of the HYV adoption distribution would have produced an extra increase in rural literacy of around 2 percentage points per cohort. The identifying assumption is that there are no omitted time-varying district characteristics correlated with HYV adoption and rural literacy that would generate spurious results. To deal with this concern in the identification strategy I use two control groups, for which I find no effects: older rural cohorts to capture pre-Green Revolution trends in literacy and same cohorts in urban areas, to capture contemporaneous district-wide trends in literacy. Results are robust to the inclusion of channels other than HYV that could explain the divergence in literacy rates, such as migration, strong presence of population belonging to Scheduled Castes or Scheduled Tribes (groups recognized by the Constitution of India as previously disadvantaged), and state trends and state-wide investments in primary schools. Additionally, I show that there is an average increase in the literacy gender gap and a decrease in female labour participation in all of India. HYV intensity mitigates these effects, but it does not reverse them: female participation in the labour force, employment in the agricultural sector and literacy gap for treated cohorts fare better in districts with greater adoption of HYV seeds. These findings reconcile conflicting evidence in the literature regarding gender issues and the Green Revolution. I suggest that not taking into account the negative trend for all of rural India in female labour and educational outcomes might lead to misleading conclusions when analysing gender status in HYV-intensive districts.

Chapter 2
Electricity Provision and Industrial Development: Evidence from India

2.1 Introduction

The adequate supply of infrastructure goods is increasingly acknowledged as one key factor in generating a conducive environment for industrial and economic development. The World Bank currently directs 35% of their lending portfolio to infrastructure projects with the idea that "infrastructure has a central role in the development agenda and is a major contributor to growth, poverty reduction and achievement of the Millennium Development Goals" (World Bank (2005)). However, average expenditure on infrastructure for developing countries is only around 3% of GDP. Among developing countries, South Asia and Sub-Saharan Africa perform particularly poorly in a range of infrastructure goods, such as water and sanitation, telecoms and electricity access. The latter, according to investment climate surveys, is one of the greatest obstacles to industrial development in India and the rest of southern Asia: around 43% of the population have access to the electricity network, compared to at least a 90% access in Latin America, Eastern Europe and East Asia.

India's poor infrastructure in general, and the power sector in particular, is seen as one of the reasons behind India's slow export growth during the 1990s, limiting their comparative advantage in labour intensive products (World Bank (2000)). Public agencies, economic journalists and sector analysts[1] are among those who point to the electricity sector's poor performance as heavily affecting the growth and development possibilities of the Indian economy since independence, to the point that "the poor quality of electricity has been the single greatest deterrent to India's economic growth and development" (US DoE (2003)). Yet the history of Indian electrification is not one of uniform failure. For example, around 10% of villages on average were electrified by 1965, with some states like Tamil Nadu close to 50% and others such as Assam or West Bengal less than 3%. In 1984, the average increased to over 75% and some states like Punjab achieved full village electrification.

I use this variation across regions and over time within India to investigate the effect of electricity provision on industrial development, by examining a panel of Indian states between 1965 and 1984. India's federal political organization gives each state full responsibility for creating, expanding and administrating the electricity network, from electricity generation to retail. This provides significant variation in the extent of the physical network and industrial outcomes, allowing me to test whether the gap in infrastructure provision is associated with unequal industrialization levels.

Assessing and quantifying the impact of investment in electrification on economic outcomes in the long run is a difficult task, since it is hard to determine the underlying driving force. The resulting endogeneity concerns arise because of re-

---

[1] For example, "(..) electricity is unusable for industry and of such poor quality that power surges wreck equipment." (The Economist, 22/9/2005). See also TERI (1999).

verse causality and unobserved state characteristics–such as the business or political climate–that might explain why some states are better prepared to provide a better electricity network than others. In such cases, OLS estimates would be biased. To overcome this problem, my empirical strategy aims to find and use appropriate instruments for the expansion of the electricity network in order to consistently estimate its impact on industrial development in a two-stage least squares (2SLS) strategy. For this purpose, I use the start of the Green Revolution–an agricultural technology intensive in irrigation introduced in India in the mid 60s–as a natural experiment. The successful introduction of High Yield Variety (HYV) seeds was determined by geographical characteristics such as soil characteristics and groundwater availability, that allowed farmers to pump water so as to provide timely irrigation to the new seeds (see Foster and Rosenzweig (2008), for example). As electric pumpsets were used to provide cheap irrigation water, the uneven availability of groundwater across states was subsequently followed by an expansion of the electricity network. That means that I can link the time-varying effects of a time invariant characteristic (groundwater availability) to the expansion of the electricity network to address the endogeneity concerns that would bias OLS estimates. After testing for the power and validity of the instruments, I find that 2SLS estimates provide a more credible estimate than least squares for this empirical question. This paper contributes to the growing literature on infrastructure, reinforcing the idea that geographic characteristics can be used to explain differential investments in expensive infrastructure projects. For example, Duflo and Pande (2007) instrument dams placement in India using information on rivers' gradients and Dinkelman (2008) instruments electrification in South Africa with land gradients. My strategy exploits the complementarity between groundwater and electrified pumpsets after the start of the Green Revolution to address the endogenous placement of rural electricity.

The main concern regarding my empirical strategy is that groundwater or HYV adoption could affect industrial outcomes through channels other than electrification. For example, greater rural incomes can increase the demand for manufacturing goods[2] or savings that translate into credit availability. An alternative story is that positive shocks to agricultural productivity release cheap labour to be employed in the manufacturing sector (see Matsuyama (1992)) where an urbanization process follows. Foster and Rosenzweig (1996) show that returns to education increased with the Green Revolution and it could be the case that an increase in overall levels of human capital or state expenditure in development projects (such as health and education) might be driving the uneven industrialization process rather than electrification. To see whether the time varying effect of groundwater availability could be linked to these alternative mechanisms, I reproduce the 2SLS specification, instrumenting for variables such as credit, expenditure, urbanization rates, literacy and development expenditure. For a series of different specifications, none of these channels shows

---

[2] See for example Voigtländer and Voth (2006) where positive agricultural shocks are associated with increases in the share of expenditure in manufacturing goods if consumers have non-homothetic preferences.

robust results, reducing the concerns that my strategy captures confounding effects. The electrification link remains the most powerful of all.

Another concern is that abundant groundwater can have a direct effect on industrialization, if firms require water for production. Using district level data within states, I show that there is no link between employment in the manufacturing sector and groundwater availability once power availability has been controlled for. I also deal with the concern that groundwater availability is spuriously correlated with electrification and that a "lucky draw" in the groundwater measure is driving the results. I carry out a placebo experiment where I randomly assign values to my measure of groundwater availability and show that the probability that my results are driven by a lucky draw is very low (i.e. 0.4%). Results are robust to these alternative explanations and suggest that a standard deviation difference in the reach of the grid explains between 10 and 15 percentage points difference in industrial output, in addition to more factories and greater output among smaller firms. Magnitudes are substantial and underline the potential economic benefits of investing in the expansion of the electricity network.

The rest of the chapter is organised as follows. Section 2.2 provides some background of the electricity sector in India and its relation to the Green Revolution. Section 2.3 develops the empirical strategy where I identify the set of instruments and test their power and validity. Subsequently, I show results of the effect of electricity indicators on manufacturing outcomes. Section 2.4 deals with alternative explanations and Section 2.5 presents additional robustness checks. Section 2.6 concludes.

2.2 Electrification in India and the Green Revolution

The empirical contribution of this paper, i.e. to quantify the effect of electrification in industrial development, draws both from the organisation of the electricity sector in India and from the uneven introduction of an agricultural technology that increased the demand for powered irrigation. From a theoretical perspective, Murphy et al. (1989b) show that infrastructure can be an important component of a "big push" industrialization process. By overcoming coordination problems that arise because infrastructure goods are used by many sectors at the same time, the public provision of infrastructure contributes to the development of other markets, since it has the effect of reducing the total production costs of the other sectors. In particular, infrastructure can reduce costs in producing and marketing (i.e. transporting) goods. In this paper, by focusing on the provision of electricity, the reduction in production costs becomes the underlying mechanism that links infrastructure to industrialization. In general, electrification could lower prices and induce more consumption of manufacturing goods[3]. In a setting with heterogeneous firms, a selection effect might be

---

[3] A look at the input-output coefficient matrix for industries in the early 1990s India shows that for many industries, such as heavy chemicals, cement or non-ferrous metal, electricity is more than 10% of their total costs and can be as high as 17% (Ministry of Industry, Government of India (1993)).

present, wherein smaller and less productive firms can break even only when electricity is available.

With these mechanisms in mind, I investigate the effect of electricity infrastructure on industrial development, by looking at a panel of Indian states between 1965 and 1984. The cross-state variation in electricity provision in India has been used to explain differences in productivity, return and investment rates (Dollar et al. (2002), (2003)), and divergence in growth rates and income levels (Bandyopadhyay (2003); Sachs et al. (2002)). To the best of my knowledge, mine is the first attempt to quantify the long run effects of electrification on industrialization by addressing the endogeneity of network expansion. The provision of electricity is analysed at the state level, since there is a constitutional arrangement in India that ensures state independence from the central government in designing electricity policies. The Indian Constitution establishes in a handful of articles the distribution of legislative powers between the Central Government and the States, specifying a concurrent list of activities where both levels of government have the power to intervene. These shared competencies include non-economic legislation (e.g., criminal law) and economic regulation (e.g., labour markets or price controls). The electricity sector falls in this last category. The Electricity Supply Act created in 1948 fully vertically integrated State Electricity Boards (SEB), in charge of coordinating electricity generation, transmission and distribution and commercialization at the state level, whose Board members are appointed by the state government (Electricity Supply Act, Chapter III, 5.2).

Figure 1 shows a correlation between changes in agricultural electrification and manufacturing output in the analysed period for 15 Indian states. Note that states where HYV seeds overtook traditional seeds (Punjab, Haryana and Tamil Nadu) are amongst the faster growing states and have experienced a larger expansion of their electricity network.

This correlation between Green Revolution and electrification is at the heart of the identification strategy. The Green Revolution followed persistent food shortages in newly independent India, where farming activity was mainly for subsistence, characterized by the use of primitive techniques and was rainfed (Chakravarti (1972)). The Third Five Year Plan laid out by the Planning Commission, covering the years from 1961-2 to 1965-6, set ambitious targets in terms of agricultural production: e.g. foodgrains should grow by 30%, other products like jute, by 55%, all in the belief that "with the achievement of these targets, the economy will become self-sufficient in the supply of foodgrains". After the failure of some programmes such as the Intensive Agricultural District Programme (IADP) and the Intensive Agricultural Areas Programme (IAAP) in 1966-7 the introduction of the High Yielding Variety Seed Programme provided the expected breakthrough, by providing farmers with hybrid seeds scientifically adapted to India's domestic conditions. The Green Revolution–in order to become a "revolution"–depended on a series of factors, especially the adequate and timely supply of water. In 1961, only around 20% of the total cropped area was under some form of irrigation. In the following twenty years, the increase in irrigated areas was more than 50%, with very uneven distribution across regions. In the period under consideration, Haryana and Punjab–states at the forefront of the Green Revolution–achieved a share of irrigated area around 60% and 90% respectively, while other relatively rich states like Gujarat and West Bengal were around the 25% mark.

The nature and the depth of irrigation development constitutes the fundamental argument in my empirical strategy. Its expansion goes hand in hand with HYV seeds adoption, meaning that states that were trying harder to introduce the new technology were creating a demand for irrigation. Irrigation could be provided by using canals supplying water from a dam or reservoir or the installation of deep or shallow powered tubewells. Bharadway (1990) notes that "the rate of increase of irrigation by wells/tubewells was higher than that by canals, and accelerated remarkably during the period 1969-1980 when there was a spurt in private tubewells, especially in the late sixties". McGuirk and Mundlak (1991) emphasize this point further by showing that in Punjab government-provided canal irrigation dominated only until 1968; afterwards, private wells and tanks became more important in terms of land coverage. As a matter of fact, this acceleration in the usage of tubewells is closely related to the availability of electricity in rural areas. As Bharadway explains, "a tubewell uses diesel or electricity operated pumps to lift the water (from the water table) and can manage greater depths and irrigate bigger commands (than dug wells.)" It has to be noted that electricity was the cheapest option for farmers: the National Commission on Agriculture, Ministry of Agriculture and Irrigation of India (1976) estimated that operating diesel pumpsets was two times more expensive. As the efficient utilization of the tubewells was conditional on electricity availability, states that were to deepen the Green Revolution had an incentive to expand their electricity network simultaneously. McGuirk and Mundlak show an example by mentioning that "from 1965/6 to 1979/80, power generation in Punjab increased by over 240% and by 1976 all villages had access to electricity. (...) The fastest-growing source of demand for electricity during this period was agriculture. A large part of this increase (from 14.5% in 1960/1 to 47% in 1979/80 of the total available electricity) was as a source of power for the growing numbers of tubewells."

A natural question to ask is why farmers' demand for electricity was satisfied in HYV intensive states. Two mechanisms could explain why farmers' need of electricity was satisfied by their state governments. The first one is linked to the preferences of the median voter. In the period analysed, rural population in any given state is at least 65% of total population, and almost 80% on average (topping 90% in three states), in a country that has been democratic and federal since its independence in 1947. There is some evidence that voters in states intensive in HYV like Punjab, Haryana and Tamil Nadu moved away from the Congress Party and voted for regional parties. However, the salience of other issues (e.g., religion) and the fact that incumbent parties could adapt their policies to the new needs of the rural population makes this case harder to support. A second mechanism is supported by anecdotal evidence: the endogenous formation of lobbies. The first circumstance that seems important to note is that "increasing food shortages and mounting concern for immediate gain in production led to the shift in developmental priorities" (Sharma and Dak, (1989)), such that expanding the production possibilities in the agricultural sector was deemed fundamental. That circumstance gave farmers, in particular big ones, an unprecedented political clout over state governments. Dhanagare (1989) estimates that "the prosperity unleashed by the Green Revolution was distributed differentially, putting the small and marginal farmers at a relative disadvantage. The high cost/high yield technology called for capital investments beyond the means of a majority of small and marginal farmers." It might be the case that a stronger economic position translated into political leverage. Tongia (2003), for example, stresses that rural electrification "has swayed in strong political winds." Gulati and Narayanan (2003) also show that what they call "the subsidy syndrome in Indian agriculture" after the introduction of the new seeds in the late 1960s became a fundamental instrument of economic policy where available electricity was one of its main channels. In the case of Punjab, Simms (1988) describes a situation in which "farmers are aware of the power they hold due to their strategic importance in the national economy (...) Political action has led to extensive changes in rural Indian Punjab".

Even though the importance of the political economy of infrastructure provision cannot be understated, this chapter does not enquire into its details. Rather, it relies on the identification of geographic characteristics, such as groundwater availability, that facilitated the expansion of the electricity network with the start of the Green Revolution. I turn to this analysis in the next section.

2.3 Identification Strategy

The main objective of this paper is to quantify the effect of electricity provision on industrial outcomes using a panel of fifteen Indian states[4] where consistent information on the electricity network is available over the period 1965-1984. That is, by obtaining a consistent estimate of λ in equation (2.1), where an indicator of industrial performance (y_{st}) is expressed as a function of an indicator of electricity supply (e_{st}):

$$y_{st} = \alpha_{0s} + \alpha_{1t} + \lambda e_{st} + \alpha_3 \mathbf{X}_{st} + \mu_{st} \tag{2.1}$$

The regression includes time varying state controls (X_{st}), state fixed effects included to control for persistent and constant features within states, and year fixed effects to control for shocks common to all states. To control for serial correlation, standard errors are clustered at the state level[5]. The explained variable, y_{st}, captures industrial development or performance per state and year. If infrastructure reduces

---

[4] Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal.

[5] As pointed out by Wooldridge (2003) and Bertrand et al. (2004), clustering standard errors might not be appropriate when the number of states is not large enough. Results shown throughout are robust to a variety of standard error corrections (e.g. collapsing time series information, bootstrap or GMM methods to correct for heteroskedasticity and autocorrelation.)

costs or improves productivity, an increase in electricity availability should have a positive impact on manufacturing output and fixed capital per capita. Additionally, electrification could provide the boost that new or small firms need to survive. I can test this by using data on the number of factories and on the small sector (e.g. value added and fixed capital).

Different variables are represented by e_{st} to capture the reach and network development in states' electricity supply at a point in time. The total number of consumers per type (farmers and factories) connected to the network relative to the state population and the average connected load (i.e. the maximum electrical power consumption per user) measure the depth of the network.

To control for demographic features and human capital, I include the log of population density and the proportion of rural and literate population. In order to investigate whether the infrastructure variable is actually picking up the effect of more credit availability (thus allowing more manufacturing firms to start up), I control for financial development by including indicators such as the log of real per capita total credit. As noted by Laffont (2005), government inefficiencies and corruption increase the marginal cost of raising funds and constrains the ability of the executive to invest in infrastructure. It follows that the indicators of electricity network development might actually be capturing these political economy variables that could drive both dependent and independent variables. To address this concern, state controls include political party outcomes (party allegiance of the Chief Minister and votes for the Congress Party in parliamentary elections) and real per capita expenditure on development sectors such as health and education. The latter is not only aimed at capturing states' investment in human capital, but also the ability and/or willingness of the state government to enhance the economic environment. All variables and sources are described in Appendix 2.A.

Descriptive statistics are presented in Table 1 for all main variables used in the regression analysis. In particular, I focus on the number of agricultural units connected to the network, manufacturing output, and the incidence of HYV at the state level. Table 1 shows that there is significant variation in all three, with high values of standard deviations relative to the mean and substantial differences between minimum and maximum. Of particular interest is the number of agricultural connections to the electricity network, whose mean increases tenfold during the period analysed. Additionally, states show substantial variation in geographic characteristics and initial conditions, a fact that will be exploited in the next section.

In the estimation of the coefficient of interest, λ, endogeneity is a major concern for quantifying the effect of infrastructure on output. The presence of correlation between the error term and the explanatory variable could be explained by reverse causality (e.g. more industrialized states can afford investments in expanding the electricity network) or omitted variables (e.g. unobserved changes in the institutional environment that drive both industrialization and electrification), and would introduce a bias to the estimation. At this point I am agnostic about the sign of the bias, since alternative stories could work in both directions. For example, phenomena such as a pro-business environment could explain a positive correlation with both the outcome and the explanatory variable, biasing the estimate upwards. Alternatively, the formation of a strong pro-rural electrification lobby clashing with urban industrialists, for example, could explain a positive correlation with electrification and a negative correlation with industrialization, resulting in a negative bias.

The challenge is to find a set of instruments that can explain the differential pattern of investment in infrastructure that is not correlated with industrial output by other means. I argue in this paper that the introduction of High Yielding Varieties (HYV) in Indian agriculture in the mid-60s is at the root of the dissimilar development of the electricity network across states, without being related to contemporaneous or potential industrial development. HYV seeds needed reliable irrigation and access to electricity was a fundamental input for farmers to pump water. Because some concerns about the endogeneity of HYV adoption remain, I will first investigate how geo-climatic characteristics and some initial conditions–such as groundwater availability and irrigation prior to the Green Revolution, respectively–have affected the diffusion of the new agricultural technology[6]. Subsequently, by exploiting the fact that states followed either a path of adoption of the new agricultural technology–creating a demand for rural electricity, or a path of traditional farming without creating the need for expansion of the electricity network, I will ex-

---

[6] A similar idea was already discussed by Evenson and McKinsey (1999) where the authors analyse the viability of HYV seeds as a function of the interaction between climatic and technological (e.g., infrastructure) variables.

plore whether the proposed instruments satisfy the assumptions that validate the IV strategy. That is, I will investigate whether state specific characteristics leading to the uneven introduction of new seeds starting in 1966 is correlated with the development of the electricity network (i.e. power of the instruments) and whether this relation can be used to assess the differential impact in manufacturing outcomes (i.e. validity of the instruments). The underlying idea is that as the electricity network expanded to satisfy farmers' needs, industrial producers have benefited.

2.3.1 Instruments

The adoption of HYV seeds in rural India has been subject to a number of studies. As noted by Sharma and Dak (1989) or Kohli and Sing (1997), among others, there are many institutional factors that may claim a share of responsibility in the adoption of this new technology. A list would include credit availability or price incentives (to buy the new seeds, fertilizers, pesticides or new machinery), land matters (titling, distribution and size) or human capital. The fact that these variables might be correlated with (or determined by) industrial output calls for an exogenous source of variation that explains HYV successful adoption. The answer lies in geographic and climatic characteristics.

In a paper analysing the importance of human capital, Foster and Rosenzweig (1996) point out that the technological change could be considered exogenous, given the fact that the innovation industries were different from the beneficiaries of the new seeds and that "a feature of the Green Revolution in India is that the ability to exploit the new seeds profitably was substantially different across India because of exogenous differentials in local soil and weather conditions". On a similar note, Evenson and McKinsey (1999) show that climatic and soil conditions, when combined with basic investments in infrastructure, resulted in the diffusion of HYV seeds. In this section I further investigate these claims by looking at the evolution of HYV adoption as a function of initial conditions, soil and climate characteristics (in particular, groundwater) and their interaction.

HYV adoption drivers: geography and initial characteristics

To investigate whether soil or climate characteristics can be linked to the adoption of the new agricultural technology when the Green Revolution started in 1966, I run a regression of the following form:

$$HYV_{st} = \beta_{0s} + \beta_{1t} + \sum_{k=1966}^{1984} \gamma_k (S_s * T_k) + \delta \mathbf{X}_{st} + \epsilon_{st} \tag{2.2}$$

where HYV_{st} is the proportion of cultivated land with HYV seeds in state s at time t, S_s captures states' time-invariant geographic characteristics, and T_k is a year dummy equal to 1 whenever $k = t$. State controls, X_{st}, are included to control for the above mentioned characteristics suspected to be correlated with the diffusion of HYV seeds, such as the proportion of literate population, the log of real per capita development expenditure, the log of real per capita credit availability and population density. Additionally, state and time fixed effects are included and errors are clustered at the state level to deal with serial correlation. The coefficients of interest in equation (2.2) are γ_k, where positive and significant values would indicate that states with more adequate S_s adopted a greater proportion of HYV seeds. Increasing values of γ_t would also indicate that the difference was growing over time. Intuitively, the appropriate characteristics for seeds to grow would trigger a path of divergence across states in HYV adoption.

Three different state characteristics are used for S_s: to account for the availability of abundant groundwater that could be obtained with electric pumpsets and used to provide timely and adequate irrigation, I use the proportion of districts at the state level that have an aquifer thicker than 150 metres. This measure is the most important for my empirical strategy, since groundwater availability has been the major source of electrified irrigation after the Green Revolution (see, for example, Foster and Rosenzweig, (2008)). Other geographic characteristics include the quality of the land, where I use the proportion of districts per state with a topsoil depth of at least 3 metres. Finally, to account for aridness, I use the average annual rainfall at the state level. Sharma and Dak (1989) maintain that "HYV was directed towards areas that were irrigated and not areas that relied on rainfall. The rainfed areas comprising 70% as against 30% of irrigated areas remained outside the fold of green revolution." That means that HYV was directed towards drier areas, leaving states with abundant rainfall to specialise (relatively) on rainfed crops.

Sharma and Dak (1989) also claim that "(..) the new technology was tilted towards areas with better quality lands, assured irrigation facilities and more developed infrastructure". Before the Green Revolution started, all agriculture was rainfed and farmers in drier states had to rely on irrigation, like canals or pumpsets. This initial "curse" that induced a state to look for alternative sources of water might explain why some states, who were better prepared to reliably irrigate their crops, ended up introducing these new high-yielding but sensitive water-dependent seeds. Chakravarti (1972) notes that "farmers under HYV cannot afford to take chances with erratic rainfall, but 70% of the cropped area in India has rainfall too low or too unreliable to permit their use even during the main cropping season, and only 20% of the area is irrigated. Even where the annual rainfall is heavy, the available moisture is insufficient for crop production during winter and premonsoonic seasons. The HYV has been adopted mainly in areas with well developed irrigation facilities." To account for this interaction between geographic characteristics and initial levels of infrastructure, I also run a regression of the form:

$$\begin{aligned} HYV_{st} &= \beta_{0s} + \beta_{1t} + \sum_{k=1966}^{1984} \gamma_{1k}(S_s * IC_{s59} * T_k) + \gamma_2(S_s * \mathrm{T}_k) \\ &\quad + \gamma_3(IC_{s59} * \mathrm{T}_k) + \delta(\mathrm{X}_{s59} * \mathrm{T}_k) + \epsilon_{st} \qquad (2.3) \end{aligned}$$

where IC_{s59} is a measure of infrastructure at least 5 years before the new seeds were introduced, captured by two variables: proportion of irrigated land and agricultural electricity connections per capita, both measured in 1959. The main distinction in the specification of equations (2.2) and (2.3) is related to the introduction of state controls. In the latter case, as I am testing the effect of initial conditions such as irrigation, I am concerned that variation across states in the initial values of other characteristics are driving the results. For example, it might be the case that what mattered for HYV adoption was the initial level of human capital (e.g., literacy). In the former regression, I am focusing on time invariant characteristics, so I want to control for changes in potentially relevant variables. Also note that equation (2.3) also includes S_s and IC_{s59}, implying that lower levels of interactions are included and results captured by vectors γ_2 and γ_3. This means that values of γ_{1k} would now represent the marginal levels of adoption for states whose more appropriate geographical characteristics for HYV diffusion were complemented with better initial infrastructure.

The estimation of regression (2.2) provides point estimates of the yearly impact of groundwater availability on HYV adoption, as captured by the coefficients γ_t. Results are shown in Figure 2 for specifications with and without state controls. All estimated coefficients are significant at the 1% level and show that states with better access to groundwater have become relatively more intensive in the use of HYV seeds[7]. Figure 2 also plots estimates for γ_{1t} in equation (2.3), where I use the measure of groundwater availability interacted with initial levels of irrigation. A similar result is obtained when using initial incidence of electricity in rural areas as a mea-

---

[7] Unreported results show a similar pattern for the measure of soil quality and average rainfall, even though significance levels drop to 5%. For the latter, coefficients are negative, sustaining the previous point that arid areas attracted more HYV adoption.

sure of initial conditions (unreported). The positive, significant and increasing coefficients show a marginal effect for states that on top of having groundwater availability also had greater initial agricultural facilities. This is consistent with the idea that a combination of good geo-climatic conditions and investment in infrastructure fosters the adoption of technological change in the agricultural sector. Additionally, the increasing gap over time suggests that initial characteristics triggered a self-reinforcing process of HYV seeds adoption.

Instruments power

The last section provided evidence that the depth in the diffusion of the Green Revolution can in part be attributed to geographic characteristics, in particular to the measure of groundwater availability. The empirical strategy here will exploit their role in the differential adoption and success of HYV seeds in Indian states as a shock to their decision to expand their electricity grid, exogenous to the needs or the performance of their industrial sector. If the instrument is not perfectly uncorrelated with the error term, the consistency of the IV estimator relative to OLS hinges on the power of the instrument, i.e. on how important the correlation between the instruments (namely, $S_s * T_k$ and $IC_{s59}*$ $S_s * T_k$), and the instrumented variable, electricity provision (e_{st}), is[8]. To test the power of the instrument, I exploit the time dimension of the panel to test whether the timing at which some states have improved their electricity reach significantly coincides with the start of the Green Revolution.

---

[8] Note that at this point I am only using variation of geographical characteristics and not of HYV adoption.

A first specification looks only at the time dimension and tests whether there is a structural change in the provision of electricity with the start of the Green Revolution, using data on the electricity network that starts in 1950. To do this, I run

$$e^{b}_{st} = \beta_{0s} + \theta_1(t - 1950) + \theta_2(t - 1966) + \varepsilon_{st} \tag{2.4}$$

where the estimated θ_1 and θ_2 provide information on whether the evolution of electricity provision before and after 1966, respectively, increases significantly. Additionally, I introduce a test of the power of the instruments by interacting them with the time terms as follows:

$$e^{bs}_{st} = e^{b}_{st} + \theta_1(t - 1950) * S_s + \theta_2(t - 1966) * S_s \tag{2.5}$$

Additionally, I run (2.4) and (2.5) using state controls, available only from 1958. Finally, to check whether there is first hand evidence of the instruments having explanatory power for the evolution of the variable of interest in the second stage–i.e. industrialization–I run similar reduced form regressions using the log of manufacturing output per capita on the left hand side (available only after 1960). Even though the pre-Green Revolution period is very short, these regressions might provide valuable information on subsequent evolution, when states are divided along geographical characteristics linked to HYV adoption.

Table 2 shows results for six different specifications. In column (1), there is no apparent change in electrification between 1950 and 1966. A different pattern emerges after 1966, where the number of agricultural connections increases with time. Column (2) interacts the measure of groundwater availability with both pre and post Green Revolution time trends. Before 1966 there was no significant difference between states with high and low levels of groundwater availability, and neither coefficient is statistically different from 0. After the Green Revolution started, however, both sets of states increased their number of agricultural connections to the electricity network, even though the trend was considerably greater in magnitude for states with higher groundwater availability. Columns (3) and (4) reproduce these results when controlling for time-varying state characteristics. Column (4) suggests that in the build up to 1966 states with abundant groundwater were on a slightly positive trend of electrification, but then accelerated after 1966. Otherwise, coefficient magnitudes remain similar to results in columns (1) and (2). When analysing changes in industrial output, columns (5) and (6) show interesting results. There is no apparent change before and after the Green Revolution when all states are put together, as in column (5). However, in column (6) it becomes apparent that the industrial output in states with thick aquifers was falling behind before 1966, but that this trend turned around with the Green Revolution. These results suggest that when states are divided along the dimension of characteristics adequate for HYV adoption, a wedge in both electrification and industrialization emerges, adding confidence to the suitability and power of the chosen instruments.

The second test consists of a difference-in-difference regression to see whether states with better characteristics S_s have observed a differential expansion of their electricity network after the Green Revolution started in 1966. The regression is of the form:

$$e_{st} = \beta_{0s} + \beta_{1t} + \gamma(S_s * P_{66}) + \varepsilon_{st} \tag{2.6}$$

where P_{66} is a dummy equal to 1 for all years after 1966. This regression provides a simple before-and-after analysis, without looking at time trends. As in the previous table, time-varying state controls (X_{st}) are included for specifications after 1958 only. A positive and significant coefficient for the interaction term $S_s * P_{66}$ would mean that states whose characteristics are associated with greater HYV adoption have also, on average, observed a greater expansion of their electricity network after 1966.

Additionally, a specification without fixed effects of the form

$$e_{st} = \beta_s S_s + \beta_P P_{66} + \gamma(S_s * P_{66}) + \varepsilon_{st} \tag{2.7}$$

would add some relevant information to the picture of the electricity expansion across states. The coefficient β_s represents the difference across groups before the Green Revolution started. A result where the coefficient is not significantly different from zero would suggest that electricity availability was not statistically different across states before the introduction of the new seeds. The coefficient β_P captures how states with relatively lower values of S_s performed after the start of the Green

Revolution. A positive and significant coefficient would mean that these states significantly expanded their network on average after 1966.

Results are shown in columns (1) to (4) in Table 3. In line with the idea that states with more abundant groundwater expanded their electricity reach more with the start of the Green Revolution, the estimated coefficients for the interaction terms are all positive and significant. Magnitudes for electrification remain similar, even in the restricted sample with state controls. Results in columns (2) and (4) refer to equation (2.7), where state and time fixed effects are dropped and estimates of the coefficients for S_s and P_s can be obtained. Only the interaction term is positive and significantly different from 0 in both cases. Other results seem to lose significance once state controls are introduced. The insignificance of β_s suggests that the mean agricultural reach of the electricity network was not substantially different across states in the years before the Green Revolution. An insignificant β_P suggests that the mean agricultural connections to the grid for the states with lower values of S_s did not change after the Green Revolution, once other factors are controlled for. This result can probably be explained by the great variation across states without abundant groundwater (which drives up standard errors.)

To see whether this wedge in the evolution of electrification according to states' geographical characteristics also holds for the second stage outcome, i.e. industrialization, columns (5) and (6) in Table 3 show results for a difference in difference regression where manufacturing output is the explained variable. In both cases, states with thick aquifers observed a differential increase in industrial output. It is also interesting to see in column (6) that before the Green Revolution, these states were on average relatively less industrialized, as the negative coefficient on groundwater availability shows.

Finally, a third test consists in analysing whether the progress of the electricity network has shown a similar pattern of divergence as the one observed for the HYV adoption results in the previous section. To test for this, I follow a similar methodology as in equations (2.2) and (2.3), but instead use the number of agricultural producers' electricity connections per 1000 people as the explained variable, and the same measures of geographical characteristics (i.e. access to abundant groundwater) and initial conditions (i.e. irrigated area in 1958).

Figure 3 shows three different specifications, all using the measure of groundwater availability. In the first one, I use the measure of electrification on the left hand side, starting in 1950, with state and year fixed effects but no state controls. In the second regression, the data start in 1958 and I include state controls. In the third, I interact groundwater availability with initial irrigation, including all lower levels of interactions. Results show that states with more suitable geographical characteristics have expanded the electricity reach in rural areas after the Green Revolution. Even though in two out of the three sets of results the divergence seems to start just before 1966, coefficients are not significantly different from zero and the slope becomes steeper after 1966 only, suggesting an acceleration in electrification in states with abundant groundwater. Coefficients are significantly different from 0–at least at the 5% level, but most at the 1% level–only after the start of the Green Revolution. These findings provide more evidence towards establishing a link between characteristics that determined the innovation process in the agricultural sector and the subsequent evolution of the electricity supply network.

This section has shown that the period following the introduction of HYV seeds in 1966 is contemporaneous with a divergence in the electricity supply across states, especially when states are divided along a measure of groundwater availability[9]. Additionally, a similar pattern emerges for a set of reduced form regressions of the outcome of interest–manufacturing output–on the same characteristics, supporting the identification strategy. The next section will build on these findings in a two-stage least square (2SLS) strategy to obtain an estimate of the impact of electrification on measures of industrial development.

2.3.2 OLS and Two-Stage Least Squares Results

In this section I present OLS estimates of λ in equation (2.1) and for the 2SLS specifications where e_{st} is replaced by $\hat{e}_{st} = f(S_s * T)$(and by $\hat{e}_{st} = f(S_s * IC_{59s} * T)$ when I include initial conditions). That is, I instrument the electricity indicator using the measure of groundwater availability that, interacted with year dummies, has shown explanatory power with respect to states' divergence in both HYV diffusion

---

[9] Indicators of rainfall and soil quality produce the same pattern of results throughout this section, even though they show lower power than the measure of groundwater availability.

and electricity expansion. Additionally, since the adoption of HYV is the channel I claim is at work, I also instrument e_{st} with $\hat{e}_{st} = f(H\hat{Y}V_{st}(S_s * T))$, where $H\hat{Y}V_{st}$ is a generated instrument. Basically, I use the cross state variation in S_s interacted with year dummies to predict the value of HYV_{st}, and use these state and time-varying predicted values ($H\hat{Y}V_{st}$) as an instrument for the electricity indicator.

In results presented in Table 4, I first look at the effect of electrification on the log of real manufacturing output per capita, by using the number of farms connected to the electricity network per 1000 people as the explanatory (and instrumented) variable. Columns (1) and (2) show the OLS result with and without state controls (respectively), and find a positive and significant correlation between electrification and industrialization. It is reassuring to find a positive and strongly significant correlation between electrification and industrial output once other time-varying factors and state fixed effects are controlled for. However, there are reasons to interpret this result with extreme care. As mentioned before, a source of positive bias could be the presence of reverse causality or unobserved state characteristics, such as the business environment, that are positively correlated with both industrial outputs and infrastructure provision. Sources of bias towards zero could be any unobserved process driving rural electrification positively, but at the expense of industrial producers (such as skewed electricity pricing policies) or the presence of attenuation bias because of measurement error in the electricity indicator (produced by illegal connections, theft of electricity, etc. that tend to be widespread in rural areas, see Tongia (2003)).

Columns (3) to (7) show the key results in this chapter, where I use the time varying effects of groundwater availability as an instrument for electrification. In Columns (3) and (4) the number of agricultural electricity connections is instrumented by the interactions between year dummies and the measure of abundant groundwater availability, controlling for state-time-varying characteristics only for the latter specification. In both cases, the coefficients are positive and strongly significant. What we learn from the different magnitudes of the estimates in columns (3) and (4) is that the nature of the bias changes if we don't control for state variable characteristics. Column (5) uses the triple interaction, adding the proportion of net irrigated area before the Green Revolution, and again finds a significant positive coefficient, similar in magnitude to Column (4). Additionally, in columns (3) to (5), I can carry out over-identification tests of the instruments that pass with a great margin, with p-values above 90% in all cases. This suggests that instruments are uncorrelated with the error term and correctly excluded from the second stage. In column (6), the number of agricultural connections is instrumented with the net proportion of area cropped with HYV seeds. There is a small drop in the magnitude and in the significance of the estimated coefficient. But HYV adoption might suffer from similar endogeneity problems that could bias the results. Column (7) uses the same instruments as Column (4) to predict the values of HYV adoption that I subsequently use as a generated instrument. Results are consistent with previous findings both in significance and magnitude. Across specifications estimated values of λ are positive and significant at the 1% level when electrification is instrumented. In all cases, first stage F-tests pass comfortably at the 1% level, in line with results in the previous section where I found a strong relationship between states' time-invariant geographical characteristics, HYV adoption and subsequent expansion of the electricity network.

The estimates of λ also suggest that the impact of electrification is substantial: an increase in one standard deviation in the number of rural connections implies an increase of around 13.5% in manufacturing output. It is important to note that, given the great variation in electricity provision across states, a state at the mean of the distribution would need to double its rural electrification to observe such an effect. For example, Bihar in 1984 had a real manufacturing output per capita of Rs 1206 and 2.64 farms connected to the network per 1000 people, while Haryana had Rs 2694 and 19.45 connections, respectively. If Haryana's network depth was reproduced in Bihar, the model predicts that the gap in manufacturing output between these two states would be reduced by almost 75 percentage points[10]. Bihar's network incidence covered barely 200,000 farms in 1984, and reaching Haryana's level would imply taking the number of total agricultural connections to more than 1,200,000. Even though serving an additional million farming units would be very expensive, the potential expected benefit derived only from the increase in the industrial output

---

[10] Haryana's manufacturing output would be 1.5 times Bihar's, rather than the observed factor of 2.25.

per year would be around 45 billion Rupees[11]. This would constitute an increase of around 7.8% of Bihar's total product per year.

2.4 Alternative Explanations and Other Robustness Checks

The main concern with respect to the identification strategy at hand is related to the exclusion restriction. This means that if the measure of rural electrification is strongly correlated with other rural outcomes that are actually driving the industrialization process, then the results obtained in the previous section would be spurious. In this section I will explore whether the variation used to identify the effect of electrification can also be used to explain alternative channels for the observed divergence in industrial development.

2.4.1 Would the same strategy work for alternative channels?

There are many stories in the literature as to why shocks to agricultural productivity would spill over to the manufacturing sector. Johnson (1997) points out that a rapid increase in agricultural productivity contributed to the start of the Industrial Revolution, by releasing (cheap) labour for the manufacturing sector and creating the agricultural surplus needed to feed the growing population. This argument could be linked to a process of urbanization as the agricultural sector lays workers off. Mat-

---

[11] This is obtained by multiplying the estimated increase in real manufacturing output per capita times total population in Bihar in 1984 (around 74 million people). The value is in 1974 Rupees.

suyama (1992) formalizes these mechanisms in a model where both sectors compete for labour and where exogenous increases in agricultural productivity benefit the manufacturing sector. This mechanism works only in closed economies, since comparative advantages in the agricultural sector might deprive industries from labour in open economies. In that case, effects on industrialization could go either way[12]. Murphy et al. (1989a) propose a mechanism wherein industrialization happens because positive shocks to agricultural productivity boost rural income and–subsequently–the demand for manufactures. This result hinges on two conditions: non-homothetic preferences, i.e. the share of expenditure on manufactures increases with income, and a fair distribution of benefits from increases in agricultural productivity to sustain a sizeable demand[13]. There is also a likely credit mechanism at work whereby the increase in rural incomes affects saving rates in rural areas and, subsequently, credit availability. This would reduce the cost of setting up firms. Finally, the instrumental variable strategy might be picking up greater investment in schooling after the Green Revolution increased returns to education (see Foster and Rosenzweig (1996)) or greater expenditure on development projects by the state government

In short, the shock to agricultural productivity introduced by the Green Revolution might have affected industrial outcomes in different ways. In all cases, the

---

[12] India, being a very large federal country, could be thought as a collection of open economies. In that case, the mechanisms described by Matsuyama might be present.

[13] See also Voigtländer and Voth (2006). This could undermine the argument when applied to India, since it is well documented (see for example, Dhanagare (1989)) that the Green Revolution has increased the rift between poor and rich farmers and, in some cases, has not even increased poor peasants' real incomes.

instruments would be correlated with the error term in the baseline regression and the instrumental strategy would lose validity. It follows that if these alternative channels are operating, variables that determined the feasibility of HYV seeds (e.g., groundwater availability) could also be used as instruments in a similar strategy to the one I used to identify λ in equation (2.1). To this end, I run OLS and 2SLS specifications as in Table 4 and check whether the same strategy could have been used to explain divergence in industrial development through other means than electrification. In particular, I run a similar set of regressions to the ones presented in columns (2), (3), (4) and (7) in Table 4, using–instead of the measure of electrification–measures of urbanization rates, changes in the demand for manufacturing goods, and changes in funds available to invest via credit, in literacy or in public expenditure in development. I am interested in testing whether the estimates show the expected sign and similar levels of significance for the point estimates, the first stage F-tests and the over-identification tests. Table 5 summarises the results.

The main conclusion is that none of the alternative stories passes all tests. In particular, the p-values of over-identification tests are low in all cases and in many cases, results don't even show the expected signs. It follows that doing a similar exercise with different variables would have given inconclusive results at best. For example, the log of rural expenditure is never significant, does not have the expected sign for some specifications, and over-id tests are not strong. The same holds for urbanization rates, rural non agricultural credit (and other unreported measures of credit) and literacy. The log of development expenditure is the one with more consistent results, even though the over-identification tests show a p-value of less than 1% in all cases. These results increase the confidence that the variation used in the instrumental strategy works through the proposed channel and not through others. It is in that sense that the implication is not that the Green Revolution had no impact whatsoever on rural expenditure or on urbanisation rates. As a matter of fact, the opposite is true. When I look at the first stage of the 2SLS specification reported in Table 5 where I instrument mean rural expenditure with "Generated HYV," controlling for state dummies, this produces a statistically insignificant coefficient for the measure of HYV, but significant differences for the mean levels captured by state dummies. For example, the mean rural expenditure per capita in Punjab and Haryana is between 2 and 2.5 standard deviations greater than in Bihar, and the proportion of rural population is between 5 and 11 percentage points lower. The importance of the information provided in Table 5 lies in ruling out alternative explanations to electrification in the empirical strategy. As mentioned before, these results are not suggestive that these other variables are unimportant or did not reflect differences across states, but simply that the variation used to instrument for the measure of electrification cannot be used to explain variation in urbanization rates, credit availability, literacy, state expenditure in development or expenditure in rural households. The use of state fixed effects in all specifications means that I exploit within state variation, and only in the case of the electricity measure does the time-varying effect of groundwater availability have explanatory power for understanding states' divergence in industrial development.

2.4.2 Do groundwater and HYV adoption have a direct effect on industrial outcomes?

The analysis presented so far looks only at variation within states across time. A good reason for focusing on states' variation is that each state has its own electricity network, run at the state level by the State Electricity Board. An implication of the story portrayed in this paper is that the expansion of the network aims at reaching areas intensive in HYV seeds, but given the network characteristic of electricity provision, intensity in HYV is not necessary. For example, a district whose soil or climatic characteristics are not suitable for HYV seeds might still benefit from this process of electrification if surrounded by districts intensive in HYV seeds. If this is not the case, then my results on industrial output could be driven by a direct effect of HYV adoption and not by electrification. An alternative story with respect to the placement of the electricity network might be that what actually matters is groundwater availability and not HYV adoption or power availability, because rural industries use groundwater as a part of their production process. To test these alternative stories, I use district level data available from the World Bank's "Agricultural and Climate Dataset," gathering information on agricultural outcomes, geographical characteristics and input availability for 271 districts across 13 Indian states from 1956 to 1987. I matched these with census data at the district level on employment in the rural manufacturing sector for every 10 years from 1961 to 1991. The measure of groundwater availability is a dummy that equals 1 if the district has aquifers thicker than 150 metres and 0 otherwise.

To test whether rural electrification follows groundwater availability or the adoption of HYV seeds, I first run a difference-in-difference estimation for a cross section of districts in 1974, where I use a measure of rural power availability at the district level[14] as the explained variable and the interaction between a measure of groundwater availability and adoption of HYV seeds. A setback to my story would be to find that power availability was not driven by the Green Revolution, but by the thickness of the watertable. Column (1) in Table 6 shows that this alternative explanation does not hold. The coefficient on groundwater availability is negative, suggesting that districts with thick watertable that were not intensive in HYV had on average less rural power. This undermines the argument that what mattered for the placement of electricity connections was water availability exclusively. Furthermore, the coefficient on HYV is positive, showing a positive correlation between HYV adoption and power in districts without a watertable thicker than 150 metres[15]. Finally, the coefficient for the interaction term is positive, significant and greater in magnitude than the estimates for each of the levels. This shows that districts adopting HYV and with thicker watertables were those with greater rural power availability.

---

[14] Collected and published for one year only by the National Commission on Agriculture, Ministry of Agriculture and Irrigation of India.

[15] Of course, water can be obtained from shallower watertables.

Overall, total effects of HYV adoption and groundwater are positive, as expected. This result provides evidence in favour of the link set out in my empirical strategy between Green Revolution and electrification.

In Column (4), I use census data on employment in rural manufacturing to test whether groundwater has a direct effect on the manufacturing sector or a differential effect for districts with more power. I interact power availability with the measure of groundwater. The coefficient on groundwater is negative and not significant, i.e. groundwater has no positive independent effect on manufacturing outcomes. Additionally, the coefficient on power availability is positive and significant at the 1% level, while the interaction term is positive but insignificant. This result supports the idea that what matters for the increase in rural manufacturing employment is electrification, and not groundwater availability.

I use a similar procedure to test the channel going from electrification to industrial outcomes against the concern that increases in manufacturing employment are a direct result of HYV adoption. As previously mentioned, districts that have not intensively adopted HYV seeds might have benefited from accessing an electricity network that expanded to serve areas intensive in HYV. By interacting power availability with HYV adoption, the difference-in-difference estimation allows me to check whether rural manufacturing employment was linked to HYV adoption in areas with low power availability, to power availability in areas with low HYV adoption or to the interaction of both. Column (2) in Table 6 finds that only the measure of electrification is positively and significantly correlated with the proportion of the rural population working in the manufacturing sector. The sign of HYV adoption is negative, even though not significant. This suggests that districts intensive in HYV with low levels of electrification have less people working in the manufacturing sector. The interaction term is positive, but insignificant. Column (3) uses a measure of the proportion of rural workers employed in the manufacturing sector. The three coefficients have the same signs but are now all significant. Again, HYV adoption without electrification was linked to less employment in the rural manufacturing sector. The interaction is positive, meaning that districts intensive in HYV and rural power had marginally more manufacturing workers. Measured at the mean of the measure of power availability (i.e. 0.15), the overall effect of HYV remains negative. The overall effect of rural electrification in the proportion of rural workers is again positive and significant, irrespective of whether the district was intensive on HYV seeds or not. These results add confidence to the claim that rural electrification can be linked to an increase in the importance of the manufacturing sector. It also reduces concerns regarding the exclusion restriction by suggesting that HYV adoption affected outcomes in the rural manufacturing sector only by means of electrification.

2.4.3 Is the measure of groundwater a "lucky draw"? A placebo experiment.

Another concern might be that the time invariant measure of groundwater availability I use as an instrument fits the data spuriously. Even though there is enough evidence that groundwater availability was an important component of irrigation for newly introduced seeds (through the use of electrified tubewells, see Sims (1988) or Foster and Rosenzweig (2008), for example), it might still be the case that the results are driven by a lucky draw. The question that I am asking here is "What are the chances that a random measure of groundwater passes all tests?" To test this hypothesis, I run 500 Monte Carlo simulations, in each case randomly assigning a value between 0 and 1 to each state[16]. I use this value, interacted with year dummies, as an instrument for electrification, replicating the main specification in Column (4), Table 4. I collect information on the sign and significance of the estimate, on the F-test and on the over-identification test. These tests provide information on whether random allocations of the measure of groundwater can systematically predict the divergence in electrification across states, have power to predict the second stage and pass the over-identification test.

Table 7 shows that out of the 500 simulations I ran, only 8% show positive and significant results for the second stage. Of those 40 cases, only half of the regressions show a powerful first stage F-test (20 cases, i.e. 4%). Finally, only 2 regressions out of the 500 also pass the over-identification test with p-values above 90%. This simulation exercise shows that the probability of obtaining the results obtained in Column (4), Table 4 by a random draw are on the order of 0.4%. This low probability increases the confidence that the econometric exercise is producing meaningful results.

---

[16] Note that the measure of groundwater availability I use in the previous section is also between 0 and 1.

2.4.4 Other indicators of industrial development.

Table 8 shows results for other indicators of industrial development. Columns (1) and (2) report results for measures of value-added and fixed capital per capita. In both cases, the instrumented measure of electrification is positive and significant. One concern related to the expansion of the manufacturing sector might be that Green Revolution states are only expanding the food industry after the surge in food production. To check whether this is the case, I use a measure of manufacturing output in non-food industries for Columns (3) and (4), available only for the years 1980-1984. Column (3) reproduces the baseline 2SLS specification and finds a positive and significant effect of electrification on production of non-food manufactures. In Column (4) I used only cross-state variation in groundwater availability as an instrument and also found a positive and significant effect.

The mechanism through which the expansion of the electricity network can affect industrialization is a simple one: better infrastructure can lower prices and induce more consumption of manufacturing goods. An alternative story could be that of a selection effect in a setting with heterogeneous firms, where smaller and less productive firms can break even only when electricity is available. To check this, I test whether new factories open and smaller firms produce more with the expansion of the electricity network. I use information on the number of factories per capita and small manufacturing firms' indicators as surveyed by the Annual Survey of Statistics and excluded from the Census Sector[17]. In Columns (5) and (6) in Table 6, I test for the number of factories with two different sets of instruments, respectively: groundwater availability interacted with year dummies and the triple interaction, including also the measure of initial irrigation. Results show that electrification is significantly associated with more factories. F-tests and over-identification tests pass. Magnitudes are of the same order as for total output: a one standard deviation difference in electrification is associated with at least 15% more factories. Similar results are obtained in Columns (7) and (8), using a measure of value-added for small firms[18]. This provides evidence of the presence of a selection effect where smaller firms also benefit by the expansion of the electricity network.

2.4.5 Other instruments and measures of electrification.

The next step is to check whether other geographic characteristics associated with adoption of HYV seeds can be used as instruments. In Columns (1) and (2), Table 9, respectively, I use a measure of soil quality and aridness interacted with year dummies as instruments. Results remain similar in magnitude to those in Table 4, even though significance, first stage F-tests and over-id tests are not as strong. This is not surprising since electrification is associated not just with soil and climatic character-

---

[17] Small firms are those with less than 50 workers (and with more than 50 and less than 100 workers that do not use power).

[18] Unreported specifications show the results hold for other measures of the small sector, such as fixed capital.

istics compatible with HYV seeds, but mainly with the availability of groundwater that can be obtained through electrified tubewells.

In columns (3), (4) and (5) I use different indicators of electricity availability. I use the total connected load in Kw per million people (i.e., maximum power available), the number of factories connected to the network and the number of factories using low and medium voltage only, respectively. In all three cases, the pattern of results hold as do the power of tests and magnitudes: a change in one standard deviation in the electricity indicator explains between 10% and 15% of the difference in manufacturing output.

2.4.6 Groundwater at the district level.

In this section, I check the robustness of the first stage using information at the district level to add confidence to the use of the Green Revolution as a natural experiment. In particular, I run a similar set of regressions as in Section 3.1 where the level of HYV adoption is explained by the interaction of groundwater availability and year dummies. Exploiting within district variation using district fixed effects, I use the following specification:

$$HYV_{dst} = \beta_{0ds} + \beta_{1t} + \sum_{k=1966}^{1984} \gamma_t (S_{ds} * T_k) + \delta_t \mathbf{X}_{dst} + \epsilon_{st} \qquad (2.8)$$

As in the analysis for the state level, S_{ds} is a measure of groundwater availability (a dummy that equals 1 if the district has aquifers thicker than 150 metres and 0 otherwise) I run (2.8) with and without district controls X_{dst}[19]. A third specification is introduced, where I replace the explained variable in equation (2.8) by a measure of pumpsets available per capita. As mentioned before, there is no information on availability of electricity at the district level, but the anecdotal evidence suggests a shift from diesel to electrified pumpsets after the start of the Green Revolution (see for example, Bharadway (1990) or McGuirk and Mundlak (1991)).

Figure 4 plots the estimated γ_t for the three specifications. As shown for the state level regressions, districts with greater groundwater abundance are associated with a greater and increasing proportion of adoption of HYV seeds and more irrigation pumpsets. Estimates for the interaction terms are positive, significantly different from 0 only after the Green Revolution and increasing, meaning that there is a positive marginal effect for districts with more groundwater. This set of results is consistent with the findings at the state level and provides more evidence validating the choice of instruments.

2.5 Conclusion

The role that infrastructure provision plays in improving economic and social outcomes in developing countries is part of the conventional wisdom among academics, policy makers and the population at large. Despite this consensus, the academic lit-

---

[19] Time varying controls at the district level include literacy, population density, proportion of rural population, bullocks and tractors per hectare. District fixed effects are also included and standard errors are clustered at the district level.

erature has not provided a substantial amount of empirical investigations aimed at understanding and quantifying the effects of infrastructure provision. The challenge of this paper is to quantify the effect of electrification in industrial output by looking at a panel of Indian states from 1965-1984. To do so, I address the endogenous placement of infrastructure by looking at the time-varying effect of time-invariant geographic characteristics on electrification after the start of the Green Revolution in India. The need for timely irrigation that could be cheaply supplied by pumping water from the water table using electrified pumpsets generated a growing demand for electricity and subsequent expansion of the electricity grid in some states. The magnitude of the results suggests that expanding the electricity network–including rural areas–should be considered seriously as a policy option for promoting industrial development. Two-stage least squares estimations show that between 10 and 15% of the differential level in manufacturing output can be explained by the extension of electricity provision across Indian states. The expansion of the network also helped the entry and performance of smaller firms. These results are robust to a series of tests regarding the power and validity of the instruments. The findings of this chapter support the idea that the lack of infrastructure is a major constraint on economic activity in developing countries, including countries with large rural populations.

2.A Appendix A: Variables and Sources

Electricity indicators: All electricity indicators are taken from "Public Electricity Supply - All India Statistics," published annually by the Central Electricity Authority between 1950 and 1985. "Agricultural Connections" is the number of agricultural units connected to the electricity network per 1000 people. "Connected Load" is the maximum consumption available per consumer, measured in KW per million people. "Industrial Connections" is the quantity of industrial users connected to the network per 1000 people. "Low voltage" refers to industrial users with connection with voltage below 33 kV.

Manufacturing outcomes: Data for industrial development and performance indicators are taken from different publications from the Department of Statistics, Ministry of Planning, Government of India. Manufacturing output comes from "Estimates of State Domestic Product" and measures of the stock of fixed capital, value added, investment and number of factories come from "Annual Survey of Industries" and were gathered by the EOPP Indian States Data, STICERD, LSE.

Agricultural variables: "HYV" is the net proportion of area cropped with HYV seeds. "Groundwater" is the proportion of districts per state with aquifers thicker than 150 mts."Topsoil" is the proportion of districts per state with topsoil deeper than 3 mts. "Rainfall" is the monthly average rainfall measured for 1958-60. Data on soil characteristics and cropped area under HYV seeds come from the "India Agriculture and Climate Dataset", compiled by Sanghi, Kumar, McKinsey, Jr. (1998).

Controls: Population density, proportion of literate population, and proportion of rural population, mean rural expenditure and price indices used to deflate variables are taken from "A Database on Poverty and Growth in India," prepared by Ozler, Datt and Ravallion (1996). Data on credit come from "Statistical Tables relating to Banks in India," Reserve Bank of India. Data on expenditure on health, education and development come from a Ministry of Finance, Government of India publication: "Public Finance Statistics." Political outcomes per state were compiled by Butler, Lahiri and Roy in their "India Decides: Elections 1952-1995." All data are available from the EOPP Indian States Data, STICERD, LSE.

District level data: The measure of power availability in rural areas is in horses per hectare and comes from the Ministry of Agriculture and Irrigation of India (1976), Report of the National Commission on Agriculture, Volume X. Measures of rural manufacturing outcomes are taken from the Indian Census for 1961, 1971, 1981 and 1991.

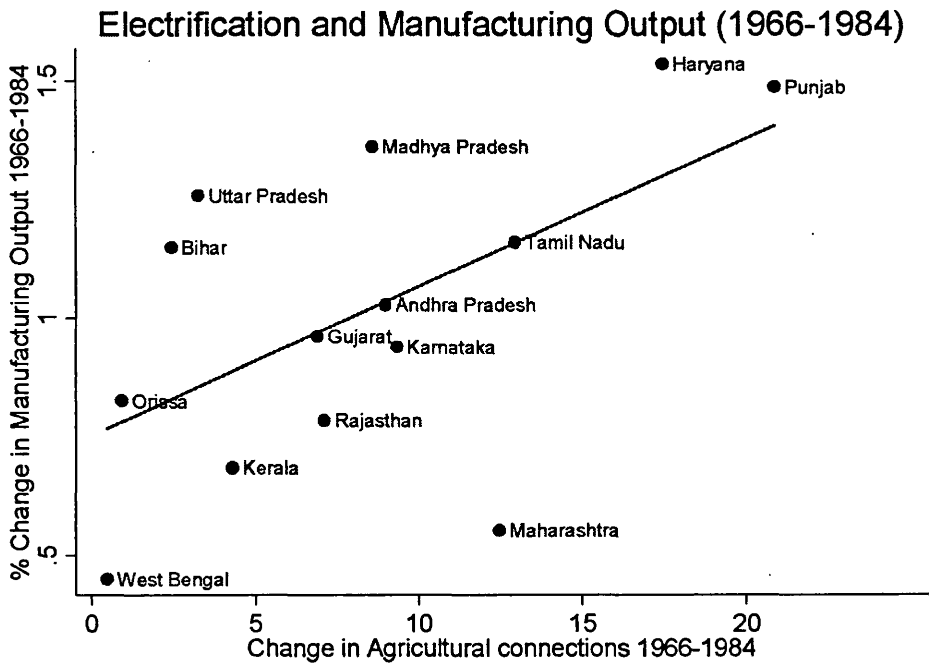

Figure 1: Changes in Rural Electrification (measured by agricultural connections per 1000 people) and Per Capita Manufacturing Output Growth Between 1966 and 1984.

Note that rates of HYV adoption were highest in Punjab, Haryana and Tamil Nadu.

Table 1: Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
Agricultural Connections (per 1000 people)	4.04	5.18	0	23.24
HYV adoption (%, from 1966)	0.23	0.17	0	0.92
Log Manufacturing Output per capita	7.00	0.55	5.98	8.30
Geographic Variables				
Abundant Groundwater (% of Districts)	0.11	0.19	0	0.6
Top Soil (% of Districts)	0.48	0.35	0	1
Average Rainfall (mm, annual)	292.5	137.4	66.3	557.7
Initial Conditions (year 1960)				
Irrigated Area (%)	0.19	0.14	0.06	0.54
Agricultural Connections (per 1000 people)	0.46	0.89	0	3.54
Controls				
Population Density	0.23	0.15	0.02	0.68
Literate Population (%)	0.37	0.13	0.13	0.82
Log Total Credit per capita	4.70	0.98	2.49	6.76
Congress Party Vote (%)	39.72	10.78	8.8	56.3
Log Development Expenditure per capita	-6.16	0.76	-8.05	-4.62

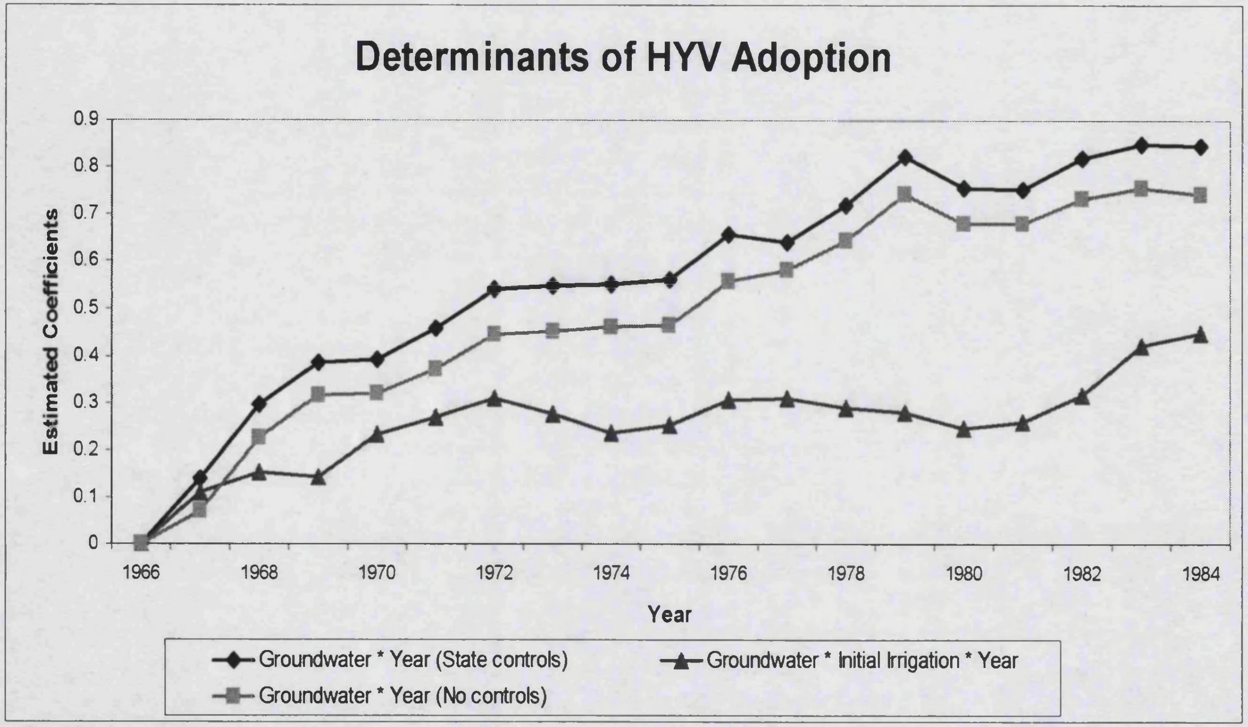

Figure 2: Drivers of HYV. Plot of estimated coefficients γ_t: Diffusion of HYV as a function of groundwater availability (with and without state controls) and its interaction with initial irrigation.

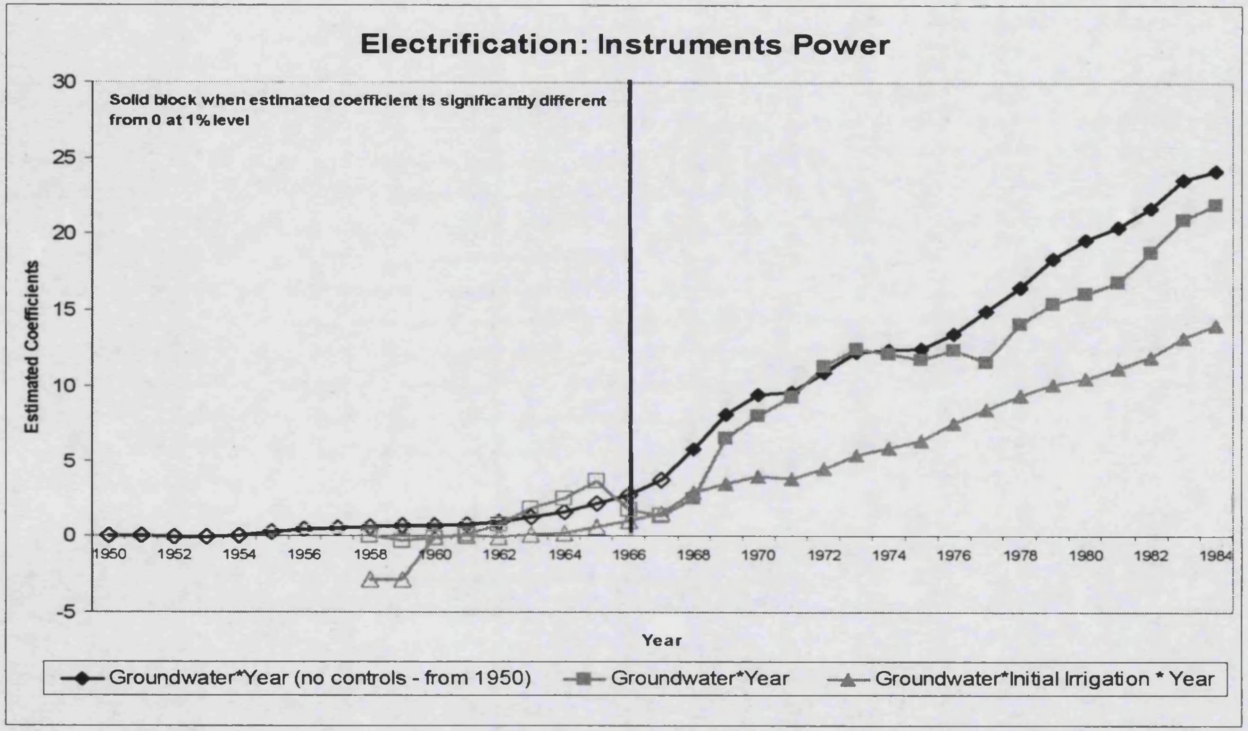

Figure 3: Instruments Power. Plot of estimated coefficients γ_t for agricultural electricity connections as a function of groundwater availability (with and without state controls) interacted with year dummies. Also includes results when interacted with initial irrigation. The vertical line denotes the start of the Green Revolution

Table 2: Rural Electrification and Manufacturing Output Time Trends

	(1)	(2)	(3)	(4)	(5)	(6)
	Agricultural Connections				Log Manufacturing Output	
Data from	1950	1950	1958	1958	1960	1960
Year-1950	-0.018 (0.012)	-0.01 (0.01)				
Year-1958			0.05 (.05)	-0.04 (0.04)	0.06 (0.09)	0.014 (0.007)*
Year-1966	0.46 (0.10)***	0.34 (0.09)***	0.47 (0.10)***	0.47 (0.09)***	-0.001 (0.016)	-0.006 (0.013)
Groundwater * (Year-1950)		-0.07 (0.09)				
Groundwater * (Year-1958)				0.39 (0.19)*		-0.09 (0.02)***
Groundwater * (Year-1966)		1.28 (0.20)***		1.09 (0.13)***		0.03 (0.01)**
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	No	No	No	No	No	No
State Controls	No	No	Yes	Yes	Yes	Yes
Observations	525	525	405	405	375	375
Adjusted R-sq	0.70	0.81	0.90	0.90	0.93	0.93

Robust standard errors in parentheses, clustered at the state level. * significant at 10%; ** significant at 5%; *** significant at 1%. "Agricultural Connections" is the number of agricultural units connected to the electricity network per 1000 people (data: 1950-1984). "Log Manufacturing Output" is the log of real per capita manufacturing output (data:1960-1984). "Year-T" is equal to the number of years after T (until 1966, unless T=1966). "Groundwater" is the proportion of districts per state with aquifers thicker than 150 mts. "Topsoil" is the proportion of districts per state with topsoil thicker than 3 mts. Controls are: population density, literacy, log of total credit pc, log of development expenditure pc, Congress Party vote, party of the Chief Minister.

Table 3: Rural Electrification and Manufacturing Output After the Start of the Green Revolution for States with More Available Groundwater

	(1)	(2)	(3)	(4)	(5)	(6)
	Agricultural Connections				Log Manufacturing Output	
Data from	1950	1950	1958	1958	1960	1960
Groundwater * Post 66	13.06 (2.41)***	13.06 (2.30)***	11.12 (1.53)***	11.75 (4.30)**	0.77 (0.19)***	0.83 (0.25)***
Groundwater		0.55 (0.21)*		0.48 (4.26)		-0.80 (0.36)**
Post 66		3.55 (1.01)***		0.13 (1.05)		-0.13 (0.08)
State Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	Yes	No	Yes	No	Yes	No
State Controls	No	No	Yes	Yes	Yes	Yes
Observations	525	525	405	405	375	375
Adjusted R-sq	0.77	0.41	0.90	0.67	0.93	0.80

Robust standard errors in parentheses, clustered at the state level. * significant at 10%; ** significant at 5%; *** significant at 1%. "Agricultural Connections" is the number of agricultural units connected to the electricity network per 1000 people (data: 1950-1984). "Log Manufacturing Output" is the log of real per capita manufacturing output (data:1960-1984). "Post 66" is a dummy equal to 1 for all years from 1966. "Groundwater" is the proportion of districts per state with aquifers thicker than 150 mts. "Topsoil" is the proportion of districts per state with topsoil thicker than 3 mts. Controls are: population density, literacy, log of total credit pc, log of development expenditure pc, Congress Party vote, party of the Chief Minister.

Table 4: OLS and 2SLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log Manufacturing Output						
	OLS		2SLS				
Agricultural Connections	0.029 (0.010)**	0.026 (0.007)***	0.037 (0.006)***	0.021 (0.007)***	0.021 (0.005)***	0.018 (0.008)**	0.028 (0.006)***
Instruments	n/a	n/a	Groundwater * Year	Groundwater * Year	Groundwater * Initial Irrigation * Year	HYV	Generated HYV
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes	Yes	Yes	Yes
Observations	300	300	300	300	300	300	300
Over-id P-value	n/a	n/a	0.99	0.97	0.94	n/a	n/a
First Stage F-Test P-value	n/a	n/a	<1%	<1%	<1%	<1%	<1%

Robust standard errors in parentheses, clustered at the state level. * significant at 10%; ** significant at 5%; *** significant at 1%. "Agricultural Connections" is the number of agricultural units connected to the electricity network per 1000 people. "Log Manufacturing Output" is the log of real per capita manufacturing output. "Groundwater" is the proportion of districts per state with aquifers thicker than 150 mts. "Initial Irrigation" is the proportion of cropped area irrigated in 1959. "HYV" is the net proportion of area cropped with HYV seeds. "Generated HYV" is the predicted value of HYV on Groundwater * Year, used as a generated instrument. Controls are: population density, literacy, log of total credit pc, log of development expenditure pc, Congress Party vote. The First Stage F-Test is a test of the power of the instruments, sustained when null is rejected (low P-value). Over-identification tests instruments validity, where not rejecting the null (high P-value) suggests that instruments are correctly excluded and uncorrelated with the error. All data are from 1965 to 1984.

Table 5: Test for Alternative Explanations

Specification	Test	Agricultural Connections	Log Mean Rural Expenditure	Proportion of Urban Population	Rural Non Agricultural Credit	Literacy	Log Development Expenditure
OLS	Significance (at 1% level)	*					
	Expected sign	*	*	*			*
2SLS (no state controls) using Groundwater * Year	Significance (at 1% level)	*			*	*	*
	Expected sign	*			*	*	*
	F-test (p-value <1%)	*	*				*
	Over-id (p-value > 90%)	*				*	
2SLS (with state controls) using Groundwater * Year	Significance (at 1% level)	*		*			*
	Expected sign	*		*	*	*	*
	F-test (p-value <1%)	*	*	*			*
	Over-id (p-value > 90%)	*					
2SLS (with state controls) using Generated HYV	Significance (at 1% level)	*					
	Expected sign	*	*			*	*
	F-test (p-value <1%)	*				*	

" * " when the specification in the first column passed the test in the second Agricultural Connections is the number of agricultural units connected to the electricity network per 1000 people. "Log mean rural expenditure" is a measure of real expenditure per capita in rural households, "Rural non agricultural credit" a measure of real per capita credit in rural areas not used in agriculture and "Log development expenditure" is the log of real per capita expenditure by the state government in health, education and other development projects. "Literacy" and "Urban population" are proportions of total population. The First Stage F-Test is a test of the power of the instruments, sustained when null is rejected (low P-value). Over-identification tests instruments validity, where not rejecting the null (high P-value) suggests that instruments are correctly excluded and uncorrelated with the error.

Table 6: District Level Data - Do Groundwater and HYV Have a Direct Effect on Industrial Outcomes?

	(1)	(2)	(3)	(4)
	Power Availability	Rural Manufacturing Workers		
		% Rural Population	% Rural Workers	% Rural Population
Groundwater	-0.25	0.06		-0.06
	(.066)***	(0.09)		(0.11)
Groundwater * HYV	1.27			
	(0.26)***			
HYV	0.52	-0.004	-0.013	
	(0.13)***	(0.003)	(0.007)*	
Power Availability		0.0024	0.005	0.003
		(0.001)***	(0.002)**	(0.001)***
Power Availability * HYV		0.003	0.015	
		(0.002)	(0.007)**	
Power Availability * Groundwater				0.002
				(0.0015)
Observations	271	763	763	763
Adjusted R-sq	0.40	0.58	0.62	0.58

Robust standard errors in parentheses, clustered at the state level. * significant at 10%; ** significant at 5%; * significant at 1%. "Power availability" is a measure, in horses per hectare, of power in rural areas. "Groundwater" is a dummy equal to 1 if the district has aquifers thicker than 150 mts. "HYV" is the net proportion of area cropped with HYV seeds. Controls are: population density, literacy, roads and bullock per hectare. Column (1) uses data for 1974. Columns (2) and (3) use Census data for 1961, 1971, 1981 and 1991.**

Table 7: A Placebo Experiment

	Simulation	Cases (% out of 500 repetitions)
(1)	Significance (at 1% level) and Expected Sign	40 (8%)
(2)	(1) + F-test (p-value <1%)	20 (4%)
(3)	(2) + Over-identification test (p-value > 90%)	2 (0.4%)

I run 500 Monte Carlo simulations where in each case I randomly assign a value between 0 and 1 to each state. I use this value, interacted with year dummies, as an instrument for electrification, replicating the main specification in Column (4), Table 4. Proportions shown in parentheses.

Table 8: Robustness Checks - Alternative Industrial Outcomes, No-Food Industries, Factories and Small Sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Value Added	Fixed Capital	Log Manufacturing Output (No food industries)		Factories		Value Added- Small Sector	Value Added - Small Sector
	2SLS				2SLS			
Agricultural Connections	0.028 (0.006)***	0.09 (0.02)***	0.08 (0.03)**	0.03 (0.01)**	0.003 (0.001)**	0.005 (0.001)***	0.014 (0.004)***	0.014 (0.002)***
Instruments	Groundwater * Year	Groundwater * Year	Groundwater * Year	Groundwater	Groundwater * Year	Groundwater * Initial Irrigation* Year	Groundwater * Year	Groundwater * Initial Irrigation* Year
State Fixed Effects	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	300	300	75	75	300	300	300	300
Over-id P-value	0.23	0.97	0.97	n/a	0.80	0.96	0.69	0.92
First Stage F-Test P-value	<1%	<1%	2%	<1%	<1%	<1%	<1%	<1%

Robust standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%. "Fixed Capital" and "Value Added" are per capita for the manufacturing sector "Log Manufacturing Output ("No food industries" is available only for 1980-1984 and excludes all manufacturing output in food processing industries "Agricultural Connections" is the number of agricultural units connected to the electricity network per 1000 people "Groundwater" is the proportion of districts per state with aquifers thicker than 150 mts All outcome variables are in real terms and per capita "Small sector" are factories with less than 50 workers using power or more than 50 and less than 100 workers without power. Controls are population density, literacy, log of total credit pc, log of development expenditure pc, Congress Party vote, party of the Chief Minister. The First Stage F-Test is a test of the power of the instruments, sustained when null is rejected (low P-value) Over-identification tests instruments validity, where not rejecting the null (high P-value) suggests that instruments are correctly excluded and uncorrelated with the error.

Table 9: Robustness Checks - Alternative Instruments and Measures of Electrification

	(1)	(2)	(3)	(4)	(5)
	Log Manufacturing Output				
	2SLS				
Agricultural Connections	0.025 (0.013)**	0.035 (0.017)**			
Connected Load			0.002 (0.0006)***		
Industrial Connections				0.18 (0.07)***	
Industrial Connections - Low Voltage					0.23 (0.07)***
Instruments	Topsoil * Year	Rainfall * Year	Groundwater * Year	Groundwater * Year	Groundwater * Year
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes
Observations	300	300	300	300	300
Over-id P-value	0.36	0.71	0.78	0.89	0.89
First Stage F-Test P-value	4%	4%	<1%	<1%	<1%

Robust standard errors in parentheses, clustered at the state level. * significant at 10%; ** at 5%; *** at 1%. "Agricultural Connections" is the number of agricultural units connected to the electricity network per 1000 people. "Log Manufacturing Output" is the log of real per capita manufacturing output. "Groundwater" is the proportion of districts per state with aquifers thicker than 150 mts."Topsoil" is the proportion of districts per state with topsoil depeer than 3 mts. "Rainfall" is monthly average rainfall measured for 1958-60. "Connected Load" is measured in KW per million people. "Industrial Connections" is measured per 1000 people. "Low voltage" is less than 33 kV. Controls are: population density, literacy, log of total credit pc, log of development expenditure pc, Congress Party vote, party of the Chief Minister. The First Stage F-Test tests instruments power, sustained when null is rejected (low P-value). Over-identification tests instruments validity, where not rejecting the null (high P-value) suggests that instruments are correctly excluded and uncorrelated with the error. All data are from 1965-1984.

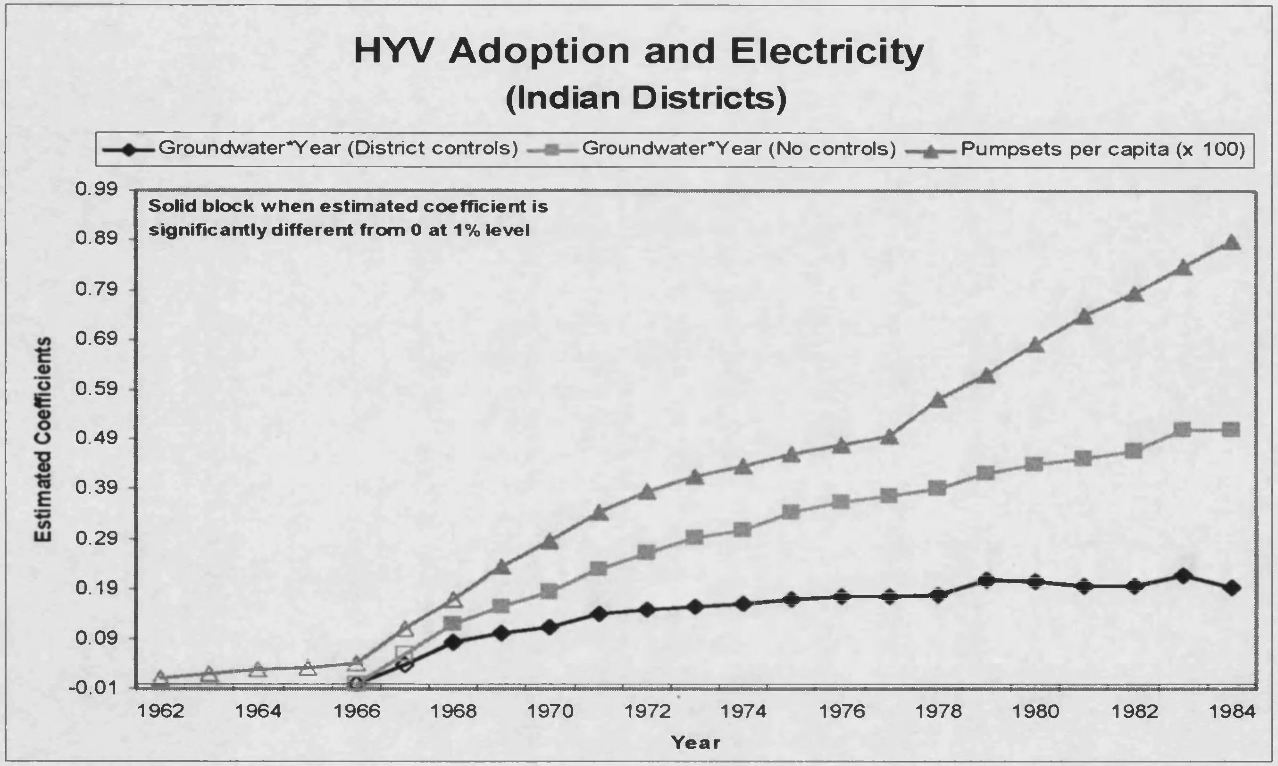

Figure 4: Drivers of HYV and Pumpsets at the District Level. Plot of estimated coefficients γ_t: Diffusion of HYV and irrigation pumpsets as a function of groundwater availability.

Chapter 3

Coping with Bad Infrastructure: Adoption of Captive Power Generators Among Indian Firms

3.1 Introduction

The lack of adequate infrastructure is one of the main challenges that firms in developing countries face on a daily basis. Infrastructure costs for firms in the developing world can be up to four times higher than they are in OECD countries, making infrastructure one of the main concerns for entrepreneurs (Gonzalez et al. (2007)). In many cases, bad infrastructure stems from countries' financial inability to invest in infrastructure (e.g. because of the stress or inefficiency of the tax system), but in others, it is the consequence of direct institutional failures such as weak governance, lack of capacity or poor regulation (Laffont (2005), World Bank (2006)). However, problems associated with high tariffs and poor quality in infrastructure can be mitigated if firms have access to outside options for the public provision of infrastructure goods. In the case of electricity, in regions like Sub-Saharan Africa and South Asia–where only 1 in 4 people and 2 in 5, respectively, have access to electricity from the network–firms turn to the installation of captive power generators. For example, in a survey of manufacturing firms in India in 2002, almost 30% of firms reported that the state of the electricity network was a major or a very severe obstacle for business and 64% owned an electricity generator. In Pakistan, firms reported 40% and 42% to the same questions, respectively. In Sri Lanka and Bangladesh, 41% and 71%, respectively, reported that electricity was a major or very severe obstacle and, in both cases, the adoption rate exceeded 70% (Asian Development Bank (2002)). Arnold et al.(2008) show data from a similar survey for firms in Sub-Saharan Africa: in Kenya, 71% of firms in the sample owned a generator; in Tanzania, 55% and almost 40% in Uganda and Zambia. In all countries, the average score for electricity as an obstacle was between "moderate" and "major".

This set of stylized facts suggests that, at least in South Asia and Sub-Saharan Africa, the inadequate provision of electricity hinders the development of the manufacturing sector and that some firms can find alternative arrangements to deal with the problem, but some cannot. It follows that if firms that adopt captive power coexist with firms that do not, then this must somehow affect the configuration of an industry (i.e. who survives, distribution of profits, etc.) In this chapter I investigate these questions by analysing the underlying determinants of captive power adoption by firms and their effects on industry outcomes. To that end, I develop a model that links the regulatory process in the upstream (e.g. electricity provision) to firms' behaviour in the downstream and the resulting industry equilibrium. I subsequently use data from firms in India to analyse their decision to set up a captive power generator to deal with inadequate provision and pricing schemes in the electricity sector. I am interested in identifying not just which firms adopt a generator to back up their network provision but also in looking at how this phenomenon affects industry-wide outcomes. Furthermore, I provide some insight into the nature of different regulatory outcomes, by using the institutional features of the electricity sector in India, where each state controls the supply and decides on the pricing schedule.

Pricing schemes in Indian states are generally biased towards agricultural and domestic consumers, who enjoy below cost or even free energy, while the industrial sector has to contend with high tariffs and unreliable supply. These regulatory practices have a political economy component (e.g. an agricultural voting base gives power to pro-agricultural political parties) and a socioeconomic component (e.g. keeping agricultural inputs cheap so agricultural and food prices remain low). They may also generate a vicious circle in which the agricultural sector, because of its importance, controls the political lever lobbying for subsidies that affect a manufacturing sector which remains politically and economically restrained. To account for this asymmetry in lobbying power, I build on the analysis of the regulatory process in the baseline model developed by Laffont and Tirole (1993, Ch. 2). The authors show that under the standard assumption that the regulator collects the money and makes transfers to the producer, optimal pricing is above marginal and below monopolistic price, because transfers to the supplier can only be done at a positive cost. The key idea is that because transferring funds from taxpayers is socially costly, society is better off by allowing the regulated monopoly to charge above marginal cost, thus reducing the need for transfers. In this setting, I complement this model by introducing a measure of the power of lobbies in the regulator's objective function: the price set will depend on the relative importance the regulator gives to the welfare of the industrial sector, the consumer of the monopolistically supplied good. Intuitively, with zero transfer costs, if the regulator cares fully about consumers' well being, then the weight given to the consumer surplus in a utilitarian social welfare function would be highest and the price should be set equal to the marginal cost. Conversely, if the regulator does not care about consumers, then the price would be set at the unconstrained monopolistic price. Intermediate values of the pro-consumer preference would imply a price somewhere in between.

I analyse how regulatory outcomes affect the behaviour of firms that have access to an electricity-cost-reducing device (such as captive power generators) in a setting where consumers have a taste for variety and firms are heterogeneous, as in Melitz (2003) and Bernard et al.(2006). I am interested in characterising firms that adopt this technology and how changes in the environment (such as trade protection and licensing schemes) affect industry and firm outcomes. In particular, the model shows that adoption of captive power follows higher electricity prices and lower quality and is more prevalent in environments that are shielded from competition. This provides an interesting finding that links infrastructure regulation with the process of market deregulation. Since it is easier for firms to recoup the cost of setting up a captive generator in a more protected environment, the outside option (e.g. a captive generator) becomes less feasible when the sector is subject to more competition.

From the model, I derive a set of predictions that are tested using a repeated cross-section of manufacturing factories in India for the years 1990, 1994 and 1997. The data are merged with information on regulatory outcomes at the state level (such as prices for industrial consumers and indicators of quality) and at the industry level (such as protection from trade or barriers to entry, i.e. licenses). By looking at data from State Electricity Boards (SEB) I find that industries have to pay relatively more for their electricity in states where farmers use an electricity-intensive technology, providing evidence for the political economy mechanism of the regulatory process described above. Other results confirm that the mechanisms related to firm and industry outcomes described in the model are at work, such as the link between industry regulation (e.g. licensing schemes and trade tariffs), and firms' adoption, survival and productivity. I also find evidence that the regulatory process has heterogeneous effects along dimensions such as location, with rural and urban firms reacting more to high electricity tariffs and metropolitan firms reacting to low quality and reliability in the electricity provision. Overall, results suggest that an industry's equilibrium after a process of opening to foreign competition and deregulation varies according to quality and pricing policies in the provision of infrastructure goods.

This investigation adds to the literature on the impact of regulation of the industrial sector in developing countries by bringing together the regulatory process in the provision of an infrastructure good and the economic environment that determines firm's behaviour and industries' performance. Others have looked at price formation in input markets and industrial outcomes. Besley and Burgess (2004), for example, analyse the drivers of labour regulation in Indian states and the effects of pro-labour regulation on investment and productivity. Aghion et al. (2007) find that pro-labour regulation has a negative effect on industrial outcomes when sectors are delicensed. From an infrastructure perspective, Dollar et al. (2003) analyse the "investment climate" by using indicators of performance of telecoms, electricity and banking sectors as perceived by firms and find that firms are affected more by inadequate or badly regulated infrastructure than by higher corruption or poor governance. In particular, they find that the performance of the electricity sector (as measured by power outages) takes the higher toll in terms of productivity and profitability. Arnold et al. (2008) show that total factor productivity in manufacturing industries is positively correlated with generator ownership and negatively correlated with power outages for eight countries in Sub-Saharan Africa. In a similar vein, but with different emphasis, and concentrating specifically on regulatory issues in developing countries, a recent strand in the regulation literature investigates how the economic environment affects the quality of public utilities regulation, underlining corruption and enforcement problems (see for ex, Guasch et al. (2003) and Laffont (2005)).

The remainder of this chapter is organised as follows. The next section presents the model and derives a set of testable predictions in the context of Indian firms. Section 3 discusses the data and sources of variation and Section 4 presents the results. Section 5 concludes.

3.2 The Model

The model is composed of two parts. In the first one, I investigate the captive power adoption decision for heterogeneous firms and the subsequent industry equilibrium for a given level of quality-adjusted price of electricity. In the second part, I model the regulatory process to determine the drivers of electricity pricing and how they affect firms' decisions and the industry equilibrium.

3.2.1 Set up: downstream

Demand

All consumers have similar Cobb-Douglas preferences over two types of goods, agricultural and manufactured goods denoted by A and M, respectively. There is a continuum range of varieties V_i of manufacturing goods, over which consumers have preferences defined by a constant elasticity of substitution (CES) utility function. The taste for variety of goods is denoted by ρ, such that $0 < \rho < 1$ and the elasticity of substitution is defined as $\sigma = 1/(1-\rho) > 1$. Consumers' utility function is

$$U = M^{\alpha} A^{1-\alpha} = [\int_{v \in V} m(v)^{\rho} dv]^{\frac{\alpha}{\rho}} A^{1-\alpha} \tag{3.1}$$

where $m(v)$ is the consumption of each variety of manufactured goods. Following Dixit-Stiglitz (1977) and Fujita et al. (1999, Chapter 2), the demand for a manufacture j is

$$m(j) = \alpha Y \frac{p(j)^{-\sigma}}{H^{1-\sigma}} \tag{3.2}$$

and the demand for the agricultural good is

$$A = (1-\alpha)\frac{Y}{p_a}$$

where Y is a given level of income, $p(j)$ and p_a are the prices of the good j and the agricultural good respectively and H is the aggregate price of manufactures such that

$$H = [\int_{v \in V} p(v)^{1-\sigma} dv]^{\frac{1}{1-\sigma}} \tag{3.3}$$

Firms

There is a continuum of heterogeneous firms, each producing a different variety using labour and electricity. For simplicity, in the short run, firms are assumed to pay a fixed cost F in labour to produce l units of output and can produce in excess of l by using electricity with different degrees of success. In particular, heterogeneity is captured by a positive productivity factor φ with uniform density function $d(\varphi)$ on $(0, \Phi)$ and cumulative distribution $D(\varphi) = \varphi/\Phi$. Demand for electricity is then a linear function of firm's output (q_f) of the form $e = \dfrac{q_f - l}{\varphi}$ if $q_f > l$ and 0 otherwise. I assume the demand of a variety represented in equation (3.2) is such that $m(j) > l$ and redefine $q = q_f - l$. It follows that total and marginal costs for variety j increase with electricity price P_e and decrease with productivity φ and can be expressed as

$$TC(\varphi_j) = F + \frac{P_e}{\varphi_j} q \tag{3.4}$$

The combination of (3.2) and (3.4) gives a profit function with a variable net revenue of $(p - \frac{P_e}{\varphi_j})\alpha Y \frac{p(j)^{-\sigma}}{H^{1-\sigma}}$. The first order condition for the profit-maximising price is $(1-\sigma) + \frac{\sigma}{p}\frac{P_e}{\varphi_j} = 0$. Replacing $\sigma = 1/(1-\rho)$ and solving for p, produces an optimal price equal to the marginal cost indexed by the productivity level and a fixed mark up $1/\rho$. This price captures the local market power producers have, derived from consumers' taste for variety ρ

$$p(\varphi_j) = \frac{P_e}{\varphi_j \rho} \tag{3.5}$$

and a profit for firm j

$$\pi(\varphi_j) = \alpha(1-\rho)Y[\frac{\varphi_j \rho H}{P_e}]^{\sigma-1} - F \tag{3.6}$$

Note that profits increase when consumers spend more on many manufactures, driven by preferences (α) or income (Y), for example. Given these parameters, firms' idiosyncratic productivity (φ) determines whether firms can cover their fixed costs and survive in the market. That means that less productive firms will not be able to remain active if productivity is below a value φ^*, such that $\pi(\varphi^*) = 0$. Average productivity and the subsequent aggregate price and market equilibrium will be determined by the pool of existing firms in the market and not by the pool of potential entrants. As a consequence, I need to redefine the distribution of productivity to account for the low productivity firms that will not remain active. The resulting conditional distribution of productivity is $g(\varphi) = \frac{d(\varphi)}{1 - D(\varphi^*)} = \frac{1}{\Phi - \varphi^*}$ if $\Phi > \varphi^*$ and 0 otherwise. To rule out the latter, I assume that there is a taste for variety large enough to allow a cut-off level below the maximum, i.e. $\Phi\rho > \varphi^*$ or $\pi(\varphi = \Phi\rho) > 0$. The aggregate price is $H(\varphi^*) = [\int_{\varphi^*}^{\Phi} p(\varphi)^{\sigma-1} g(\varphi) d\varphi]^{\frac{1}{1-\sigma}}$. Also note that higher marginal cost will increase firms' profit per unit sold but, because consumers' expenditure in manufacturing goods is fixed, it will reduce demand for each variety. The role of the aggregate price and the electricity price will be explored further below.

Market equilibrium in a closed economy

In a closed economy, only domestic firms with a similar technology compete. As firms in an industry are assumed to differ only in their productivity, changes affecting the marginal cost of one firm will affect symmetrically the whole industry, meaning that not just the final price of a firm will increase with a higher marginal cost, but also the aggregate price level H. Note that by combining (3.5) for all firms and (3.3), H can be rewritten as

$$H^c = p(\tilde{\varphi}) M^{\frac{1}{1-\sigma}} = \frac{P_e}{\rho \tilde{\varphi}} M^{\frac{1}{1-\sigma}} \tag{3.7}$$

where $p(\tilde{\varphi})$ is the price of the average firm, as $\tilde{\varphi}(\varphi^*) = [\int_{\varphi^*}^{\Phi} \varphi^{\sigma-1} g(\varphi) d\varphi]^{\frac{1}{\sigma-1}}$ is the weighted average productivity, and M is the mass of surviving firms in the market. By combining equation (3.7) with (3.6), profits can be expressed as

$$\pi^c(\varphi_j) = \alpha(1 - \rho) \frac{Y}{M} [\frac{\varphi_j}{\tilde{\varphi}}]^{\sigma-1} - F \tag{3.8}$$

which is independent of the price of electricity. The fact that equilibrium profits are independent of changes in the marginal cost reproduces a result found in the classical microeconomic literature. When final goods' markets are assumed to be perfectly competitive, the supply of final goods depends on an input's price via its demand. An increase in the factor's price would simply shift upwards the supply curve. This means that changes in variables increasing firms' marginal cost would simply induce a new market equilibrium with a higher price and lower quantity traded in the final good's market. In brief, in a closed economy, changes in the marginal cost that affect all firms will not affect the cut-off productivity level φ^*, i.e. $\frac{\partial \varphi^*(P_e)}{\partial P_e} = 0$.

Following Melitz (2003) and Bernard et al. (2006), market equilibrium will be characterized by a Zero Cut-off Profit Condition (ZCP) where the average profit can be written

$$\tilde{\pi} = \pi(\tilde{\varphi}) = [\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*}]^{\sigma-1}[\alpha(1-\rho)Y[\frac{\varphi_j^* \rho H}{P_e}]^{\sigma-1}] - F = Fk(\varphi^*) \qquad (3.9)$$

where $k(\varphi^*) = [\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*}]^{\sigma-1} - 1$ is a monotonically decreasing function of φ^*, from infinity to zero under density distribution $d(\varphi)$[20].

The Free Entry Condition (FEC) implies that the expected net value of entry for the average firm (v_e)–considering the discounted (by a factor δ) probability of successful entry after paying entry fee f_e–should be equal to 0: $v_e = \frac{[1-D(\varphi^*)]}{\delta}\tilde{\pi} -$

---

[20] Under the uniform distribution $[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*}]^{\sigma-1} = \frac{\Phi^{\sigma-1} - \varphi^{*\sigma-1}}{\varphi^{*\sigma-1}(\Phi - \varphi^*)}$ that is decreasing in φ^* under the assumption that $\Phi\rho > \varphi^*$.

$f_e = 0$, or

$$\tilde{\pi} = \frac{f_e \delta}{[1 - D(\varphi^*)]} \tag{3.10}$$

which is monotonically increasing in φ^*. This is because the lower the probability of succeeding, the higher the average profit has to be. The cut-off equilibrium level φ^* will satisfy simultaneously both ZCP and FEC: as $k(\varphi^*)[1 - D(\varphi^*)] = f_e\delta/F$. The existence of equilibrium is ensured by the fact that k decreases from infinity to zero and $[1 - D(\varphi^*)]$ decreases from 1 to 0.

3.2.2 Adoption of captive power generators

The following part of the model draws from anecdotal evidence on the adoption of captive power generators by Indian firms. In particular, this section introduces the possibility that firms set up a technology which allows them to reduce their marginal cost by reducing the price P_e they pay for their electricity of a determined quality. In this setting, an excessive price or reduced quality in the provision of electricity could increase the costs of industrial production and affect the competitiveness of firms when some firms have an outside option such as a captive generator and others do not. It is well documented (see for example TERI (1999) and Tongia (2003)) that Indian firms choose to set up their own electricity generator to hedge against power failures or voltage fluctuations, thereby ensuring adequate quality (small generators) or lower costs (bigger power plants), and that this practice has increased over the years. As Biswas et al. (2004) point out when analysing the captive generation in Gujarat, some industries like metal producers rely critically on the provision of energy. Sometimes, as in the aluminium industries, it can determine up to 40% of total production costs.

Assume firms can pay a fixed amount G to introduce a marginal cost-reducing technology that scales down the tariff paid to the electricity by $\Theta(P_e)$, so that the price of electricity enters marginal cost as $P_e\Theta(P_e)$. The idea is that firms can buy a generator to hedge against erratic quality in supply. A few assumptions are needed to characterize this technology: if the electricity supplier charges the marginal cost of producing a unit (adjusted by quality) of electricity c, then using a generator does not reduce firms' marginal cost ($\Theta(c) = 1$). The greater the price paid for electricity, the more effective is this technology at reducing cost ($\partial\Theta(P_e)/\partial P_e < 0$), even though marginal cost is still increasing in P_e ($0 < \partial P_e\Theta(P_e)/\partial P_e < 1$)[21] . An example of such a function would be $\Theta(P_e) = \theta^{\frac{c-Pe}{P_M}}$ where $\theta > 1$ and P_M is the maximum price the electricity supplier can charge (i.e. monopolistic price).

Adoption and productivity cut-off

An individual firm j will be willing to adopt this technology if the reduction in marginal cost that drives profits up compensates the extra fixed cost G, or if

$$\begin{aligned}\Delta\pi(\varphi_j) &= \pi(P_e\Theta(P_e);\varphi_j) - \pi(P_e;\varphi_j) \\ &= \alpha(1-\rho)Y[\frac{\varphi_j H\rho}{P_e}]^{\sigma-1}[(\Theta(P_e))^{(1-\sigma)} - 1] - G \geq 0\end{aligned} \tag{3.11}$$

---

[21] Since $\partial P_e\Theta(Pe)/\partial Pe = (\partial\Theta/\partial Pe)P_e + \Theta$ where $\partial\Theta/\partial Pe < 0$ and $\Theta \leq 1$

There is a cut-off level of productivity $\hat{\varphi}$ for which a firm is indifferent between adopting or not, such that $\Delta\pi(\hat{\varphi}) = 0$. The productivity of the marginal firm adopting the captive power can be obtained as a function of φ^* by noting that $\Delta\pi(\hat{\varphi}) = \pi(\varphi^*) = 0$:

$$\hat{\varphi} = \varphi^*(\frac{G}{F})^{\frac{1}{\sigma-1}}[(\Theta(P_e))^{(1-\sigma)} - 1]^{-\frac{1}{\sigma-1}}$$

Rearranging the second term, if $\frac{F+G}{F} < (\Theta(P_e))^{(1-\sigma)}$, the productivity of the marginal adopter would be lower than the productivity of the marginal surviving firm, and all firms in the market would adopt the generator. Using $\Theta(P_e) = \theta^{\frac{c-Pe}{P_M}}$, I can obtain a threshold

$$\hat{P}_e = c + \frac{\log(1+\frac{G}{F})}{\log\theta}\frac{P_M}{(\sigma-1)}$$

such that if $P_e > \hat{P}_e$, all firms will adopt the technology. Note that the threshold is increasing in the cost of the generator (G) and decreasing on the effectiveness of the technology (θ). The intuition is that adoption is profitable when the cost of the generator relative to fixed costs is lower than the extra profits obtained for reducing the marginal cost. If the generator was freely available, all firms would adopt it even when the price of electricity was at its lowest possible level, c. On the other hand, when $P_e = c$, the technology does not reduce costs and no firm finds it profitable to pay a positive set-up cost. However, I assume that the most productive firms find adoption worthwhile for any small increase ε above c, i.e. $\Delta\pi(\Phi; c+\varepsilon) > 0$. At the firm level, it follows that for values $\hat{P}_e \geqslant P_e > c$, adoption is increasing in the price of electricity or reductions in quality ($\partial\Delta\pi/\partial P_e > 0$), and for $P_e > \hat{P}_e$, all surviving firms adopt.

I next analyse how the presence of a generator plays out at the industry level when some firms adopt and other don't. The uniform distribution of productivity provides a simple way of measuring the proportion of firms adopting (A), as $A = \frac{\Phi - \hat{\varphi}}{\Phi - \varphi^*}$. Note that firms adopting the captive power generator (with weighted average productivity $\tilde{\varphi}_a$) will coexist with firms relying exclusively on the network (with weighted average productivity $\tilde{\varphi}_n$) and the aggregate price will be

$$H^g = M^{\frac{1}{1-\sigma}}[(\frac{P_e}{\rho\tilde{\varphi}_n})^{1-\sigma} + (\frac{\Theta(P_e)P_e}{\rho\tilde{\varphi}_a})^{1-\sigma}]^{\frac{1}{1-\sigma}} \tag{3.12}$$

In this situation, a price change in the electricity supply will not affect all firms symmetrically. As a matter of fact, firms that have adopted the captive generator will suffer the price hike less and their marginal cost will increase less than firms that have not adopted the generator. The change in profits for the marginal firm φ^* will depend on the sign of the change in the aggregate price relative to its own price, namely the quotient $\frac{H^g}{P_e}$. By inspection of equation (3.12) it follows that

$$sign\ \partial(\frac{H^g}{P_e})/\partial P_e = sign\ \partial(\Theta(Pe))/\partial P_e < 0$$

The marginal consumer will not be able to remain in the market and will exit. Whenever most productive firms have an alternative to network provision of electricity, less productive firms will be more vulnerable to changes in the input price and may face market exit: the productivity threshold for survival will increase, i.e.

$\partial\varphi^*(P_e)/\partial P_e > 0$. The intuition is as follows: when the price of electricity increases in the case with no captive generator it does so for all firms. This means that the marginal cost and the aggregate price move together and the marginal firm will not be affected. When the cost-reducing technology is available more productive firms find it worthwhile to adopt, and increases in the electricity price are not matched by an increase in the aggregate price. As a consequence, the marginal firm makes negative profits and exits the market.

On the other hand, whenever there is an increase in price P_e firms in the adoption margin (i.e. with productivity $\hat{\varphi}$) will now find it worthwhile to adopt unambiguously, since the captive generator absorbs part of the increase in P_e. It follows that the cut-off productivity is decreasing in P_e, i.e. $\partial\hat{\varphi}(P_e)/\partial P_e < 0$. As a consequence, an increase in the price of electricity will affect the population and composition of firms, both by inducing the exit of firms on the margin of survival and by inducing more adoption of the marginal cost-reducing technology within the pool of surviving firms, i.e. $\partial A/\partial P_e > 0$.

Market equilibrium

The average productivity when some firms adopt a captive generator can be expressed as

$$\tilde{\varphi}^g = [\tilde{\varphi}_n{}^{\sigma-1} + (\Theta^{-1}\tilde{\varphi}_a)^{\sigma-1}]^{\frac{1}{\sigma-1}} \tag{3.13}$$

By replacing (3.13) in the profit function, the new average product can be expressed as $\pi(\tilde{\varphi}^g) = \pi(\tilde{\varphi}_n) + \pi(\tilde{\varphi}_a)$, which is greater than $\pi(\tilde{\varphi})$ for a given cut-off productivity level φ^*, since profits for users of the captive generator are greater than the profit they would have obtained had they not adopted. This would imply a ZPC shifted up with respect to the case where the captive generator was not available. Note that this holds even for the case when all firms adopt. As the expression for FEC remains the same, the new equilibrium will happen at a greater cut-off productivity[22].

3.2.3 Changes in the environment

Barriers to entry: open economy and entry fees

The equilibrium outcomes in terms of survival and adoption of captive power depend on the economic environment that determine the parameter values in the model. For example, when the economy is opened to foreign competition, the entry of new firms will affect the aggregate price H and, subsequently, sales and profits (see Fujita et al. (1999)). This would affect the Zero Cut-Off Profit Condition and market equilibrium, since in a more competitive environment, the required productivity increases. In a similar vein, the level of the entry fee, f_e, determines the required probability of success needed for firms to attempt entry, and changes in its level correspond to shifts in the Free Entry Condition curve.

To look at changes in H, assume the economy is open and transport costs are modelled as firms having to send $\tau > 1$ units to export 1 unit of output. For simplic-

---

[22] This result is similar to Melitz (2003), but rather than it being exporting firms profits that push the ZPC up, it is the availability of a cost-reducing technology which produces the increase in profits for a set of more productive firms.

ity, I concentrate on the effect of new products entering the market on domestic firms (abstracting from their export decision, since I am interested in looking at the reaction of domestic firms to bad regulation in a more competitive environment). Foreign firms will be able to enter the market if they can meet the equilibrium cut-off price of entry $p(\varphi_f) = \frac{\tau P_e^f}{\varphi_f \rho} \leq \frac{P_e}{\varphi^* \rho}$[23]. This will happen for very productive firms or for firms coming from countries with cheaper or more reliable electricity provision (lower P_e^f). Whatever the case, the entry of new firms will expand the set of varieties available to consumers and, by increasing the mass of competing firms M, a more open environment will reduce the aggregate price H and the demand for all domestic varieties.

It follows that trade barriers are associated with greater aggregate prices and profits for domestic firms, while the productivity cut-off level falls:

$$\frac{\partial \pi}{\partial H} = (\sigma - 1)\alpha(1 - \rho)Y[\frac{\varphi_j \rho}{P_e}]^{\sigma-1}H^{\sigma-2} > 0; \partial\varphi^*/\partial H < 0$$

The productivity needed to cover the fixed costs and to survive is greater in an open economy than in a protected economy, i.e. $\varphi_c^* < \varphi_o^*$. This replicates a result obtained in the literature on trade liberalization (e.g., Melitz (2003)) where an open environment forces firms at the bottom of the productivity distribution out of the market, raising the average productivity. Note that this would happen both when the generator is not available and in the situation where all firms adopt. This finding underlines the importance for domestic firms of well-functioning input markets. In such

---

[23] $\frac{\Theta P_e}{\varphi^* \rho}$, if all firms adopt the generator

settings, the impact of foreign competition will depend mostly on the price of electricity as follows: first, the relative efficiency of the domestic country at producing electricity, i.e. whether $c \geq c^f$. Second, on the regulatory process that determines P_e relative to P_e^f. It follows that if $P_e = c = c^f$, then only very productive foreign firms will be able to sell their varieties and the impact on marginal firms in the domestic markets might be lower.

In terms of adoption of the technology, the condition represented in equation (3.11) shows that $\partial \Delta\pi / \partial H > 0$, implying that the marginal firm in the adoption decision has a greater productivity level in an open economy. This is explained by the fact that as the economy opens, a firm sells less units of output. If before, at a given level of productivity and P_e, there was just enough to cover the fixed cost of the generator, now firms need to be more productive to make up for the lower sales. Under the assumption of uniform productivity distribution, and considering that the change in the adoption margin is greater than the change in the survival margin, since $|\partial\hat{\varphi}/\partial H| = (\partial\varphi^*/\partial H)\ (\frac{G}{F})^{\frac{1}{\sigma-1}}[(\Theta(Pe))^{1-\sigma} - 1]^{-\frac{1}{\sigma-1}} > |\partial\varphi^*/\partial H|$ when $P_e < \hat{P}_e$, it follows that the proportion of firms adopting increases in more protected environments, i.e. $\partial A/\partial H > 0$. This stems from the fact that adopting the technology implies an extra fixed cost, meaning that adopting firms have to be proportionally more productive than non-adopting firms to compensate for reductions in sales after a fall in the aggregate price.

A similar point could be made regarding the entry fee, f_e, firms pay to participate in the market. Larger fees shift the FEC to the left and equilibrium happens at a lower φ^*, because firms need a greater probability of survival to compensate for the increase in fees. It follows that, using the same reasoning as with more protected markets, the proportion of adopting firms should increase, $\partial A/\partial f_e > 0$.

Borrowing constraints

An additional concern is the existence of imperfections in the credit markets that might induce borrowing constraints. As an illustration, assume a competitive financial sector, i.e. where banks meet a zero-profit constraint. In the case where adopters and non-adopters of the generator co-exist in the market, firms seek a loan L to buy the generator. Whenever a firm's productivity is observable and there are not incentives to default for operating firms[24], the bank will charge a gross interest rate r equal to the opportunity cost ξ. The fixed cost for the firm will be $G = Lr$, defining a productivity cut-off for the adoption of the generator $\hat{\varphi}$ such that $\Delta\pi(\hat{\varphi}) = 0$ as in equation (3.11). If banks cannot determine the productivity of a given firm and cannot recoup anything from a bad loan, their zero-profit condition requires that the expected value of the loan equals the opportunity cost, i.e. $p(\varphi \geq \varphi^*)Lr = L\xi$ or $r = \xi/(1 - D(\varphi^*)) > \xi$. In the presence of information or enforcement problems, loans would be more expensive, which is equivalent to an increase in the fixed cost of the generator G that raises the adoption productivity cut-off. In the presence of credit

---

[24] Assume, for example, that for operating firms the probability of getting caught is equal to 1, and that cheating firms pay a hefty fine. However, firms not operating could "take the money and run," even though that is not a problem with full information.

market imperfections, more productive firms will still find it worthwhile to adopt, but firms on the margin of adoption will remain in the market, but will not use a captive generator, i.e. $\partial A/\partial G < 0$

Expenditure

Consumers derive their utility from consuming food and manufactures, and their shares in total consumption are captured in the model by α. When consumers change their preferences towards manufacturing goods (i.e. an increase in α), both the survival and the adoption margins are reduced. As the magnitude of the change is greater for adopting firms, then the ratio of adoption is greater, the greater the expenditure on manufacturing goods, i.e. $\partial A/\partial \alpha > 0$.

3.2.4 Upstream

This section of the model builds on Laffont-Tirole (1993) and analyses how the regulator of the electricity sector determines P_e. I assume that the electricity market is run by a state monopoly whose decisions are affected by the political process. Following India's example, the regulator can price discriminate between different types of consumers–agricultural and industrial–and that the farmer's lobby is such that an agricultural electricity tariff is always set equal to the marginal cost. The regulator is left with a Utilitarian Social Welfare function equal to the sum of industrial consumers and producer utilities in the electricity market (V and U, respectively):

$$W = V + U \tag{3.14}$$

Consumer surplus is weighted by a "sector-specific consumer bias" ($\lambda \in [0;1]$) that captures the relative lobbying power of industrial consumers. The monopolist's utility is simply the revenue from selling to both sectors, less a separable cost function. The cost of producing a unit of electricity is $C = cq + F_M$. Combining firms' technological use of electricity, $e = \frac{q_t - l}{\varphi}$, and equations (3.2) and (3.5), an individual demand for electricity can be defined, as $e_j^d = \alpha Y \rho \frac{P_e^{-\sigma}}{(\rho H \varphi_j)^{1-\sigma}} - l/\varphi_j$. This expression shows that demand for electricity will depend on the level of productivity. Additionally, demand price elasticity of electricity is greater than 1.

The regulator can use transfers to expand the set of possible outcomes by taxing consumers and compensating the monopolist for low levels of prices that do not meet the participation constraint, namely $U \geq 0$. When such transfers are allowed, consumers are liable to be taxed in order to fund the transfer. I assume linear pricing and define $S(P_e;\varphi)$ as the net surplus obtained by an industrial consumer with productivity φ, where the surplus is 0 for the cut-off level of productivity φ^*, i.e. $S(P_e;\varphi^*) = 0$. Equation (3.15) defines the consumer surplus

$$V = \lambda \int_{\varphi^*}^{\Phi} S(P_e;\varphi) g(\varphi) d\varphi - (1+\gamma)\tilde{t} \tag{3.15}$$

where $\tilde{t}$ is the gross transfer ($\tilde{t} \geq 0$) and γ the cost of transferring public funds. The monopolist's utility is $U = \tilde{t} + (P_e - c)\int_{\varphi} e(P_e;\varphi) g(\varphi) d\varphi - F_M$. By replacing $\tilde{t}$ in

(3.15), adding up U, and replacing in (3.14), social welfare is:

$$\begin{aligned} W &= \lambda \int_{\varphi^*}^{\Phi} S(P_e;\varphi)g(\varphi)d\varphi + \\ &\quad (1+\gamma)[(P_e - c)\int_{\varphi*}^{\Phi} e(P_e;\varphi)g(\varphi)d\varphi - F_M] - \gamma U \end{aligned} \tag{3.16}$$

The regulator chooses the price that maximizes (3.16) subject to the monopolist's participation constraint, $U \geq 0$. As the transfer of funds to the monopolist is socially costly, the utility of the monopolist will be kept at $U = 0$. The regulator also considers that the price set might affect the cut-off productivity, i.e. φ^* can be written as $\varphi^*(P_e)$.

The first order condition is

$$\begin{aligned} & -\lambda \int_{\varphi^*}^{\Phi} e(P_e;\varphi)g(\varphi)d\varphi + (1+\gamma)\int_{\varphi*}^{\Phi} e(P_e;\varphi)g(\varphi)d\varphi \\ & +(1+\gamma)(P_e - c)[\int_{\varphi*}^{\Phi} \frac{\partial e}{\partial P_e}(P_e;\varphi)g(\varphi)d\varphi - e(P_e;\varphi^*)g(\varphi^*)\frac{\partial \varphi^*}{\partial P_e}] \\ = {} & 0 \end{aligned} \tag{3.17}$$

By defining $E(P_e;\varphi^*) = \int_{\varphi^*}^{\Phi} e(P_e;\varphi)g(\varphi)d\varphi$, the price elasticity of demand for a given cut-off level of productivity φ^* as η_d[25], and the price elasticity of participation $\eta_p = \frac{\partial \varphi^*}{\partial P_e}\frac{P_e}{\varphi^*}$, the Lerner index becomes

---

[25] Where $\eta_d = -P_e \int_{\varphi*}^{\infty} \frac{\partial e}{\partial P_e}(P_e;\varphi)g(\varphi)d\varphi) / \int_{\varphi^*}^{\infty} e(P_e;\varphi)g(\varphi)d\varphi$

$$\frac{P_e - c}{P_e} = \frac{1+\gamma-\lambda}{(1+\gamma)[\eta_d + \frac{e(P_e;\varphi^*)g(\varphi^*)}{E(P_e;\varphi^*)}\varphi^*\eta_P]} \tag{3.18}$$

Condition (3.18) is a general expression for the Lerner index for cases with anti-consumer bias, consumer heterogeneity and transfer costs. When setting prices with consumer heterogeneity, the regulator has to include in the maximization programme the fact that price changes result in shifts for the marginal consumer; specifically by adding to the price elasticity of the demand (η_d), the sensitivity of the frontier (η_P) weighted by marginal consumer's relative importance in aggregate demand. In the absence of these three features, the price chosen by the regulator will be equal to marginal cost. Deviations from marginal cost pricing come from low levels of λ (i.e. industries' well-being in the electricity market does not enter fully in regulator's objective function), or from high levels of γ (i.e. as the cost of government transfers increases, it becomes optimal to recoup provision costs by charging electricity users more). As previously shown, in a closed economy the productivity of the marginal consumer is independent of the price of electricity ($\frac{\partial\varphi^*}{\partial P_e} = 0$). From (3.18), it follows that the regulator would set $P_e = \frac{c\eta_d}{\eta_d - 1 + \frac{\lambda}{1+\gamma}}$. There are two extreme cases: when the well being of consumers does not enter regulator's objective function ($\lambda = 0$), or when transfers are extremely expensive ($\gamma \to \infty$) such that the regulator sets the monopolist price, either because it would be simply maximizing the monopolist's profits or because avoiding costly transfers implies choosing the highest possible price.

Firms, industries and regulation

This result links to the firm and industry analysis in a very simple way: the parameters that determine the price level of electricity, namely c, γ and λ, will determine whether a firm adopts, as well as the composition of survival and adoption at the industry level. Everything else equal, the proportion of firms adopting will be greater in places where the monopolist is less efficient at producing electricity ($\partial A/\partial c > 0$) or the government is less efficient at taxing and redistributing ($\partial A/\partial\gamma > 0$) or where the regulator sets prices with less regard for the well-being of industrial producers ($\partial A/\partial\lambda < 0$). The introduction of the latter constitutes the most important mechanism in linking the political economy of regulation and industry outcomes. This relationship is explored further below, when applied to the Indian experience.

Quality

The discussion of prices could be extended to consider quality issues as well. In its simplest interpretation, and following the Laffont-Tirole framework (2003, see Section 2.5, Chapter 2), the results obtained above would remain the same when price and quality are perfect substitutes for consumers and the electricity producer. I could redefine P_e as the quality adjusted price such that $P_e = p_e - s$, where p_e is the actual tariff and s is quality (normalized to 0 for a standard service) and $c = c_t + s$, where c_t is the technological marginal cost. In that sense, predictions related to increases in the price of electricity and reductions in its quality should be the same. If quality is not verifiable, it provides the regulator with more freedom to set her preferred policies. For example, if the regulator has a big anti-industry bias ($\lambda = 0$) and no transfer costs in a closed economy, then the optimal price for the regulator would be the monopolistic price $P_e^M = \frac{c\eta_d}{\eta_d - 1}$. If additional constraints arise, such as a price cap $\bar{p}_e$ from the central government (or pressures for a maximum price from industrial producers, for example), the regulator would have to maximise $W = (p_e - c_t + s)Q(p_e - s; \varphi^*)$ subject to $p_e \leq \bar{p}_e$. By setting negative (i.e. below standard) quality $s = \bar{p}_e - P_e^M$, the regulator can avoid complying with the price cap and behave as if unconstrained.

3.2.5 Predictions in the Indian context

Electricity prices

The model predicts that regulatory outcomes are linked to industrial outcomes by means of the determinants of P_e. To test this in the context of Indian provision of electricity, it is important to understand the institutional framework. The Electricity Supply Act in 1948 created agencies like the Central Electricity Authority (CEA), in charge of formulating national policies and assessing the technical and economic viability of power projects, and the State Electricity Boards (SEB), whose board members are appointed by the state government (Chapter III, 5.2). The SEBs are fully vertically integrated, owning the four segments of the industry–generation, transmission, distribution and commercialization– and set their own pricing and investment policies.

In terms of linking the electricity tariffs to the government's anti-industry bias (λ) and inefficiency (γ), sectorial analysis for Indian SEBs deemed the energy provision as "more of a social service than a business. Prices do not reflect costs and state institutions are dependent on allocations from the public budget" (TERI (1999)). Also, "the SEBs became bastions of political patronage rather than true business enterprises (...) Reformers faced political opposition from farmers, who had come to rely on enormous quantities of low-cost electricity for pumping water" (Tongia (2003)). Overall, subsidies to farmers and domestic consumers by states via the electricity provision have been estimated to be in the order of 1 to 1.5% of India's GDP (World Bank (2000)). In all circumstances described by the previous model(e.g. availability of a generator, open economy, borrowing constraints), other than a closed economy, an increase in the price of electricity would induce the exit of firms at the margin ($\eta_P > 0$). The loss of consumers would reduce the monopolist's revenue regardless of the weight assigned to industrial consumers and would reduce the price set by the regulator. In all cases, states with more anti-industry bias and with less efficient governments should show higher electricity prices. Unfortunately, there is not enough information to test whether prices set by SEBs are sensitive to changes in the cut-off level for firms' survival.

The model predicts that cross-subsidization should be more prevalent in states with a stronger rural lobby. In particular, the measure of rural lobby I use reflects the adoption of high yield varieties (HYV) after the start of the Green Revolution. HYV seeds are known to be high-cost/high-yield because they multiply agricultural productivity if timely irrigation, among other inputs, is provided. The cheapest way for farmers to secure water is by means of electric pumpsets that suction water from the water table via tubewells. The importance of food production for reducing the possibility of famines gives farmers, in particular larger and richer ones, the upper hand. Dhanagare (1989) estimates that "the prosperity unleashed by the Green Revolution was distributed differentially, putting the small and marginal farmers at a relative disadvantage. The high cost/high yield technology called for capital investments beyond the means of a majority of small and marginal farmers." With a stronger economic position, rural rich farmers also developed political leverage that was translated into further subsidies. In particular, Gulati and Narayanan (2003) show how what they call "the subsidy syndrome in Indian agriculture" after the introduction of the new seeds in late 60s, became a fundamental instrument of economic policy that was mainly channelled through available and affordable fertilizers, irrigation and electricity. Dhanagare also points out that "only the needs of rich farmers were catered by the electricity boards," since while the unit price was heavily subsidized, connection charges were so expensive that only rich farmers could afford them. For example, in the case of Punjab, the state at the forefront of HYV adoption, Simms (1988) describes a situation in which "farmers are aware of the power they hold due to their strategic importance in the national economy (...) Political action has led to extensive changes in rural Indian Punjab." Describing a rally organized by one organization of farmers, Simms points out that "the state government was thrown off balance by their show of force and conceded many of their demands." This description of the Indian context combined with the upstream section of the model laid out earlier suggests that the power of rural lobbies and government's efficiency should be linked to the pricing strategy of the regulator, i.e. states with more adoption of HYV seeds and more budget deficits charge a higher price to industrial consumers of electricity (Prediction 1).

Adoption: individual characteristics

Poor regulatory decisions resulting in excessive prices or reduced quality necessarily increase the costs of industrial production and affect the competitiveness of firms, ultimately distorting their choice of technology or industry. As a consequence, firms set up their own electricity generator to hedge against power failures or voltage fluctuations to ensure the adequate quality or to reduce costs. This practice has increased over the years in India. As mentioned earlier, Biswas et al. (2004) point out that some industries like metal producers rely critically on the provision of energy and electricity is often a sizeable part of their total production cost. To ensure the quality of the provision and to remain competitive in the market, firms tend to install captive power plants. The installed captive capacity varies considerably across states and industries. An infrastructure report produced by Price Waterhouse Coopers for the government of Chhattisgarh shows that in 1998, the average captive capacity was around 17%, with significant variation. In some states–like Karnataka (30.5%) and

Haryana (38%)–captive power is significant while in others–like Jammu & Kashmir and Himachal Pradesh– firms use exclusively SEB's electricity. Industry-wise, there is considerable variation as well, with industries like electronics producers using self-provision of energy just up to a 0.5% of total captive generation capacity, while engineering or metals and minerals industries surpass 20%.

The model predicts adoption of the captive generator according to a set of parameters that varies across firms. For example, as noted earlier, the electricity price P_e could be understood as a quality-adjusted tariff. In that case, some characteristics of the firms, such as their location (urban vs. rural), might be correlated with the quality of supply received and have an impact on their decision to adopt a generator. Additionally, measures that could be capturing an idiosyncratic productivity (φ) or access to finance (G)–such as whether a firm is public or multiproduct–should also be positively correlated with adoption. The model also predicts that if adopting firms coexist with non-adopting firms, measures of output and size (e.g., workers) should be greater for adopting firms. It follows that firm characteristics, including exposure to lower quality of electricity, greater productivity, access to credit and size are all correlated with more adoption of captive power (Prediction 2).

Adoption: state characteristics

As previously mentioned, regulatory outcomes vary only at the state level. As the model predicts that adoption of captive power increases with the price of electricity and decreases with quality, in-house generation of electricity should be more prevalent in states where industries are charged relatively more for electricity and where the quality of provision is poorer. Additionally, the model identifies variables that drive pricing policies. In particular, the lobbying power of farmers and government's inefficiency should be positively correlated with adoption. Finally, I use data on urbanization rates to proxy for the share of consumption in manufacturing goods (represented by α in the model) and test whether it is positively correlated with adoption. The incidence of adoption of captive power should be greater in states where the SEB charges higher electricity tariffs to industrial consumers or provides lower quality. Adoption should also be positively correlated with farmers' lobbying power and measures of government and SEB inefficiency (Prediction 3).

Adoption: industry characteristics

The impact of the regulatory context for the adoption decision in Indian industries can also be captured by some features of the model. In particular, adoption should be greater in sectors more protected from international competition via trade tariffs. This comes from the finding that in industries where some firms adopt and others do not, the margin for adopting firms is more sensitive to changes than the margin for surviving firms. Even though India considerably reduced its trade tariffs in 1991, there is significant variation in tariff levels across industries both before and after the reforms. Another widespread industrial policy in India after independence was the establishment of a licensing scheme that was dismantled progressively from 1985 onwards[26]. The license system can be understood in the framework of this model as the entry fee f_e firms have to pay before knowing whether they will survive or not. As with tariffs, greater fees should be correlated with a greater proportion of firms adopting the power generator. In short, the fourth prediction is that the incidence of adoption of captive power should be greater in industries that are protected by greater trade tariffs and subject to licenses (Prediction 4).

Productivity

The prediction regarding productivity is in line with the trade literature where protection allows for the survival of less productive firms. This paper also links productivity with the adoption of an in-house generator of electricity. In particular, firms which adopt are expected to be more productive and bigger, on average. The model also predicts that the margin of survival and adoption will be lower for more protected sectors. Industrial sectors that are more protected (whether by trade tariffs or licensing) should also produce less and be less productive, on average. Additionally, within an industry, the average productivity and output should be greater for firms adopting the captive generator (Prediction 5).

---

[26] For a information on the delicensing process and its effects, see Aghion et al. (2007)

3.3 Data and Descriptive Statistics

To test the predictions laid out above, I use data on the adoption of captive power generators at the factory level. Tests on the effect of regulatory outcomes are done at the state level, while measures of openness vary at the industry level.

Factories: The dataset is a repeated cross-section of factories for the years 1990, 1994 and 1997 collected by the Annual Survey of Industries, Department of Statistics, Government of India. The basic sample consists of more than 145,000 observations, but is lower when combined with state or industry measures. Since the sample includes firms in the "census sector" (i.e. with more than 100 workers) and firms that are sampled (the sampling design adopts 3-digit-industry groups per state as stratum and covers all the units in a span of three years), all regressions weigh units of observation accordingly. Table 1 provides some summary statistics, where it can be seen that around 1 in 3 firms have a captive generator of electricity. Some factory characteristics capture features of the model. For example, location (around 30% rural firms, 58% urban and the rest, in one of the metropolitan areas, i.e. cities with more than 1 million people as defined by the 1991 Indian census) might reveal information about the quality of the electricity that firms are exposed to (e.g. blackouts, pilferage, technical losses, power surges). Other characteristics might reveal idiosyncratic productivity, such as whether the factory produces more than one good (61.8% do) or size (25% are in the census sector). Finally, credit constraints might be captured by whether the firm has access to credit (a large number of firms–74%–have access to some kind of credit). Moreover, if firms are public (around 14% are), we should expect them to access formal (i.e. cheaper) sources of funds.

States: The variables capturing State Electricity Boards' regulatory outcomes include data on prices such as tariffs charged to industries and agricultural users, average tariffs (the sum of prices paid by industries, farmers, households and commercial users, weighted by their share in total sales). Quality measures include transmission and distribution losses (as a percentage of total electricity traded) and outages in thermal plants. Table 1 shows significant variation in both measures. Other relevant information related to SEBs efficiency is captured by variables such as a measure of network extension (i.e. the number of consumers connected to the electricity network), unit cost of electricity production, quantity of employees per Kwh produced and subsidies received from the central government. Other state-level variables, aimed at capturing the relative power of farming lobbies (λ), government efficiency (γ) and consumers' preference for manufacturing goods (α) are proxied by the incidence of the Green Revolution (the proportion of cultivated land cropped with HYV seeds), state governments' budget deficits and the proportion of urban population, respectively.

Industries: India-wide industrial policies are captured by measures of trade tariffs at the 3-digit industry level and by a dummy at the 4-digit industry level equal to one if firms require a license to enter that industry. The reduction process of barriers to entry for foreign products and new domestic producers started slowly in 1985 and accelerated after 1991. However, the process affected different industries at different points in time and in different degrees (in particular, in terms of trade tariffs reduction). Table 1 shows that factories were exposed to different degrees of protection. For example, around 76% of the firms in the sample were subject to a licensing process in 1990 while in 1994 and 1997, it was only around 37%. The mean imports tariff also dropped considerably, from around 1.3 in 1990 to 0.89 in 1994 and 0.47 in 1997.

3.4 Results

3.4.1 Pricing

The model predicts (Prediction 1) that in states where the anti-industry bias is stronger, the price for industrial consumers will be higher. The anecdotal evidence points to regulators having a pro-farmer bias and an anti-industry bias across the board in Indian states. However, the intensity of that bias could be associated with farmers' need for electricity. A measure of this can be obtained by looking at the adoption of an agricultural technology that is intensive in electricity. In particular, the use of high yield variety (HYV) seeds that were introduced with the Green Revolution required timely irrigation and electric pumpsets were the cheapest way of accessing groundwater.

Table 2 compares measures of relative prices and quality for states where HYV seed adoption was above the median [27], and the rest of the states. The first measure is the ratio between average tariff and unit cost, available from 1974 to 1997. Even though all SEBs struggled in that period to recoup their production costs, states intensive in HYV had a significantly lower proportion of their costs covered by sales revenue. All other measures are only available for the period 1990-1997, and show that states intensive in HYV charged higher tariffs for industrial producers, relative to the average tariff or the unit cost. Additionally, the tariff charged to farmers was much lower for both groups, relative to the average tariff or the industrial tariff respectively, but both ratios are significantly lower for states intensive in HYV. Table 2 suggests that states where farmers relied more on electricity tended to set pro-farmer pricing schemes in the electricity sector, and that industrial users tended to be charged more. However, measures of quality, such as thermal outages or transmission and distribution losses were similar, on average, for both groups of states.

To test whether this pricing policy started right at the beginning of the Green Revolution, I use data on the ratio of tariffs charged to industrial users of electricity over the tariff charged to farmers and a continuous measure of the proportion of land under HYV seeds for three years in the 70s (1973-1975). Column (1) in Table 3 shows that industries were charged relatively more, the greater the proportion of HYV seeds, even when controlling for other time varying state variables such as pop-

---

[27] High HYV states are Andhra Pradesh, Bihar, Haryana, Kerala, Punjab, Tamil Nadu and Uttar Pradesh. These states were above the median of adoption both in 1974 (at least 27%) and in 1984 (at least 41%).

ulation, governments' real development expenditure and the party of the state Chief Minister. The model also predicts that electricity tariffs will be affected by governments' efficiency at transferring resources The intuition is that as governments become less efficient, the cost of transfers increases and the regulator has to rely on higher prices to recoup costs. Column (2) shows that states with higher budget deficits tended to charge higher prices to industrial producers, keeping HYV adoption fixed. One concern is that of reverse causality: i.e. because some states provided cheaper electricity, then HYV adoption follows. To mitigate this concern, in Column (3) I instrument HYV adoption with the proportion of districts with abundant groundwater, a geographical characteristic associated to the successful adoption of HYV (as in the previous chapter), and find that results hold. Columns (4) to (6) use 11-year information for 15 states on average electricity tariff relative to average cost. Results show a similar pattern: states intensive in HYV seeds and with larger budget deficits tended to have less sustainable pricing policies. That suggests that even if industries were carrying the burden of pro-farmer pricing, SEBs had a harder time recouping electricity production costs. Column (6) uses the same instrument as in column (3), this time interacted with year dummies to capture divergence in HYV adoption across states. Overall, the data shows that Indian industries faced steeper prices in states with more powerful farming sectors and less efficient governments.

3.4.2 Adoption

The adoption of a captive generator in the model is driven by characteristics that can vary at the firm, industry and state level. In particular, the model predicts (Prediction 2) that factory characteristics that capture exposure to low quality of electricity, more access to credit and greater size and productivity are all associated with greater adoption of a generator. To look at firm level characteristics that capture each of these phenomena and see whether they are correlated with adoption, I run a linear probability model of the form

$$A_{fisy} = \alpha_{isy} + \beta \mathbf{X}_f + \varepsilon_{fisy}$$

where $A = 1$ if factory f in industry i in state s in year y generates in-house some of the electricity it needs to produce and 0 otherwise. Individual characteristics are captured by X_f and state-industry-year fixed effects are captured by α. This means that the variation comes from firms producing in the same 3-digit NIC industry located in the same state and producing at a given point in time.

Anecdotal evidence shows that the supply of electricity in rural areas tends to be subject to more blackouts, power surges and erratic tension. Urban and metropolitan areas tend to suffer this problem less, in particular cities. This means that I should expect factories in rural areas to adopt more generators than those in urban and metropolitan areas. Column (1) in Table 4 shows that a rural firm is around 6% more likely to use a generator. Column (2) also includes a dummy for urban firms and shows that both rural and urban factories are respectively 11% and 6% more likely to own a generator than firms in metropolitan areas (defined as cities with a population of 1 million people or more).

The ownership structure of a firm can provide information on both its productivity and its access to external finance at a relatively cheap cost. Column (3) shows that factories belonging to a public company are around 24% more likely to adopt generators than firms that are privately (whether by individuals or societies) or state-owned. Another characteristic that might be linked to productivity is whether a factory produces many products or just one. Column (4) shows that the probability of owning a generator is around 9% greater for multiproduct factories. As predicted in the model, more productive firms tend to produce more. More productive firms, according to the model, should also be larger. To tell apart large and small firms, I use a dummy variable equal to 1 for factories that employ more than 100 employees (and are included in the census sector). Column (5) shows that the probability of owning a generator is 14% greater for those firms. The coefficient on public companies has dropped, suggesting that it was previously capturing some of the size effect, since around 60% of public firms are also in the census sector. Another measure of access to credit is also associated with more adoption of captive power. Column (6) shows that factories that have outstanding loans are 7% more likely to produce some of their consumed electricity. Finally, Column (7) shows that the larger the number of workers, the larger the rate of adoption even when controlling for whether a firm is in the census sector or not. In short, as predicted by the model, adoption of a captive generator of electricity is more likely for firms whose characteristics are associated with more exposure to lower electricity quality, greater productivity, larger size and access to credit.

An important feature of the model is the link between the upstream regulatory process and downstream behaviour, as reported by Prediction 3. The model predicts that a firm's decision to set up a captive power generator is positive in electricity prices and negative in quality. Table 5 presents results of the impact of State Electricity Boards' regulatory outcomes on the adoption decision, controlling for industry-year fixed effects and for all individual characteristics examined in Table 3. Column (1) shows that states that charge industries relatively more than other electricity consumers observe significantly greater levels of adoption. In terms of magnitudes, a standard deviation increase in the tariff charged to industries is associated with an increase in the probability of adoption of 4 percentage points. Column (2) finds a similar result using a measure of anti-industry bias relative to agricultural and domestic consumers only. Column (3) includes a measure of the average quality provided by the SEB, namely Transmission and Distribution losses as a proportion of total electricity sent into the network. The coefficient is positive and significant, meaning that a worse-performing electricity network is associated with more factories turning to self-provision of at least some of their electricity. Note that the coefficient on pricing remains similar, meaning that lower quality is not necessarily working through prices. Also note that among individual characteristics, there is a significant drop in the coefficients associated with rural and urban dummies only. When SEB quality provision is controlled for, the difference in adoption between metropolitan areas and rural and urban areas is smaller, suggesting that location was indeed picking up exposure to lower quality (unreported). Overall, the magnitude is also significant: an increase of one standard deviation in the measure of transmission and distribution losses is associated with additional 8 percentage points in the probability of adoption. Column (4) uses a more restrictive measure of quality (percentage of outages in thermal plants only) and again obtains a positive correlation between lower quality and adoption. Column (5) includes a measure of network extension that might be driving the ability to price or to provide good quality. More electricity connections per capita are negatively correlated with the adoption of captive power. The coefficient on prices is slightly larger and the coefficient on transmission and distribution losses drops slightly, and both remain strongly significant.

Column (6) includes the percentage of HYV adoption, which proxies for the anti-industry bias parameter in the regulator's objective function. Its coefficient is strong and positively correlated to adoption. The channel linking HYV to adoption of captive power through the electricity price seems to be at work since the coefficient on the electricity tariff charged to industries drops by almost 50% in column (6) with respect to column (5). With the thought that areas with a greater urban population might consume more manufacture–in turn affecting the margins for adoption and survival–Column (7) includes a measure of urban population at the state level and finds, as predicted by the model, that adoption increases with urbanization. The last column in Table 5 includes some measures of government and SEB efficiency that should affect adoption through price, according to the model. All variables remain significant and similar in magnitude, except the electricity tariff for industries, which drops in both. Factories in states with a greater budget deficit and with SEBs that require more subsidies from the state government tend to adopt generators more. Firms in states with larger electricity unit costs tend to adopt less, suggesting that is not the productive efficiency of the SEB that drives the results, but its anti-industry bias.

In the next set of regressions in Tables 6 and 7, I include an interaction between regulatory outcomes and firms characteristics, controlling for 3-digit industry-year fixed effects. If the predictions of the model are correct, then firms should react to bad regulatory outcomes when they are able to do so. Additionally, interactions can be used to identify who is most affected by regulatory outcomes. Table 6 looks at the two measures of pricing, namely the ratio between industrial and average tariffs and the burden on industries relative to the agricultural and domestic consumers. Columns (1) and (2) interact these measures with the dummies for urban and rural location. In both cases, the positive coefficients on the interaction terms suggest that adoption is more likely in rural and urban areas than in metropolitan factories, but only if the price is greater for industries. The negative coefficients on the rural and urban dummies suggest that if SEBs' pricing policy is not biased against them, then there is less adoption in those regions than in metropolitan areas, if any at all. Additionally, the negative and significant coefficients on the measure of pricing suggests that the larger the price, the lower the adoption in metropolitan areas. That suggests that pricing policies seem to influence the adoption decision only in non-metropolitan areas.

All interactions with other individual characteristics are not significant for the ratio of industrial tariff over average tariff (see, for example, Column (3)). However, in the remaining regressions shown in Columns (4) to (6), the interaction of individual characteristics with the measure of relative burden on industries vis-a-vis the domestic and agricultural sectors is positive and significant. Larger firms, multiproduct firms and firms with access to credit seem to be adopting more, the worse the anti-industry bias. This result suggests that factories' indicators of productivity and access to credit allow firms to react to bad regulatory outcomes, in line with Predictions 2 and 3. Note that as the generator take-up among public firms and census sector firms is very high (above 50%), there is not enough variation to obtain a significant coefficient in the interactions.

Table 7 looks at firms' characteristics interacted with measures of quality, i.e. transmission and distribution losses and outages in thermal plants. Columns (1) and (2) look at interactions with urban and rural dummies. With both measures of quality a similar pattern emerges: quality seems to matter more for metropolitan firms, since the coefficients on the quality measures are positive and significant. And even though rural and urban firms tend to adopt more captive generators, they adopt slightly less in areas with lower quality. As in Table 6, Columns (3) to (5) show that larger firms, multiproduct firms and those with access to credit tend to adopt more when the quality provided is lower, suggesting again that productivity and credit help firms deal with bad regulatory outcomes. Two patterns emerge from the last two tables: first, metropolitan firms react more to quality in the electricity supply and urban and rural firms react more to prices. Second, as predicted by the model, productivity, size and access to credit are positively correlated with adoption, especially when regulatory outcomes are more unfavourable.

To test Prediction 4–i.e. adoption increases with protection–I look at how the competitive environment affects the adoption decision by using a measure of whether the sector (at the 4-digit industry level) is licensed or not, and the average trade tariff (measured at the 3-digit industry level). In all cases, 2-digit industry-state-year fixed effects are included and all individual characteristics are controlled for. As predicted by the model, Column (1) and (2) in Table 8 show that adoption is more likely in more protected environments. The first column shows that firms in licensed sectors are 9% more likely to adopt. The coefficient drops to 7% when the trade tariff is included in the second column, but both measures are positive and significant. Column (3) controls for a measure of technological needs of electricity (at the 3-digit industry level), to check that results in previous columns are not capturing a correlation between protection and technology. The coefficient is positive and significant but barely affects the coefficients on license and tariffs. Columns (4) and (5) include some interactions with factory characteristics. Column (4) shows that the licensing scheme only affects non-metropolitan firms' decision to adopt. Rural and urban firms in non-licensed sectors are more likely to produce captive power than their metropolitan counterparts, but the probability is significantly higher in licensed industries. Column (5) finds that larger firms tend to adopt more in licensed industries. As predicted by the model, for a given characteristic such as size, adoption increases with protection.

To see how protection interacts with regulatory outcomes, I split the sample between licensed and delicensed sectors and run the linear probability model of adoption on price and quality measures, controlling for 3 digit industry-year fixed effects and individual characteristics. Table 9 shows that the effect of higher electricity prices and lower quality are very similar in magnitude and the statistical significance for both licensed and unlicensed industries.

3.4.3 Production and Productivity

The model predicts that average output and productivity of observed firms should be lower in more protected environments (Prediction 5). Column (1) in Table 10 shows that the log of real output is lower, on average, for firms in licensed sectors and in industries with greater trade tariffs, controlling for firms' characteristics. Column

(2) isolates a rough measure of total factor productivity by controlling for the log of skilled and unskilled workers, the log of real fixed capital and the log of real materials consumed. Firms in licensed sectors tend to be less productive, even though the result loses significance for the measure of trade protection. Columns (3) and (4) include a dummy equal to 1 if the firm has a captive generator. The prediction of the model that firms which adopt are on average larger and more productive also holds. In Columns (5) and (6), I split the sample between firms with captive power and firms without and find that for both sets of firms, productivity tend to be lower in sectors that are licensed. The model also predicts that the average product and productivity of adopting firms should be greater in sectors that are more open to the entry of new firms. To test this idea, I split the sample between firms in licensed and unlicensed sectors. Columns (7) and (8) show that output is greater for adopters in both sectors. However, the difference between adopters and not adopters is significantly higher in unlicensed sectors. Columns (9) and (10) look at the effect on residual productivity and show that factories that have captive power are significantly more productive than non-adopting firms, but only in delicensed industries. This result suggests that protection reduces average productivity for adopters, as predicted by the model.

3.5 Conclusion

Developing countries aiming at increasing their industrial base have often disregarded the importance of infrastructure provision. By analysing how the regulation of a monopolistic supplier of a fundamental input in production affects producers of manufactures, this paper explores a specific mechanism that explains how firms react to the inadequate provision of infrastructure. This chapter shows that only firms with high productivity and access to credit can actually reduce the negative effects of expensive and erratic electricity, by adopting a captive power generator. The Indian data also suggests a heterogeneous impact of regulation according to location. Firms in metropolitan areas appear to be more sensitive to quality while firms in rural and urban areas react more to prices and protection. On average, firms find it easier to cope in more protected environments–e.g. ones with high trade tariff and entry fees–where less competition raises their profits and makes the outside option affordable. Otherwise, in a more competitive environment, firms that could have survived with cheaper and better inputs will exit the market, while firms that were better prepared to cope with provision failure by adopting a captive generator will not be able to afford it any more. It follows that countries undergoing a process of opening to foreign competition and deregulation should pay particular attention to the negative impact of low quality and biased pricing policies in the provision of infrastructure goods and to the distributional effects that follow.

Table 1: Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
Adoption of Captive Generator	30.5%	0.46	0	1
Rural	30.6%	0.46	0	1
Urban	57.8%	0.49	0	1
Public Company	14.1%	0.35	0	1
Multiproduct	61.8%	0.49	0	1
Census Sector	25.1%	0.43	0	1
Access to Credit	73.9%	0.44	0	1
Log Workers	3.63	1.35	0	10.75
Log Output	15.98	2.09	0.35	24.74
Ind. Tariff / Avg. Tariff	1.83	0.40	0.97	2.89
T&D Losses	20.8%	4.43	15.3	33.4
Licensed	48.7%	0.50	0	1
Trade Tariff	0.86	0.45	0	3.44

Table 2: Differences in Electricity Pricing Schemes According to Green Revolution intensity at the State Level

Incidence of Green Revolution	Avg. Tariff / Unit cost	Ind.Tariff / Avg. Tariff	Ind. Tariff / Unit Cost	Agr. Tariff / Avg. Tariff	Agr. Tariff / Ind. Tariff	Outages	T&D Losses
Low	0.76	1.46	1.05	0.28	0.24	15.5	24.7
High	0.71	1.80	1.28	0.17	0.09	18.4	22.7
Difference	0.05**	-0.34***	-0.23***	0.11**	0.15***	-2.9	2

Significance level of differences: * significant at 10%; ** at 5%; *** at 1%. Data for column (1) is for the period 1974-1997. All other columns use years 1990-1997. All tariffs and unit costs are in Paise (cents of rupee) per KWh. "T&D losses" are Transmission and Distribution losses as a % of total electricity traded. "Outages" are forced outages of thermal stations (%). States with high incidence of Green Revolution: Andhra Pradesh, Bihar, Haryana, Punjab, Tamil Nadu, Uttar Pradesh.

Table 3: Electricity Pricing and Green Revolution

	(1)	(2)	(3)	(4)	(5)	(6)
	Ind Tariff / Agr Tariff			Avg Tariff / Unit Cost		
	OLS		IV	OLS		2SLS
HYV Adoption	0.85	1.21	0.66	-0.54	-0.73	-1.28
	(0.31)***	(0.35)***	(0.35)*	(0.11)***	(0.10)***	(0.15)***
Budget Deficit		82.1	59.5		-122.1	-159.1
		(36.7)**	(28.3)**		(17.2)***	(19.1)***
Observations	45	45	45	165	165	165
Adj. R square	0.66	0.71	0.69	0.54	0.62	0.55
F-test P-value			<1%			<1%
Over-id test P-value						0.31

Robust standard errors in parentheses, * significant at 10%; ** at 5%; *** at 1%. All tariffs and unit costs are in Paise (cents of rupee) per KWh. "HYV Adoption" is the net proportion of cropped area under High Yield Variety seeds. "Budget deficit" is the deficit as a proportion of state s product. Column (3) uses the "proportion of districts with abundant groundwater" as an instrument and Column (6) uses the same instrument interacted with year dummies. Low F-Test P-values suggest a strong correlation between instruments(s) and instrumented variables. Low Over-id P-values suggest that the instruments are correctly excluded from the second stage. State controls include development expenditure, proportion of rural population, credit availability, proportion of votes to the Congress Party and the party of the Chief Minister. Columns (3)-(6) include year dummies.

Table 4: Firm Characteristics and Adoption of Captive Generator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Adoption of Captive Generator						
Rural	0.06	0.11	0.09	0.09	0.09	0.09	0.07
	(0.006)***	(0.008)***	(0.008)***	(0.008)***	(0.008)***	(0.008)***	(0.008)***
Urban		0.06	0.05	0.05	0.05	0.05	0.04
		(0.006)***	(0.006)***	(0.006)***	(0.007)***	(0.007)***	(0.007)***
Public Company			0.24	0.23	0.18	0.19	0.12
			(0.007)***	(0.007)***	(0.007)***	(0.006)***	(0.007)***
Multiproduct				0.09	0.08	0.07	0.04
				(0.007)***	(0.006)***	(0.007)***	(0.006)***
Census Sector					0.14	0.14	0.07
					(0.007)***	(0.006)***	(0.006)***
Access to Credit						0.07	0.06
						(0.005)***	(0.005)***
Log Workers							0.08
							(0.003)***
Observations	146251	146251	146251	146251	146251	146251	143797
R-squared	0.35	0.35	0.37	0.38	0.38	0.39	0.42

Robust standard errors in parentheses clustered at the state-industry level, * significant at 10%; ** significant at 5%; *** significant at 1%. The regressions are linear probability model where the dependent variable "Adoption" is equal to 1 if the firms produces some of the electricity used in production. All variables are dummies, except "Log Workers." For location, the excluded category is "Metropolitan." "Multiproduct" is equal to 1 if the factory produces more than 1 product. "Census sector" includes factories with more than a 100 employees. The sample sector follows a sampling design adopting State- 3 digit industry group as stratum so as to cover all the units in a span of three years. "Log workers" refers to unskilled workers. State* 3-digit industry * year fixed effects are included in all cases.

Table 5: Regulatory Outcomes at the State Level and Adoption of Captive Power

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adoption of Captive Generator							
Ind. Tariff / Avg. Tariff	0.095		0.094	0.121	0.116	0.067	0.069	0.018
	(0.016)***		(0.014)***	(0.016)***	(0.014)***	(0.012)***	(0.012)***	(0.012)
Relative Burden on Industries		0.20						
		(0.014)***						
T&D Losses			0.017		0.012	0.008	0.010	0.015
			(0.002)***		(0.001)***	(0.001)***	(0.001)***	(0.001)***
Outages				0.010				
				(0.001)***				
Electricity Connections					-0.039	-0.045	-0.051	-0.078
					(0.004)***	(0.003)***	(0.004)***	(0.005)***
HYV Adoption						0.67	0.68	0.63
						(0.04)***	(0.04)***	(0.03)***
Urban Population							0.23	0.84
							(0.12)*	(0.13)***
Budget Balance								-0.91
								(0.17)***
Unit Cost								-0.02
								(0.002)***
Net Subsidies								0.02
								(0.002)***
Firm Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	133186	133186	133186	127940	133186	133186	133186	133186
R-squared	0.22	0.25	0.24	0.24	0.26	0.30	0.31	0.32

Robust standard errors in parentheses, clustered at the state-industry level * significant at 10%; ** at 5%; *** at 1%. The regressions are a linear probability model where the dependent variable "Adoption" is equal to 1 if the firms produces some of the electricity used in production and includes 3-digit industry * year fixed effects. All tariffs and unit costs are in Paise (cents of rupee) per KWh. "Relative Burden on Industries" is percentage sales revenue to percentage sales for industries relative to agriculture and domestic sectors only "T&D Losses" are Transmission and Distribution losses as a % of total electricity traded. "Outages" are forced outages of thermal stations (%) "Electricity Connections" are per capita. "HYV Adoption" is the net proportion of cropped area under High Yield Variety seeds. "Urban Population" is % of total population. "Budget Balance" is the budget superavit as a proportion of a state s product" "Net Subsidy" is the percentage of subsidies from State Government over revenues

Table 6: Interactions Between SEB Pricing Policies and Firm Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Adoption of Captive Generator					
Ind. Tariff / Avg. Tariff (IT/AT)	-0.062		0.075			
	(0.03)**		(0.02)***			
Relative Burden on Industries (RBI)		-0.67		0.11	0.14	0.14
		(0.11)***		(0.02)***	(0.01)***	(0.02)***
T&D Losses	0.016	0.10	0.017	0.10	0.10	0.10
	(0.002)***	(0.001)***	(0.002)***	(0.001)***	(0.001)***	(0.001)***
Rural	-0.14	-1.27	0.12	0.12	0.12	0.12
	(0.05)	(0.18)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
IT/AT * Rural	0.16					
	(0.03)***					
RBI * Rural		0.82				
		(0.11)***				
Urban	-0.19	-1.35	0.09	0.09	0.09	0.09
	(0.05)***	(0.18)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
IT/AT * Urban	0.17					
	(0.03)***					
RBI * Urban		0.85				
		(0.11)***				
Log Workers	0.07	0.07	0.06	0.04	0.07	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
IT/AT * Log Workers			0.01			
			(0.01)			
RBI * Log Workers				0.015		
				(0.006)**		
Multiproduct	0.05	0.04	0.05	0.04	-0.04	0.04
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.03)	(0.01)***
RBI * Multiproduct					0.04	
					(0.01)***	
Access to Credit	0.06	0.04	0.05	0.04	0.04	-0.02
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.03)
RBI* Access to Credit						0.03
						(0.015)**
Public Company	0.13	0.14	0.14	0.14	0.14	0.14
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Census Sector	0.06	0.07	0.07	0.07	0.07	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Observations	133186	133186	133186	133186	133186	133186
R-squared	0.26	0.26	0.24	0.26	0.26	0.26

Robust standard errors in parentheses, clustered at the state-industry level * significant at 10%; ** at 5%; *** at 1% The regressions are a linear probability model where the dependent variable "Adoption" is equal to 1 if the firms produces some of the electricity used in production and includes 3-digit industry * year fixed effects All tariffs and unit costs are in Paise (cents of rupes) per KWh. "Relative Burden on Industries" is percentage sales revenue to percentage sales for industries relative to agriculture and domestic sectors only "T&D Losses" are Transmission and Distribution losses as a % of total electricity traded.

Table 7: Interactions Between SEB Quality Provison and Firm Characteristics

	(1)	(2)	(3)	(4)	(5)
	Adoption of Captive Generator				
Ind. Tariff / Avg. Tariff	0.11	0.13	0.09	0.09	0.09
	(0.01)***	(0.02)***	(0.01)***	(0.01)***	(0.01)***
T&D Losses	0.05		0.01	0.01	0.01
	(0.002)***		(0.003)***	(0.001)***	(0.001)***
Outages		0.03			
		(0.002)***			
Rural	0.79	0.37	0.12	0.12	0.12
	(0.06)***	(0.03)***	(0.01)***	(0.01)***	(0.01)***
T&DL * Rural	-0.04				
	(0.003)***				
Outages * Rural		-0.02			
		(0.001)***			
Urban	0.69	0.31	0.09	0.09	0.09
	(0.06)***	(0.03)***	(0.01)***	(0.01)***	(0.01)***
T&DL * Urban	-0.034				
	(0.002)***				
Outage * Urban		-0.014			
		(0.002)***			
Log Workers	0.07	0.07	0.03	0.07	0.07
	(0.01)***	(0.01)***	(0.02)*	(0.01)***	(0.01)***
T&DL * Log Workers			0.002		
			(0.0001)**		
Multiproduct	0.05	0.04	0.04	-0.05	0.04
	(0.01)***	(0.01)***	(0.01)***	(0.04)	(0.01)***
T&D L * Multiproduct				0.004	
				(0.001)***	
Access to credit	0.05	0.05	0.05	0.04	-0.14
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.03)***
T&DL * Access to Credit					0.009
					(0.001)***
Public Company	0.14	0.14	0.14	0.14	0.14
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Census Sector	0.07	0.07	0.07	0.07	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Observations	133186	127940	133186	133186	133186
R-squared	0.25	0.25	0.26	0.26	0.26

Robust standard errors in parentheses, clustered at the state-industry level. * significant at 10%; ** at 5%; *** at 1% The regressions are a linear probability model where the dependent variable "Adoption" is equal to 1 if the firms produces some of the electricity used in production and includes 3-digit industry * year fixed effects. All tariffs and unit costs are in Paise (cents of rupee) per KWh. "T&D Losses" are Transmission and Distribution losses as a % of total electricity traded. "Outages" is forced outages of thermal stations (%).

Table 8: Competitive Environment and Adoption of Captive Power

	(1)	(2)	(3)	(4)	(5)
	Adoption of Captive Generator				
Licensed	0.09	0.07	0.07	0.02	0.01
	(0.01)***	(0.02)***	(0.01)***	(0.01)	(0.01)
Trade tariff		0.07	0.06	0.06	0.06
		(0.02)***	(0.02)***	(0.02)***	(0.02)***
Electricity intensity			0.005	0.005	0.005
			(0.003)*	(0.003)*	(0.003)*
Rural	0.08	0.08	0.08	0.04	0.08
	(0.01)***	(0.01)***	(0.01)***	(0.015)**	(0.01)***
Licensed * Rural				0.09	
				(0.02)***	
Urban	0.05	0.05	0.05	0.03	0.05
	(0.01)***	(0.01)***	(0.01)***	(0.01)**	(0.01)***
Licensed * Urban				0.04	
				(0.01)***	
Log workers	0.08	0.08	0.08	0.08	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Licensed * Log workers					0.02
					(0.006)***
Public Company	0.16	0.17	0.17	0.17	0.17
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Multiproduct	0.05	0.05	0.05	0.05	0.05
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Census Sector	0.07	0.07	0.07	0.07	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Access to Credit	0.07	0.07	0.07	0.07	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Observations	119054	119054	119054	119054	119054
R-squared	0.35	0.36	0.36	0.36	0.36

Robust standard errors in parentheses, clustered at the state-industry level. * significant at 10%; ** at 5%; *** at 1%. The regressions are a linear probability model where the dependent variable "Adoption" is equal to 1 if the firms produces some of the electricity used in production and include state * 2-digit industry * year fixed effects. "Licensed" is a dummy equalt to 1 if the 4-digit industry is licensed. "Trade tariff" is the ad-valorem import tariff (%). "Electricity intensity" is a measure of electricity needs of a 3-digit industry, as a proportion of total inputs.

Table 9: Effect of SEB Regulatory Outcomes in Licensed and Unlicensed Industries.

	(1)	(2)	(3)	(4)
	Adoption of Captive Generator			
	if Licensed = 0		if Licensed = 1	
Ind. Tariff / Avg. Tariff	0.097 (0.02)***	0.093 (0.02)***	0.093 (0.02)***	0.094 (0.02)***
T&D Losses		0.018 (0.002)***		0.016 (0.002)***
Observations	67745	67745	65441	65441
R-squared	0.17	0.20	0.21	0.26

Robust standard errors in parentheses, clustered at the state-industry level. * significant at 10%; ** at 5%; *** at 1%. The regressions are linear probability model where the dependent variable "Adoption" is equal to 1 if the firm produces some of the electricity used in production and includes 3-digit industry * year fixed effects. All tariffs and unit costs are in Paise (cents of rupee) per KWh."T&D Losses" are Transmission and Distribution losses as a % of total electricity traded.

Table 10: Product and Productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log Real Output									
					if A=0	if A=1	if L=0	if L=1	if L=0	if L=1
Licensed (L)	-0.11	-0.07	-0.17	-0.07	-0.07	-0.07				
	(0.04)**	(0.02)***	(0.04)***	(0.02)***	(0.02)***	(0.02)***				
Trade Tariff	-0.32	-0.01	-0.38	-0.01	0.01	-0.01				
	(0.10)***	(0.03)	(0.10)***	(0.03)	(0.03)	(0.03)				
Adoption of Captive Generator (A)			0.98	0.04			0.97	0.86	0.05	0.01
			(0.03)***	(0.01)***			(0.03)***	(0.03)***	(0.01)***	(0.01)
Log Unskilled Workers		0.14		0.14	0.16	0.09			0.12	0.09
		(0.01)***		(0.01)***	(0.01)***	(0.01)***			(0.01)***	(0.01)***
Log Skilled Workers		0.10		0.10	0.11	0.09			0.12	0.09
		(0.01)***		(0.01)***	(0.01)***	(0.01)***			(0.01)***	(0.01)***
Log Fixed Capital		0.04		0.04	0.04	0.04			0.04	0.02
		(0.004)***		(0.004)***	(0.004)***	(0.004)***			(0.004)***	(0.004)***
Log Materials		0.77		0.77	0.75	0.83			0.78	0.84
		(0.01)***		(0.01)***	(0.01)***	(0.01)***			(0.01)***	(0.01)***
Individual Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	96448	96448	96448	96448	62678	33770	62674	53547	59725	51689
R-squared	0.44	0.90	0.44	0.90	0.89	0.90	0.51	0.53	0.92	0.92

Robust standard errors in parentheses, clustered at the state-industry level * significant at 10%; ** at 5%; *** at 1% "Adoption" is equal to 1 if the firms produces some of the electricity used in production columns (1)-(6) include state * 2-digit industry * year fixed effects and Columns (7)-(10), state * 3-digit industry * year fixed effects "Licensed" is a dummy equalt to 1 if the 4-digit industry is licensed. "Trade tariff" is the ad-valorem import tariff (%). Individual controls are the same as used in all other tables.

Chapter 4

Literacy and Female Status in Green Revolution India

4.1 Introduction

The introduction of High Yield Varieties (HYV) during the Green Revolution in India in the mid-1960s significantly increased food production in a short period of time. In just five years, all India foodgrain production increased by almost 50% and average yields increased by around 30%. However, the heterogeneous effects associated with the introduction of the new technology generate conflicting evidence in terms of whether it brought sizeable benefits to the rural population. In particular, many authors (see, for example, Sharma and Dak (1989) for a review) argue that only relatively well-off farmers and large landowners benefited, while the many poor and small farmers, tenants or agricultural labourers might only have experienced marginal benefits at best. As the new technology was intensive in irrigation, fertilizers and mechanization, a successful introduction of the new seeds required not just an initial level of prosperity, but also an adequate level of education. That means that even if the states provided subsidised seeds, fertilizers and machinery to farmers–relaxing the financial constraint to adopt the new technology–educated farmers would still be better prepared to process information and enjoy the advantages of technical change.

In the context of the Green Revolution in India, Foster and Rosenzweig (1996) suggest that more educated farmers become aware earlier or simply were more able to manage the new technologies. They show that, in a sample of around 4,000 rural households, the likelihood of adoption of HYV seeds was significantly larger for households with at least an adult that had completed primary education, controlling for measures of wealth such as land size. Additionally, after showing that profits were greater for educated households, controlling for other inputs such as access to irrigation and machinery, the authors found that enrollment rates in primary schools increased in areas that experienced a larger increase in agricultural yields. In short, Foster and Rosenzweig show that because returns to primary education increased, private investments in schooling also increased. On the other hand, returns to education increased at a much faster pace for areas or individuals that started off relatively better, sending them into a path of more growth and higher incomes. This relationship between technical change and inequality raises the concern as to whether the increased returns to education at the micro level generated effects of any significance at the aggregate level, considering that a majority of the rural population might have been excluded from the process of technical change, where inputs were complementary to education[28].

---

[28] Arora et al.(1989) show that between 1971 and 1981, the average operational holding increased from 2.89 hectares to 3.79 hectares. In particular, small and marginal holdings (up to one hectare) were 37.63% in 1971 and 19.21% in 1981, while large holdings (between 3 and 10 hectares) increased from 26.11% to 37.01% of all holdings. The authors suggest that "small and marginal holders were leasing out or selling their lands, (...), since farming on a small scale had become an uneconomic proposition."

In this chapter, I test the relationship between HYV adoption and aggregate literacy at the district level in India, using census data for 254 Indian districts in 13 major Indian states for the census years of 1961 (pre-Green Revolution), 1971 (less than five years after its start) and 1981. The use of pre-Green Revolution data, census information and actual data on HYV adoption (as opposed to agricultural yields) are the main differences with Foster and Rosenzweig's paper. The main specification I use is similar to the one developed by Duflo (2001) to test the impact of school construction in Indonesia in schooling outcomes, i.e. a difference in difference regression where I compare the change in literacy for rural cohorts that were in primary schooling age during the Green Revolution in HYV intensive districts, with respect to the same cohort in 1961 and in districts less intensive in HYV. The identifying assumption is that there are no omitted time-varying district characteristics correlated with HYV adoption and rural literacy that would generate spurious results. To deal with this concern in the identification strategy, in addition to controlling for district and state varying characteristics, I use two control groups: older rural cohorts that would capture pre-Green Revolution trends in literacy and the same cohorts in urban areas to capture contemporaneous district-wide trends in literacy. In short, I look both at the average treatment effect on the treated (where the treatment is districts' HYV intensity) and at the average treatment effect on the untreated (older rural and urban cohorts) to test whether the treatment is picking up other unobserved phenomena that could have improved literacy even in the absence of the Green Revolution.

I find robust evidence that the increase in adoption of HYV seeds is associated with substantial increases in literacy levels. Literacy increased in all districts on average, but a district at the mean of the HYV adoption distribution would have produced an extra increase in rural literacy of around 2 percentage points per cohort. The reduced form estimation means that I cannot tell the different channels driving the increase in rural literacy apart. It could be the case that local governments built schools or hired more teachers in HYV intensive areas, increasing the supply of education. Additionally, higher returns to education could have increased demand. However, I do find evidence that when controlling for state supply of education, HYV adoption predicts greater literacy. My results suggest that a greater complementarity between economic activity and educational levels is an effective way of creating demand-side incentives for education. This can be of particular importance in rural areas, where the presistence of traditional farming might be associated with low returns to education.

I subsequently explore an additional dimension that can affect investments in education, related to the effect of the Green Revolution on gender issues and the status of women. Evidence here is somewhat conflicting. Some authors, such as Rosenzweig and Schultz (1982) and UNDP (2003), for example, show that women employment actually increased with HYV adoption, sometimes at a faster pace than men's. Others, in contrast–such as Boserup (1990), Harriss (1989), Mencher (1988) and Mazumdar and Sharma (1990)–have argued that the technological change had a pro-male bias, because of the complementarity between mechanization and male labour. If that is the case, then the advent of the new seeds should have been associated with a reduction of women in the workforce or, at least, in the agricultural sector. Displacement from the agricultural sector would also imply that returns to education for women were lower, and that educational improvements should have been lower for girls than for boys in rural areas. In the paper mentioned above, Foster and Rosenzweig (1996) found that the effect of HYV on education was significantly larger for boys than for girls in their sample. This would suggest not just that the benefits of the Green Revolution created an additional wedge in gender status in rural India, but that technical change had heterogeneous effects on education along other dimensions than initial economic prosperity. Employing a similar empirical strategy to my earlier analysis on literacy, I find no evidence of a Green Revolution-related increase in the gender gap but quite the opposite: a greater intensity in HYV is associated with a decrease in the literacy gap between boys and girls and with more women in the labour market, in the agricultural sector in particular. The interesting story the results show is that in rural areas there is an average increase in the literacy gender gap and a decrease in female labour participation. However, a more intensive level of HYV adoption mitigates these negative effects, without offsetting them completely. These findings could shed some light on the conflicting evidence regarding gender issues and the Green Revolution: research looking at a small number of districts over time might conclude that the rural gender gap has increased with and because of the Green

Revolution, without noting that the trend in most Indian districts was the same and even worse in areas with lower or no HYV adoption.

The data for this chapter are presented in Table 1. There is a great variation in the intensity of HYV adoption across districts and across time. The mean proportion of cropped area under HYV seeds was around 10% in 1971, only four years after the introduction of the new seeds and rose to above 25% ten years later. Still, some districts never introduced the new technology, while others did so very quickly and deeply. Rural literacy overall also increased over time, for both the male and female populations, even though the levels of literacy were much higher for men than for women, and the gap in literacy rates evaluated at the mean increased from 24 to 27 percentage points by 1981, compared to the pre-Green Revolution gap. This suggests a deteriorating gender gap in rural areas that is also showing in the participation of rural women in the labour force. It is striking to see that, on average, almost 80% of the women in rural areas in 1971 were not in the labour force, jumping from 66% in 1961. At first glance, the last two described patterns in the data would seem to support the idea that there was an increase in the gender gap in rural India after the start of the Green Revolution. However, I show below that districts intensive in HYV were not generating this process.

4.2 Literacy and the Green Revolution

4.2.1 Difference in differences analysis by cohort

As a first step in identifying whether the introduction of HYV brought about an increase in literacy, in this section I carry out a difference in difference analysis of the evolution of literacy before and after the start of the Green Revolution. The census data provide information on rural literacy for different age groups at different points in time which can be used to distinguish those cohorts that went through the educational age during the Green Revolution years, from those cohorts of individuals who did not. Additionally, I use information on the proportion of area that was cropped with HYV seeds at the district level to distinguish districts that were intensive in HYV (above the median) and those districts that were not.

The baseline year I use is 1961, five to six years before the first HYV seeds were introduced. For 1971, I use as the treated cohort those children between the age of 5 and 14–i.e. that were at most 10 years of age when the Green Revolution started. Table 2, Panel A shows that areas which were above the HYV adoption median in 1971 had significantly higher levels of rural literacy in 1961 (3.5 percentage points). In line with the idea that education was an important input for the successful adoption of the new seeds, this first difference suggests that the Green Revolution was introduced more rapidly in areas with better initial education. By comparing the difference in 1961 between high and low HYV intensity areas to the difference in 1971, Column (3) shows that the wedge not just persists, but experiences a significant increase, both statistically (at 1% level) and in magnitude (the difference is more than

65% larger). It is interesting to note that literacy increases over time for both groups of districts. This approach controls for group characteristics with time-invariant effects (by comparing the same group at two points in time) and common time shocks. However, this approach relies on the identifying assumption that there are no time-varying district specific effects that are correlated with literacy other than by HYV adoption, i.e. that rural literacy would not have improved more in these districts if the HYV seeds were never introduced.

To test whether the pattern of increase in literacy is driven by systematic differences across both groups of districts, I carry out two placebo experiments within the simple difference in difference framework. In the first one, I look at the literacy levels in the cohort of rural individuals aged 15 to 24–who were not exposed to primary education under the Green Revolution. Given the breakdown in age groups provided by the census data, this measure is very conservative, since many children in this group were 11 or 12 when the Green Revolution started and might have been able to improve their educational efforts at that time. In the second control experiment, I run a similar difference in difference approach by looking at the cohort of urban children aged between 5 and 14. In both cases, if regional trends were at work and were confounded with the adoption of HYV, I should observe a similar pattern of literacy increases in either older rural cohorts, in urban children or in both. The second part of Panel A and Panel B in Table 2 show that there is no systematic difference between high and low HYV areas in literacy changes for rural individuals aged 15 to

24 and for urban children aged between 5 and 14. For the former, there is a significant difference between high and low HYV intensity regions in 1961 (supporting the idea mentioned previously, that HYV was more successful in better educated areas), but this difference does not change after the start of the Green Revolution (see Column (6)). This suggests that the results obtained for the younger rural cohort were not picking up pre-existing trends in rural literacy in areas that ended up adopting more HYV seeds[29]. A difference in difference across cohorts in a given year shows that in 1971, the difference between high and low areas in the treated cohort is significantly higher than the difference for the older cohort. Interestingly, the cross-cohort difference in difference was negative in 1961 and turned positive after the start of the Green Revolution. Overall, Column (7) shows that the change between 1961 and 1971 is 2.1 percentage points (significant at the 1% level).

For the cohort of urban children aged 5 to 14, Panel B shows three interesting features. First, there is no difference in literacy before the Green Revolution in urban areas for regions high and low in HYV adoption. Second, urban literacy has increased between 1961 and 1971 by similar magnitudes to the ones I found in rural areas (around 5 or 6 percentage points). This suggests that increases in literacy were not an exclusively rural phenomenon at the time. Finally, however, this increase over time is very similar in both high and low HYV districts. In short, for children aged 5 to 14, I don't find any evidence that the increase in literacy in urban areas was more

---

[29] Unreported results show that the same pattern of no difference in high and low HYV areas holds for the group of people older than 25.

prevalent in districts intensive in HYV. In terms of the identifying assumptions, it is encouraging to find that changes that affect treated cohorts are not found in non-treated cohorts.

In Table 3, I carry out a similar exercise, this time comparing data for 1961 and 1981. This time, the cohort that grew up after the Green Revolution started includes those aged between 5 and 24 in 1981, and I compare them with the same group in 1961. Districts are again sorted according to whether the proportion of HYV adoption is above or below the median, measured in 1981. Panel A shows that the difference in rural literacy for those aged between 5 and 24 between both regions is 2.9 percentage points greater (at the 1% significance level) in 1981 than in 1961.

As in the previous table, I look for systematic differences across regions by testing whether the same difference in difference approach can explain the evolution of two groups that are not treated under my identifying assumptions: i.e. rural populations aged over 25 and urban populations aged below 25. Results in Panels A and B in Table 3 show that the increasing levels of rural literacy found for the treated group are not present for the other two control groups. The difference in literacy between high and low HYV areas for the group that had their educational years after the start of the Green Revolution is significantly greater than the difference in literacy for the same group in 1961, the older age group in 1981 and the same age group in urban areas[30].

---

[30] The results hold when comparing the 15 to 24 age group in 1971 (not treated) and 1981 (treated), showing that the change in literacy in HYV intensive districts between 1971 and 1981 is on average 1.5 percentage points greater (significant at 1% level) than in low HYV areas. Older rural and urban

This section has provided evidence that rural literacy significantly improved in HYV intensive areas. The claim that the difference in difference results are identifying causal effects seems more robust when results for untreated cohorts in rural and urban areas do not follow the same pattern, alleviating concerns that the results are driven by time-varying regional characteristics. In the following section, I extend the analysis to use a continuous measure of HYV adoption and to control for district-specific and time-varying characteristics that could be correlated with HYV adoption and literacy and might be biasing the difference in difference results in Tables 2 and 3.

4.2.2 Effect of HYV adoption by cohort

In this section, I turn to a regression framework to introduce a continuous measure of HYV adoption and control for district characteristics that might be driving the surge in rural literacy. I first consider the average effect on literacy of the Green Revolution by running an OLS regression of the following form

$$L^{cl}_{dst} = \alpha_{ds} + \beta_t + \gamma HYV_{dst} + \delta \mathbf{X}_{dst} + \varepsilon_{dst} \tag{4.1}$$

where L^{cl}_{dst} is the proportion of literate people in district d, state s and year t for a cohort c in location l (i.e. rural or urban), and HYV_{dst} is the time varying proportion of cropped area cultivated with HYV seeds at the district level. I also include district fixed effects to control for time-invariant district characteristics and year dummies

---

cohorts don't show any significant difference.

to control for average changes in literacy between census years. I also include a set of district controls $\mathbf{X}_{dst}$ that capture demographic changes (e.g. proportion of female population, proportion of rural population, log of population) and economic characteristics (e.g. proportion of rural workers in the manufacturing sector, log of roads per square kilometre) that could be correlated with both HYV adoption and literacy. To account for autocorrelation of the error term, I cluster standard errors in all regressions at the district level.

Even though the previous section has shown that the successful adoption of HYV seeds was more prevalent in areas with higher literacy, these estimates are not capturing this reverse causality. The fact that the variable HYV_{dst} is only positive and different from 0 from 1971, implies that district fixed effects are absorbing the differences in initial levels of literacy across districts and that the estimated coefficient γ only captures the additional improvement in literacy associated with increases in HYV. Note that the estimation is a reduced form regression, without being able to delineate specific mechanisms, such as an increase in demand for education or a Green Revolution-induced increase in the supply of schools in high HYV areas. As suggested by Foster and Rosenzweig (1996), both likely happened to some extent. However, the main econometric concern would be that, in the absence of technological change, these districts would have experienced supply or demand shocks at that time specific to rural areas. In the absence of a clean counterfactual, the use of a continuous measure of HYV adoption allows me to compare districts with different treatment intensity over time, controlling for district fixed effects and time-varying characteristics. Additionally, running similar specifications for non-treated cohorts in rural and urban areas allows me to check whether the literacy changes were happening only at the right time and in the expected areas.

Panel A in Table 4 shows results for three cohorts, with and without district controls in each case. In columns (1) and (2), increases in HYV adoption are positively associated with greater literacy rates for the group of rural children aged 5 to 14. The inclusion of district controls reduces the magnitude of the coefficient, but not its significance. Additionally, the year dummies are positive and significant, capturing an average improvement in rural literacy among the youngest cohort. The magnitude is sizeable: an increase in one standard deviation in HYV results in around 1.4 percentage points increase in rural literacy for children aged 5 to 14. Taken at the average district in the sample, that means having almost 5400 more literate children in that cohort. For the whole sample, this is equivalent to an increase in the number of literate children aged 5 to 14 by an extra 1.4 million, as a result of the introduction of HYV seeds.

It is important to note that the cohort of children aged 5 to 14 were of educational age in both HYV years (1971 and 1981). That means that I expect the effect of HYV on this cohort to be stronger than for the cohorts analysed in the subsequent columns. In columns (3) and (4), I run regression (4.1) using the cohort of individuals in rural areas aged 15 to 24. Given that the Green Revolution started in 1966/7, I expect many people in the 1971 cohort to be beyond educational age, probably in the labour market already. However, all individuals in the 1981 cohort were born (and then were eligible to receive primary education) after the start of the Green Revolution. Because the cohort in 1971 includes many people that were not treated, I expect the effect of HYV on literacy to be smaller than for the youngest cohort. Results in Columns (3) and (4) show that this is the case, with coefficients remaining strongly significant. When pooling both cohorts together in columns (5) and (6) and running on rural literacy for individuals aged 5 to 24, I include all treated individuals (aged 5 to 14 in 1971 and aged 5 to 24 in 1981) and some that I expect not to be treated (aged 15 to 24 in 1971). The estimates are positive and strongly significant (at the 1% level) and, as expected, the magnitude of the coefficients is smaller than the estimates in columns (1) and (2), but greater than in columns (3) and (4).

Panel B in Table 4 looks at the correlation between literacy and some district controls to shed some light on some of the drivers of rural literacy. The proportion of population that is female is strongly correlated with higher levels of literacy while more rural districts don't seem to perform worse. However, more densely populated areas do have lower literacy rates across cohorts. As expected, measures of rural economic development and infrastructure–such as the proportion of rural workers in manufacturing jobs and road density–are both positively correlated with higher levels of literacy.

In Table 5, I reproduce regression (4.1) but this time using the control groups, as in the previous section. As in Table 4, I include regressions with and without district controls. In columns (1) and (2) the left hand side variable is the rate of literacy for the rural cohort of individual that was of educational age before the start of the Green Revolution, i.e. those aged 25 or more. As expected, the coefficient on HYV adoption is not significantly different from 0. The year dummies are positive and significant, suggesting an increasing improvement in rural literacy across generations that predates independence, but that was not associated with the Green Revolution. Columns (3) to (6) use the groups aged 5 to 14 and 5 to 24 in urban areas to see whether districts' intensity in HYV is correlated with urban literacy among the cohorts that were of educational age after the introduction of HYV seeds. A positive and significant coefficient would imply that HYV adoption was correlated with omitted district developments which were driving literacy up in both rural and urban areas and that the results are capturing a spurious correlation between HYV adoption and rural literacy. The failure to produce statistically significant coefficients for urban literacy alleviates these concerns somewhat.

Next, I combine the baseline OLS regression and the difference in difference approach to look at the additional effect per year of HYV adoption with respect to 1961 in a regression of the following form:

$$L^{cl}_{dst} = \alpha_{ds} + \beta_t + \gamma_{71} HYV_{ds71} + \gamma_{81} HYV_{ds81} + \delta \mathbf{X}_{dst} + \varepsilon_{dst} \qquad (4.2)$$

where γ_{71} and γ_{81} respectively capture the differential increase in literacy associated with higher HYV intensity in 1971 and 1981 with respect to 1961. The previous OLS specification provided information on the average effect of HYV on literacy. Equation (4.2) allows me to deal with at least three concerns. The first, related to the magnitude of the effect of the Green Revolution on rural literacy, is to deal with the possibility that in order to get sizeable improvements in literacy, the adoption of HYV seeds should increase beyond the bounds of the data at a point in time. I can check that by having a measure of the effect per year. A second concern is that a cohort that was treated shows significant results because the same cohort in another year is pulling the average up. For example, if all literacy improvements for rural children aged 5 to 14 are significant only in 1981, and insignificant or negative in 1971, this could raise doubts as to whether the improvements came from other sources than HYV adoption, which only became active many years after the start of the Green Revolution. Similarly, this specification shows whether a cohort that was not supposed to be treated (e.g. individuals aged 15 to 24 in 1971) produces significant results. A third concern is that the effects observed in the OLS regression were short-lived, thus capturing a pre-Green Revolution trend in districts that ended up adopting more HYV, rather than its effects. If that is the case, the differential increase in rural literacy associated with HYV intensity with respect to 1961 could be greater in 1971 than in 1981 (where HYV adoption was greater). In short, since HYV adoption increased steadily over time, finding an unbalanced distribution of the effects of HYV on rural literacy over time might raise some concerns about the described channel: specifically, it might be revealing that the measure of HYV in a given year is capturing some other omitted time-varying district-specific effects that are driving rural literacy up.

Table 6 shows results for the same three cohorts analysed earlier, with and without district controls. Columns (1) and (2) show that for rural children aged 5 to 14, a greater proportion of HYV seeds is associated with an additional increase in rural literacy over and above the average increase captured by the year dummies. As expected, the effect holds both in 1971 and 1981, meaning that the effects of the Green Revolution on literacy are present from the beginning. In terms of magnitude, a district at the mean of HYV adoption in 1971 has a 1.1 percentage points extra increase in rural literacy (on top of the average increase of 3.7 percentage points) with respect to 1961. When comparing 1981 with 1961, a district at the mean of the adoption distribution has increased between 2.3 and 3.2 percentage points more than the average increase of 9 percentage points. Note that even though the point estimates are similar, the fact that HYV was more prevalent in 1981 increases the mean effect and alleviates the concern that the effects on literacy were not distributed over time in a similar way to HYV adoption. Columns (3) and (4) show interesting results regarding the cohort of rural individuals aged between 15 and 24. Since most of the individuals in this cohort were not supposed to be treated in 1971, I expect the effects to be insignificant or small at best. This is what the results show: I find no effect in the specification without district controls, and I find a coefficient significant at the 10% level only when I control for district characteristics. If present, the magnitude of the effect is low (0.5 percentage points, less than half the effect found for the younger cohort). This contrasts with the positive, sizeable (around 2 percentage points) and significant (at the 1% level) effect found for this cohort in 1981. These results are good news in terms of the identifying assumptions, since the effects on rural literacy seem to come not just from districts that were more intensive in HYV, but also from the cohorts that were supposed to be affected by the Green Revolution. Columns (5) and (6) show that the results hold when pooling together the two cohorts, i.e. individuals aged 5 to 24, and that the magnitudes are very similar to those found in the first two columns.

The same exercise is done for control groups, namely rural people that were in schooling age before the Green Revolution and younger urban cohorts. A concern with respect to the OLS results for these control groups might be (again) that OLS is averaging out some opposing effects that might happen in HYV-intensive districts over time, revealing some unobserved phenomena that have affected different cohorts at different points in time. If during the first years of the Green Revolution, adopting districts were drawing resources from urban areas to serve rural areas, then we should expect intensity in HYV to be associated with a drop–or less than average increases– in urban literacy. Alternatively, resources generated by HYV in the late stages might have been used to fund a district-wide schooling programme that would affect both urban and rural children. If stories of this sort happened over time, it could be the case that an OLS estimation would show an insignificant effect of HYV on urban areas, but a difference in differences approach might show a positive effect for 1971, and a negative effect for 1981, for example. Results in Table 7 alleviate the concern that a more complex non-linear story was generating insignificant results in urban or older rural cohorts. Even though the year dummies capture increasing literacy in all groups, there is no strong evidence that intensity in HYV was associated with heterogeneous effects for the untreated population at any point in time.

Finally, I check whether the estimates are confounding the effects of other phenomena that could be generating a spurious correlation between HYV adoption and rural literacy[31]. The first alternative story is one related to migration: the initial difference in difference tables showed that HYV adoption was more successful in areas with greater levels of literacy. It might well be the case that, as a result, areas suitable for HYV adoption were attracting more literate farmers. If that were the case, the increase in average rural literacy would not be the consequence of increased returns to education generating incentives to younger generations to get more education, but instead the consequence of a selection of better educated migrants arriving in the HYV-compatible districts. Column (1) in Table 8 controls for the interaction of the proportion of rural migrants in the district with year dummies. The estimates for

---

[31] I present results for the cohort of children aged 5 to 14, even though the following results hold for the other cohorts as well.

HYV adoption in 1971 and 1981 remain positive, significant and similar in magnitude, suggesting that HYV was not capturing the effects of migration.

Another concern is that HYV adoption was correlated with the proportion of population belonging to Scheduled Castes (SC) or Scheduled Tribes (ST). Dhanagare (1989) points out that the benefits for these groups were smaller, in particular because it was poorer farmers that on average benefited less from the introduction of agricultural change. The story I am testing here is one where the evolution of literacy in a district is linked to the economic and social composition of the population, rather than to HYV. The presence of a poorer or more disenfranchised population could be correlated both with less HYV adoption and with a deteriorating trajectory for rural literacy. Columns (2) and (3) check whether the effect of HYV vanishes when controlling for the proportion of population belonging to either group, interacted with year dummies. Since the proportion of people belonging to either group in a district tends to be very stable, the interaction with year dummies will capture the time evolution of rural literacy for a level of SC/ST population. The effects of HYV remain positive and significant, even though the magnitude drops a little, in particular when the proportion of population belonging to Scheduled Tribes increases[32].

Finally, the increase in literacy might come as a consequence of improved provision of education supplied by the state government and unrelated to the adoption of

---

[32] Interestingly, the unreported coefficients on the interaction of the ST measure with year dummies are negative and significant. This implies that the upward bias comes from a negative correlation between successful adoption of HYV and more population classified as Scheduled Tribe, as suggested by the anecdotal evidence.

the new technology. In column (4), I include as a control the log number of schools per capita at the state level[33]. In Column (5), I also include state time-trends to check whether changes in literacy were actually driven by state-wide policy trajectories. In both cases, results on the interaction terms are positive and significant, and the magnitude of the coefficients does not drop much. These specifications reduce concerns that literacy improvements were not driven by HYV adoption but by other district or state phenomena that would have improved rural literacy in the absence of the Green Revolution.

4.3 Gender, literacy and the Green Revolution

In this section, I analyse whether the improved literacy associated with HYV intensity had heterogeneous effects by gender. It is generally argued that women lost out in the process of technical change associated with the Green Revolution in India, either by being displaced from wage earning opportunities or by receiving even lower wages than before (see FAO (1997), for example). Theoretically, if the new technology was a substitute rather than a complement to female labour, I should observe fewer women employed in rural areas or, if employed, that they moved away from the agricultural sector. Unfortunately, the data available does not allow me to investigate whether women with particular characteristics (e.g. landowners or from wealthier households) benefited more. But the data do allow me to investigate if the

---

[33] I alternate this measure with the log of expenditure in primary education per capita and the log of primary teachers per capita, without major changes in the results.

claims that female labour in general–and in agriculture in particular–was displaced in areas intensive in HYV, hold on aggregate. The existing evidence is not conclusive about whether women were displaced by the introduction of technical change (as suggested by Harriss (1989) or Mencher (1988)) or whether female employment actually increased (as in UNDP (2003) or Rosenzweig and Schultz (1982), for example). Kaur and Sharma (1989) suggest that there are heterogeneous effects from mechanization for women based on socioeconomic characteristics. They find that the overall effect is a displacement of female labour in rural areas in two districts in the state of Haryana.

Additionally, if HYV reduced female opportunities in rural areas and their wages were depressed, relative to male wages it follows that their returns to education should have been lower as well. If so, I should find that the interactions of HYV and year dummies in equation (4.2) are negative or insignificant when the explained variable is rural female literacy for a treated cohort. After checking labour market outcomes for women, I test whether improvements in literacy were gender specific by looking separately at the effects on gender cohorts and the gap between male and female literacy.

4.3.1 Labour market and gender

The effect of the Green Revolution on female labour could theoretically go either way. A simple mechanism for increasing female labour is that the new technology increased female wages and raised the opportunity cost of not working. However, if the technology was intensive in inputs that were not complementary to female labour, then women's wages would have dropped. Unless rural women were in a backward-bending part of the labour supply schedule, a drop in wages should have reduced the supply of female labour. However, other mechanisms might have been at play if the decision to work was made at the household level or depended on other characteristics, such as land ownership. For example, if HYV adoption increased the income of male household members, women might have stopped working altogether–or moved away from the agricultural sector. On the other hand, the adoption of the new seeds by the household might only have been possible if women raised some extra cash to buy seeds, fertilizers, etc. The district data do not provide enough detail to tell these alternative mechanisms apart. However, a reduced form regression of employment outcomes on HYV adoption might be informative in terms of labour market outcomes, particularly since the data set includes districts with little or no HYV adoption. To that end, I use information on the proportion of women not working and, among those who did work, the proportion employed in agriculture as a proportion of total rural females and working females.

In column (1), Table 9, the year dummies in the OLS regression show that the proportion of women that were not part of the working force in rural areas was greater in 1971 than in 1961, but not in 1981. However, a greater level of HYV adoption was associated with a reduction in the proportion of women that did not participate in the labour market. Note that the proportion of HYV adoption that offsets completely the effect captured by the year dummy should be above 1. That means that HYV adoption could at best mitigate average changes in female employment. This suggests that following a district intensive in HYV before and after the introduction of the new technology would show that female employment dropped but would not reveal what is apparent from results in column (1): that the effect was actually lower than it would have been in the absence of the new technology.

Column (2) looks at the interaction of HYV adoption with year dummies and finds that a greater part of the effect of HYV adoption in column (1) happened by 1971. However, even though the 1981 dummy is still insignificant, intensity in HYV was associated with a slight increase in female participation in the labour force. The magnitude of the coefficients and the distribution of HYV adoption per year suggest that, with the exception of a small number of districts, female participation in labour markets dropped significantly between 1961 and 1971, but the effect was increasingly mitigated by a greater adoption of HYV seeds. By 1981, the proportion of women not working fell again, but the increase in female labour (when compared to 1961 levels) came from districts with greater adoption of HYV only.

In columns (3) to (6) I look at the participation of women in the agricultural sector. When measured as a proportion of the total female population, columns (3) and (4) show that intensity in HYV attracted more women into the agricultural sector. The HYV coefficient in column (3) is positive and significant and column (4) shows that the effect was stronger in 1971. It is apparent again that HYV adoption mitigated the decrease in the proportion of women in the agricultural sector, as captured by the 1971 dummy. Not including a proper counterfactual (i.e. evolution of districts with less HYV or without HYV altogether) might lead to the conclusion that HYV was associated with a displacement of female force from agricultural labour. Columns (5) and (6) look at the proportion of rural women in agriculture, conditional on being in the labour force, to check the sectorial allocation of female workers. Over time, there seems to have been almost no significant movement towards or away from agriculture among female workers. However, an increase in HYV adoption was strongly associated with women working in the agricultural sector. This time, the effects are not just strongly significant but also within the bounds of the data: the increase in the proportion of female workers in the agricultural sector was 2.3 percentage points higher for a district at the mean of the HYV adoption distribution in 1971, and 5.6 percentage points higher in 1981 than in a district with no HYV.

4.3.2 Literacy

Table 9 provides evidence that the Green Revolution not just attracted women to the labour market but to the agricultural sector, in particular. To test whether the increase in female employment was associated with increasing returns to education for females, in the absence of information on wages for female agricultural workers, I can only run a reduced form regression and check whether literacy among females increased with HYV. If an increase in the labour market gap between males and females resulting from the Green Revolution was not captured by Table 9, but was present and affected the returns to education, then it should be reflected in the evolution of the literacy rates by gender.

To analyse whether the literacy results found in previous sections were gender specific, I again ran regression (4.1), this time for men and women separately. Information on literacy rates per age groups and per sex allows me to test the hypothesis that the Green Revolution induced a gender bias in educational outcomes. Columns (1) and (2), Table 10 present results for the OLS estimation of HYV adoption on literacy for rural individuals aged 5 to 24 for males and females, respectively. An increase in 1 percentage point in the proportion of land cropped with HYV seeds was associated with an increase in rural literacy of 0.06 percentage points for males and 0.10 percentage points for females. As the average effect obtained in the OLS estimate for this cohort was 0.08, it follows that there were heterogeneous effects in rural literacy associated with HYV adoption, but, if anything, in a pro-female direction. To check whether this differential impact in literacy came from gender specific trends, I will again use the two control groups (rural population older than 25 and the population aged 5 to 24 in urban areas). The average effect in these groups was not significantly different from zero in the OLS estimation. But it could be the case that female literacy was following a different trajectory in rural and urban areas at the district level (through a catch up effect, for example, since rural male literacy was more than two times greater than female literacy, on average, or through a district-wide programme that also reached urban girls). So it could be that column (2) was simply reflecting these effects. Columns (3) and (4) show that for the older rural cohort HYV was associated with changes in literacy that are insignificantly different from zero. Similarly, results for the urban cohort in columns (5) and (6) are also insignificant. In both cases, the results on HYV in the first two columns do not seem to be picking up a gender-specific rural or generational effect.

As pointed out before, looking at the average effect of HYV on rural literacy might be hiding some heterogeneous gender effects behind the positive and significant estimates within cohorts. The steady increase in HYV adoption over time should be reflected in the year specific effects in order to reduce concerns that the measure of HYV adoption is correlated with year and region-specific shocks. To test this, I ran a set of difference in difference regressions as in equation (4.2), but this time for male and female cohorts, to see whether the gender effect on literacy found in the OLS specification followed the deepening of the Green Revolution. For the cohort of rural individuals aged between 5 and 24, the point estimates confirm that the effect of HYV was stronger in 1981 than it was in 1971, for both men and women. In particular, intensity with HYV was associated with a larger and increasing effect on female literacy. In a district with a mean level of HYV adoption, the increase in rural literacy was similar for men and women in 1971 (i.e. 0.8 and 1 percentage points, respectively), but larger for females in 1981 (1.6 percentage points for males, 2.6 for females)[34]. When looking at the control groups, columns (3) to (6) show no results that could suggest that omitted characteristics at the district levels were driving the general improvement in literacy or the relatively improved performance of female literacy in rural or urban areas.

The positive and significant coefficients for year dummies across generations and locations show that literacy did improve for men and women. Moreover, improvements above the average in areas with higher HYV intensity only happened for the cohorts that were supposed to be affected by the change, i.e. individuals that were educated after the start of the Green Revolution in rural areas. Additionally, the improvements in literacy seem to be evenly distributed between men and women; if anything, with a tilt towards larger benefits for girls. In the next set of regressions, I check whether these differences are statistically significant, by running a regression of the form:

$$\Delta L^{cl}_{dst} = \alpha_{ds} + \beta_t + \gamma_{71} HYV_{ds71} + \gamma_{81} HYV_{ds81} + \delta \mathbf{X}_{dst} + \varepsilon_{dst} \tag{4.3}$$

where ΔL^{cl}_{dst} is the literacy gap between men and women of cohort c in location l for a district d in state s and year t. The difference should cancel out all district characteristics that had a similar effect on male and female literacy (e.g. teacher availability or quality), but would not deal with characteristics that had a differential impact on literacy, even if these characteristics were fixed (e.g. cultural gender bias). To account for this possibility, I use district fixed effects that would take care of

---

[34] Unreported estimates show a very similar pattern for the cohort of rural children between 5 and 14, suggesting that the choice of cohort is not relevant in identifying gender effects.

gender-varying effects of district fixed characteristics. Similarly, year dummies will capture the change in the gender gap with respect to 1961. Additionally, I replace the left hand side variable with the difference between the gender gap in rural areas with the gender gap in urban areas, to account for the possibility that the adoption of HYV seeds might have been correlated with some unobserved district-wide changes in gender attitudes.

The positive and significant estimates for the year dummies in column (1) in Table 12 suggest that the rural literacy gap increased between males and females aged between 5 and 24. However, the gender gap increase was mitigated in districts with more HYV adoption, i.e. the increase in the literacy gap was lower, the greater the incidence of HYV seeds. Only in a few cases was the level of HYV adoption large enough to offset the overall increase in the literacy gap. Column (2) shows the gender gap in literacy for the same cohort in urban areas. The year dummies show that, on average, the gap decreased with respect to 1961. If anything, in 1971 the gap decreased slightly less in districts more intensive in HYV, even though the magnitude of the coefficients is small and a reversion in the decrease captured by the year dummy is beyond the bounds of the HYV adoption distribution. Columns (3) and (4) show results for the rural and urban cohort of people aged over 25 and prove interesting: as in the previous two columns, the gender gap decreased in urban areas and increased in rural areas. Most notably, districts intensive in HYV showed a reduction in the gender literacy gap. This would be bad news in terms of the identification strategy, since there was an effect in HYV-intensive districts in a cohort that should not have been affected. However, column (4) shows that a similar phenomena was happening in urban areas, suggesting that districts intensive in HYV were closing the gender literacy gap in rural and urban areas at the district level, probably capturing pre-Green Revolution trajectories. The next four regressions use the difference between rural and urban gender gaps as the explained variable, with the objective of absorbing common trends in gender gap literacy in urban and rural areas. In columns (5) and (7) I run an OLS regression using HYV intensity as the explanatory variable and in columns (6) and (8), I run regression (4.3). Column (5) shows that the gender gap in rural areas increased over time vis-a-vis urban areas (as shown by the point estimates of the year dummies), but the coefficient on HYV adoption is negative and significant. This means that the difference between the rural and urban gender gaps decreased with the Green Revolution. Column (7) shows that this was not the case for the older cohort which grew up before the Green Revolution started. Columns (6) and (8) show the same pattern: the relative increase in the gender gap in rural areas was mitigated in areas with greater HYV adoption for the cohort that was treated, i.e. of educational age after the start of the Green Revolution, and not for the older cohorts.

In conclusion, results in this section suggest not just that the intensity of the Green Revolution was associated with an increase in literacy in rural areas for both men and women, but that the effect was significantly stronger for women. The evidence suggests that HYV adoption actually mitigated the increasing gap between male and female rural literacy. This observation is at odds with the widespread belief that the Green Revolution worsened the status of women, at least in terms of access to education (see FAO report on "Women and Food Security" (1997) and Mazumdar and Sharma (1989)). Since the gender literacy gap in rural areas seems to have worsened over time, looking at the evolution of districts intensive in HYV only might lead to the conclusion that the increase in HYV was driving this phenomenon. However, by looking at a large sample of Indian districts over time and using control groups in rural and urban areas, I can distinguish the general direction of gender differences in rural areas and the performance associated with a deeper incidence of the Green Revolution across India.

4.4 Conclusions

The economic and social changes in rural areas associated with the introduction of high yield seeds in mid-1960s in India–known as the Green Revolution–still generate conflicting evidence and controversy. The overarching concern is that in the presence of strong heterogeneous effects, only the more prosperous or the more educated farmers benefited from the introduction of a new technology and many remained excluded. In particular, the prerequisites for successful adoption of the new technology, such as mechanization, irrigation and fertilizers, seemed to be very restrictive for a country with more than 70% rural population and high levels of poverty, landlessness and illiteracy.

This paper shows that, despite the possibility of heterogeneous effects within districts, the Green Revolution was associated with sizeable aggregate effects on rural literacy at the district level. Even though pre-HYV means show that adoption was more prevalent in districts with higher levels of rural literacy, the effects identified capture changes over and above the average increase in rural literacy experienced between census years. A district at the mean of the distribution of HYV intensity experienced an extra increase in rural literacy of around 1 percentage point in 1971 and more than 2 percentage points in 1981 with respect to pre-Green Revolution levels. Some identification concerns regarding the link between literacy and HYV adoption are alleviated by the lack of evidence that cohorts whose educational decisions should not have been affected by the level of HYV intensity experienced the same changes in literacy. I find no evidence of pre-Green Revolution trends in rural areas (as captured by older rural cohorts) or simultaneous district wide changes (as captured by the cohort of young urban individuals). Additionally, the identifying assumption–i.e. that these changes would not have happened by other means in the absence of the Green Revolution–is robust to the inclusion of time-varying district characteristics, state investment in primary schools and time trends.

The concerns that the female population might have been left out of the benefits of technological change are not supported by the data in terms of literacy rates and labour participation, on aggregate. Even though there is evidence that the average number of rural women excluded from labour markets increased after the Green Revolution, I find that this phenomenon was less prevalent, the greater the intensity in HYV. Similarly, the gender gap in rural literacy rates increased over time since the introduction of HYV seeds, but the gap widened less in districts that adopted more HYV seeds. Without dismissing the possibility of heterogeneous effects among rural women, the aggregate data presents a picture that reconciles the conflicting evidence around the status of women after the start of the Green Revolution.

In many cases, the provision of infrastructure such as schools and teachers may be the leading constraint to education and its provision could spur increases in educational levels (see Duflo (2001), for example). But in other cases, low education may be the natural consequence of economic activities with low returns to education, such as traditional farming. The improvement in literacy for men and women and the reduction of the gender gap associated with the adoption of a new agricultural technology in rural areas suggest that improvements in human capital can be obtained in rural areas, if the returns to education increase through technological change.

4.A Appendix A: Data

The census data for years 1961-1981 comes from the Indian District Database prepared by R.Vanneman and D.Barnes (2000), Indian District Data, 1961-1991: data file and codebook, College Park, Maryland: Center on Population, Gender, and Social Inequality, available at www.inform.umd.edu/~districts/index.html. I merged the data with the India Agriculture and Climate Data set, prepared by Apurva Sanghi, K.S. Kavi Kumar and James W. McKinsey, Jr. for the World Bank (available at chd.ucla.edu/dev_data/datafiles/india_agric_climate.htm), accounting for changes in districts' boundaries that left me with 254 districts in 13 states (Andhra Pradesh, Bihar, Gujarat, Haryana, Karnataka, Maharashtra, Madhya Pradesh, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal).

"HYV adoption" is the proportion of area under HYV seeds and "HYV intensity" in the difference in difference tables is a dummy equal to 1 for all districts with HYV adoption above the median in a particular year. "Rural literacy" for all age groups, gender and location is the proportion of literate people within that group. "Female population" and "Rural population" are the proportion of women and individuals living in rural areas in total population, respectively. "Rural workers in manufactures" is the proportion of main workers in rural areas working in the manufacturing sector, and "Log Roads" is the log of roads per square kilometre. "Rural migration" is the proportion of population in a district that has come to rural areas from other districts. "Scheduled Caste" and "Scheduled Tribe" are the proportion of population that belongs to either group. I also include "Primary schools", which is the log of primary schools per capita at the state level, and I alternate it with the log of number of primary school teachers per capita at the state level and the log of state expenditure on primary schools per capita. Finally, "Rural women not workers" is the proportion of women that are not employed in any main or marginal activity. "Rural women in agriculture" is the number of women working in the agricultural sector taken either as a proportion of rural female or of rural female workers.

Table 1: Summary Statistics

Variable	Year	Mean	Std. Dev.	Min	Max
HYV Adoption (%)	1961	0	0	0	0
	1971	0.102	0.111	0	0.678
	1981	0.255	0.160	0	0.767
Rural Literacy (%)	1961	0.196	0.075	0.044	0.547
	1971	0.248	0.094	0.062	0.651
	1981	0.308	0.108	0.085	0.699
Male Rural Literacy (%)	1961	0.310	0.099	0.070	0.637
	1971	0.366	0.111	0.096	0.720
	1981	0.439	0.118	0.132	0.752
Female Rural Literacy (%)	1961	0.077	0.059	0.009	0.456
	1971	0.122	0.085	0.014	0.580
	1981	0.168	0.109	0.019	0.645
Rural Women Not Workers (%)	1961	0.655	0.154	0.394	0.979
	1971	0.798	0.117	0.445	0.991
	1981	0.750	0.143	0.472	0.994

Table 2: Mean Rural Literacy by Cohort and Year (1961-1971)

Panel A:	Rural Literacy							Panel B:	Urban Literacy		
	Aged 5 to 14			Aged 15 to 24					Aged 5 to 14		
	HYV intensity (in 1971)			HYV intensity			Difference in difference across cohorts		HYV intensity (in 1971)		
	High	Low	Difference	High	Low	Difference			High	Low	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7) = (3) - (6)		(1)	(2)	(3)
1971	0.299 (0.010)	0.241 (0.008)	0.058 (0.013)***	0.385 (0.011)	0.336 (0.011)	0.049 (0.016)***	0.009 (0.005)**	1971	0.550 (0.010)	0.548 (0.008)	0.002 (0.011)
1961	0.239 (0.008)	0.204 (0.007)	0.035 (0.010)***	0.273 (0.008)	0.226 (0.008)	0.047 (0.012)***	-0.012 (0.003)***	1961	0.489 (0.008)	0.495 (0.007)	-0.005 (0.011)
Difference	0.060 (0.013)***	0.037 (0.004)***	0.023 (0.006)***	0.112 (0.014)***	0.110 (0.014)***	0.002 (0.006)	0.021 (0.004)***	Difference	0.061 (0.012)***	0.053 (0.010)***	0.007 (0.005)

Robust standard errors in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV intensity" is the net proportion of area cropped with HYV seeds. "High" is for districts above the median and "low" for those below. "Rural Literacy" and "Urban literacy" are the proportion of literate people in rural and urban areas, respectively, for different age groups.

Table 3: Mean Rural Literacy by Cohort and Year (1961-1981)

Panel A:	Rural Literacy							Panel B:	Urban Literacy		
	Aged 5 to 24			Aged 25 to 59					Aged 5 to 24		
	HYV intensity (in 1981)			HYV intensity			Difference in difference across cohorts		HYV intensity (in 1981)		
	High	Low	Difference	High	Low	Difference			High	Low	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7) = (3) - (6)		(1)	(2)	(3)
1981	0.403	0.336	0.066	0.252	0.232	0.020	0.046	1981	0.643	0.650	-0.007
	(0.012)	(0.011)	(0.016)***	(0.008)	(0.007)	(0.011)*	(0.008)***		(0.010)	(0.008)	(0.011)
1961	0.250	0.213	0.037	0.169	0.154	0.015	0.022	1961	0.536	0.535	0.001
	(0.008)	(0.008)	(0.011)***	(0.005)	(0.006)	(0.008)*	(0.004)***		(0.008)	(0.007)	(0.010)
Difference	0.153	0.123	0.029	0.084	0.079	0.005	0.024	Difference	0.108	0.115	-0.008
	(0.014)***	(0.013)***	(0.008)***	(0.009)***	(0.009)***	(0.004)	(0.006)***		(0.012)***	(0.010)***	(0.006)

Robust standard errors in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV intensity" is the net proportion of area cropped with HYV seeds. "High" is for districts above the median and "low" for those below. "Rural Literacy" and "Urban literacy" are the proportion of literate people in rural and urban areas, respectively, for different age groups.

Table 4: Literacy and HYV Adoption (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: HYV Adoption and Year Effects					
	Rural Literacy					
	Aged 5 to 14		Aged 15 to 24		Aged 5 to 24	
HYV Adoption	0.121 (0.029)***	0.093 (0.026)***	0.069 (0.026)***	0.063 (0.024)***	0.105 (0.028)***	0.084 (0.025)***
1971	0.035 (0.004)***	0.061 (0.013)***	0.104 (0.005)***	0.109 (0.013)***	0.059 (0.004)***	0.077 (0.012)***
1981	0.091 (0.009)***	0.125 (0.025)***	0.144 (0.008)***	0.149 (0.025)***	0.110 (0.008)***	0.134 (0.024)***
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	No	Yes	No	Yes	No	Yes
Observations	762	762	762	762	762	762
R-squared	0.95	0.96	0.96	0.97	0.96	0.97
	Panel B: Controls					
Female Population (%)		1.49 (0.55)***		1.52 (0.54)***		1.56 (0.52)***
Rural Population (%)		-0.02 (0.12)		-0.03 (0.12)		-0.02 (0.12)
Log Population		-0.17 (0.06)***		-0.15 (0.06)***		-0.16 (0.06)***
Rural workers in Manufacturing (%)		1.76 (0.25)***		1.63 (0.23)***		1.75 (0.23)***
Log Roads		0.03 (0.006)***		0.02 (0.008)**		0.03 (0.006)***

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV intensity" is the net proportion of area cropped with HYV seeds at the district level. "Rural literacy" is the proportion of literate people in rural areas for different age groups. "Female Population" and "Rural Population" are the proportion of women and individuals living in rural areas in total population, respectively. "Rural Workers in Manufacturing" is the proportion of main workers in rural areas working in the manufacturing sector and "Log Roads" is the log of roads per square kilometre.

Table 5: Literacy and HYV Adoption in Untreated Cohorts

	(1)	(2)	(3)	(4)	(5)	(6)
	Rural Literacy		Urban Literacy			
	Aged 25 to 59		Aged 5 to 14		Aged 5 to 24	
HYV Adoption	0.007	-0.0003	0.027	0.011	-0.019	-0.028
	(0.012)	(0.012)	(0.021)	(0.021)	(0.018)	(0.021)
1971	0.031	0.035	0.054	0.062	0.074	0.081
	(0.002)***	(0.007)***	(0.004)***	(0.011)***	(0.003)***	(0.010)***
1981	0.079	0.084	0.095	0.100	0.117	0.122
	(0.004)***	(0.014)***	(0.007)***	(0.022)***	(0.005)***	(0.020)***
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	No	Yes	No	Yes	No	Yes
Observations	762	762	762	762	762	762
R-squared	0.97	0.97	0.93	0.94	0.95	0.96

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV intensity" is the net proportion of area cropped with HYV seeds at the district level. "Rural Literacy" and "Urban Literacy" are the proportion of literate people in rural and urban areas, respectively, for different age groups. District controls are Female Population, Rural Population, Rural Workers in manufacturing and Log Roads.

Table 6: Literacy and HYV Adoption (Year Effects)

		(1)	(2)	(3)	(4)	(5)	(6)
		Rural Literacy					
		Aged 5 to 14		Aged 15 to 24		Aged 5 to 24	
HYV * 1971		0.110 (0.029)***	0.104 (0.029)***	0.026 (0.027)	0.045 (0.026)*	0.086 (0.026)***	0.089 (0.026)***
HYV * 1981		0.125 (0.031)***	0.089 (0.027)***	0.081 (0.028)***	0.069 (0.025)***	0.110 (0.030)***	0.082 (0.026)***
1971		0.037 (0.004)***	0.060 (0.013)***	0.108 (0.005)***	0.110 (0.013)***	0.061 (0.004)***	0.077 (0.012)***
1981		0.091 (0.009)***	0.126 (0.025)***	0.141 (0.008)***	0.147 (0.026)***	0.109 (0.008)***	0.135 (0.024)***
Additional effect of mean HYV (in percentage points)	1971	1.1	1.1	0.3	0.5	0.9	0.9
	1981	3.2	2.3	2.1	1.8	2.8	2.1
District Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
District Controls		No	Yes	No	Yes	No	Yes
Observations		762	762	762	762	762	762
R-squared		0.95	0.96	0.96	0.97	0.96	0.97

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. "Rural Literacy" is the proportion of literate people in rural areas for different age groups. District controls are Female Population, Rural Population, Rural Workers in Manufacturing and Log Roads.

Table 7: Literacy and HYV Adoption (Year Effects, Untreated Cohorts)

	(1)	(2)	(3)	(4)	(5)	(6)
	Rural Literacy		Urban Literacy			
	Aged 25 to 59		Aged 5 to 14		Aged 5 to 24	
HYV * 1971	-0.018	-0.011	0.057	0.052	-0.004	-0.006
	(0.012)	(0.012)	(0.030)*	(0.030)*	(0.023)	(0.026)
HYV * 1981	0.014	0.003	0.019	-0.0003	-0.023	-0.035
	(0.014)	(0.013)	(0.021)	(0.022)	(0.019)	(0.021)*
1971	0.034	0.036	0.051	0.059	0.073	0.079
	(0.002)***	(0.007)***	(0.005)***	(0.011)***	(0.004)***	(0.010)***
1981	0.078	0.082	0.097	0.106	0.118	0.126
	(0.004)***	(0.014)***	(0.007)***	(0.022)***	(0.005)***	(0.019)***
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	No	Yes	No	Yes	No	Yes
Observations	762	762	762	762	762	762
R-squared	0.95	0.96	0.93	0.94	0.95	0.96

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. "Rural Literacy" is the proportion of literate people in rural areas for different age groups. District controls are Female Population, Rural Population, Rural Workers in Manufacturing and Log Roads.

Table 8: Literacy and HYV Adoption (Year Effects, Additional Controls)

	(1)	(2)	(3)	(4)	(5)
	Rural Literacy: Aged 5 to 14				
HYV * 1971	0.098 (0.030)***	0.085 (0.029)***	0.078 (0.030)***	0.098 (0.029)***	0.045 (0.027)*
HYV * 1981	0.092 (0.027)***	0.086 (0.027)***	0.072 (0.030)**	0.085 (0.029)***	0.064 (0.023)***
1971	0.041 (0.025)*	0.053 (0.015)***	0.068 (0.013)***	0.056 (0.014)***	
1981	0.074 (0.033)**	0.124 (0.027)***	0.135 (0.026)***	0.128 (0.027)***	
Additional Controls	Rural Migration * year	Scheduled Caste * year	Scheduled Tribe * year	Primary Schools (State)	State Time Trends + Schools (State)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	762
R-squared	0.96	0.96	0.96	0.96	0.97

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. "Rural Literacy" is the proportion of literate people in rural areas for different age groups. District controls are Female Population, Rural Population, Rural Workers in Manufacturing and Log Roads. "Rural Migration" is the proportion of population in a district that has come to rural areas from other districts. "Scheduled Caste" and "Scheduled Tribe" are the proportion of population that belong to either group. "Primary Schools" is the log of primary schools per capita at the state level.

Table 9: Rural Women and Labour Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Rural Women Not Workers (%)		Rural Women in Agriculture			
			(% of rural females)		(% of rural female workers)	
HYV Adoption	-0.086 (0.030)***		0.085 (0.029)***		0.218 (0.070)***	
HYV * 1971		-0.193 (0.046)***		0.187 (0.045)***		0.206 (0.079)***
HYV * 1981		-0.055 (0.030)*		0.056 (0.030)*		0.222 (0.072)***
1971	0.099 (0.014)***	0.107 (0.014)***	-0.071 (0.014)***	-0.078 (0.014)***	0.040 (0.023)*	0.041 (0.023)*
1981	0.021 (0.028)	0.006 (0.029)	0.001 (0.028)	0.015 (0.029)	0.013 (0.042)	0.011 (0.043)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	762	762
R-squared	0.94	0.94	0.93	0.93	0.87	0.87

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. District controls are Female Population, Rural Population, Rural workers in Manufacturing and Log Roads. "Rural Women Not Workers" is the proportion of women that are not employed in any main or marginal activity, "Rural Women in Agriculture" is the number of women working in the agricultural sector taken either as a proportion of rural female or of rural female workers.

Table 10: Literacy and HYV Adoption (Effects by Gender)

	(1)	(2)	(3)	(4)	(5)	(6)
	Rural Literacy				Urban Literacy	
	Aged 5 to 24		Aged 25 to 59		Aged 5 to 24	
	Male	Female	Male	Female	Male	Female
HYV Adoption	0.064 (0.023)***	0.101 (0.033)***	-0.018 (0.016)	0.014 (0.013)	-0.020 (0.020)	-0.036 (0.022)
1971	0.062 (0.01)***	0.090 (0.02)***	0.035 (0.01)***	0.041 (0.01)***	0.052 (0.01)***	0.113 (0.01)***
1981	0.118 (0.022)***	0.147 (0.031)***	0.096 (0.016)***	0.079 (0.016)***	0.084 (0.02)***	0.166 (0.02)***
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	762	762
R-squared	0.97	0.95	0.98	0.94	0.94	0.96

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. "Rural Literacy" and "Urban Literacy" are the proportion of literate people in rural and urban areas, respectively, for different age groups and genders. District controls are Female Population, Rural Population, Rural Workers in Manufacturing and Log Roads.

Table 11: Literacy and HYV Adoption (Year Effects by Gender)

		(1)	(2)	(3)	(4)	(5)	(6)
		Rural Literacy				Urban Literacy	
		Aged 5 to 24		Aged 25 to 59		Aged 5 to 24	
		Male	Female	Male	Female	Male	Female
HYV * 1971		0.074	0.101	-0.022	-0.001	0.006	-0.016
		(0.021)***	(0.030)***	(0.016)	(0.011)	(0.026)	(0.027)
HYV * 1981		0.062	0.102	-0.017	0.018	-0.028	-0.041
		(0.023)***	(0.035)***	(0.017)	(0.015)	(0.021)	(0.022)*
1971		0.061	0.091	0.036	0.042	0.050	0.111
		(0.011)***	(0.016)***	(0.008)***	(0.008)***	(0.009)***	(0.011)***
1981		0.119	0.147	0.096	0.077	0.088	0.169
		(0.022)***	(0.032)***	(0.016)***	(0.017)***	(0.018)***	(0.023)***
Additional effect of mean HYV (in percentage points)	1971	0.8	1.0	-0.2	0.0	0.1	-0.2
	1981	1.6	2.6	-0.4	0.5	-0.7	-1.1
District Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
District Controls		Yes	Yes	Yes	Yes	Yes	Yes
Observations		762	762	762	762	762	762
R-squared		0.97	0.95	0.94	0.98	0.94	0.94

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. "Rural Literacy" and "Urban Literacy" are the proportion of literate people in rural and urban areas, respectively, for different age groups and genders. District controls are Female Population, Rural Population, Rural Workers in Manufacturing and Log Roads.

Table 12: Gender Gap in Rural and Urban Areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Gender Gap				Rural Gender Gap - Urban Gender Gap			
	Aged 5 to 24		Aged 25 to 59		Aged 5 to 24		Aged 25 to 59	
	Rural	Urban	Rural	Urban				
HYV Adoption					-0.074		-0.024	
					(0.024)***		(0.019)	
HYV * 1971	-0.042	0.027	-0.042	-0.039		-0.068		-0.003
	(0.023)*	(0.016)*	(0.012)***	(0.015)**		(0.025)***		(0.020)
HYV * 1981	-0.059	0.017	-0.052	-0.022		-0.076		-0.030
	(0.027)**	(0.015)	(0.017)***	(0.017)		(0.025)***		(0.020)
1971	0.011	-0.063	0.025	-0.018	0.074	0.073	0.045	0.043
	(0.005)**	(0.004)***	(0.003)***	(0.004)***	(0.0051)***	(0.005)***	(0.005)***	(0.005)***
1981	0.051	-0.079	0.074	-0.018	0.130	0.130	0.090	0.091
	(0.010)***	(0.006)***	(0.006)***	(0.007)**	(0.009)***	(0.009)***	(0.009)***	(0.009)***
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	762	762	762	762
R-squared	0.83	0.93	0.95	0.90	0.90	0.94	0.90	0.94

Robust standard errors in parentheses, clustered at the district level. * significant at 10%; ** significant at 5%; *** significant at 1%. "HYV" is the net proportion of area cropped with HYV seeds at the district level. "Gender Gap" is the difference between the proportion of male and female that are literate for different age groups and in rural and urban areas. District controls are Female Population, Rural Population, Rural Workers in Manufacturing and Log Roads.

References

[1] Aghion, P., Burgess R., Redding, S. and Zilibotti, F. (2004), Entry Liberalization and Inequality in Economic Performance, mimeo, London School of Economics.

[2] Aghion, P., Burgess R., Redding, S. and Zilibotti, F. (2007), The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India, London School of Economics, mimeo.

[3] Ahluwalia, S. and Bhatiani (2000), Tariff Setting in the Electric Power Sector, mimeo, TERI.

[4] Arnold, J.M., Mattoo, A. and Narciso, G. (2008), Service Inputs and Firm Productivity in Sub-Saharan Africa: Evidence from Firm-Level Data, Journal of African Economies, forthcoming.

[5] Arora, D.R., Agarwal, B.K. and Gupta, A.K.(1989), Green Revolution and Social Change in Punjab, in Green Revolution and Social Change, edited by Sharma and Dak,Ajanta Publications, Delhi.

[6] Arrow, K. J. (1975), Vertical Integration and Communication, The Bell Journal of Economics, Vol. 6(1), 173-183.

[7] Asian Development Bank and The World Bank Group (2002), ICS - Investment Climate Surveys, mimeo.

[8] Bandyopadhyay, S. (2003), Convergence Club Empirics: Some Dynamics and Explanations of Unequal Growth Across Indian States", Discussion Paper 69, DARP, London School of Economics.

[9] Banerjee, A. and Duflo, E (2003), Reputation Effects and the Limits of Contracting: A study of the Indian Software Industry, Quarterly Journal of Economics, Vol. 115 (3), 989-1017..

[10] Baron, D. and Myerson, R. (1982), Regulating a Monopolist with Unknown Costs, Econometrica, Vol. 50, No. 4, 911-930.

[11] Bernard, A. , Redding, S. and Schott, P. (2006), Multi-Product Firms and Trade Liberalization, CEP Discussion Paper 769, London School of Economics.

[12] Bharadway, K. (1990), Irrigation in India: Alternative Perspectives, Indian Council of Social Science Research.

[13] Bertrand, M., Duflo, E. and Mullainathan, S. (2004), How Much Should We Trust Differences-in-Differences Estimates?, Quarterly Journal of Economics, 119(1), 249-75.

[14] Besley, T. and Burgess R. (1999), Land Reform, Poverty Reduction and Growth: Evidence from India, Quarterly Journal of Economics, 115(2), 389-430.

[15] Besley, T. and Burgess R. (2002), The Political Economy of Government Responsiveness: Theory and Evidence from India, Quarterly Journal of Economics, 117(4), 1415-1452.

[16] Besley, T. and Burgess R. (2004), Can Labor Regulation Hinder Economic Performance? Evidence from India, Quarterly Journal of Economics, 119(1), 91-134.

[17] Biswas, D., Heller, T, Nag, T., Shukla, P.R., Victor, D.and Yajnik, A (2004), Captive Power Plants: Case Study of Gujarat, India, WP 22, Program on Energy and Sustainable Development, Stanford University.

[18] Boserup, E.(1990), Economic Change and the Roles of Women, in Tinker I. (ed.) Persistent Inequalities: Women and World Development. Oxford: Oxford University Press.

[19] Burgess, R and Pande, R. (2005), Can Rural Banks Reduce Poverty? Evidence from the Indian Social Banking Experiment, American Economic Review, 95 (3), 780-795.

[20] Carlton, D. (1979), Vertical Integration in Competitive Markets Under Uncertainty, The Journal of Industrial Economics, Vol.27(3), 189-209

[21] Caves, R. and Bradburd, R. (1987), The Empirical Determinants of Vertical Integration, Journal of Economic Behaviour and Organization, Vol. 9, 265-279.

[22] Chakravarti, A.K. (1972), Green Revolution in India, Annals of the Association of American Geographers, 63 (3), 319-330.

[23] Department of Energy, USA (2003), An Energy Overview of India, mimeo.

[24] Department of Power, Ministry of Energy, India (1991), Performance Budget, mimeo, various years.

[25] Dhanagare, D.N. (1989), The Green Revolution and Social Inequalities in Rural India, in Green Revolution and Social Change, edited by Sharma and Dak, Ajanta Publications, Delhi.

[26] Dinkelman, T. (2008), The effects of Rural Electrification on Employment: New Evidence from South Africa, University of Michigan, mimeo.

[27] Dixit, A. and Stiglitz, J. (1977), Monopolistic Competition and Optimum Price Diversity", American Economic Review, Vol. 67, 297-308.

[28] Dollar, D., Iarossi, G. and Mengistae, T. (2002). Investment Climate and Economic Performance: Some Firm Level Evidence from India , WP No. 143, Center for Research on Economic Development and Policy Reform, Stanford University.

[29] Dollar, D., Hallward-Driemeier, M. and Mengistae, T (2003). Investment Climate and Firm Performance in Developing Economies, mimeo, The World Bank.

[30] Duflo, E.. (2001), Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment, American Economic Review ; 91(4), 795-813.

[31] Duflo, E. and Pande, R. (2007), Dams, Quarterly Journal of Economics, 122 (2), 601–646.

[32] Estache, A. (2004), Emerging Infrastructure Policy Issues in Developing Countries: A Survey of the Recent Economic Literature, INFVP, The World Bank

[33] Evenson, R.E. and McKinsey, J.W. (1999), Technology-climate interactions in the Green Revolution in India, Economic Growth Center Discussion Paper 805, Yale University.

[34] FAO (1997), Women and Food Security, FAO FOCUS on the Issues series.

[35] Foster, A and Rosenzweig, M. (1996), Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution,, American Economic Review 86(4), 931-953.

[36] Foster, A. and Rosenzweig, M. (2003a), Agricultural Productivity Growth, Rural Economic Diversity and Economic Reforms: India, 1970-2000, mimeo, Brown University.

[37] Foster, A. and Rosenzweig, M. (2003b), Agricultural Development, Industrialization and Rural Inequality, mimeo, Harvard University.

[38] Foster, A. and Rosenzweig, M. (2008), Inequality and the Sustainability of Agricultural Productivity Growth,: Groundwater and the Green Revolution in Rural India, mimeo.

[39] Fujita, M., Krugman, P. and Venables, A. (1999), The Spatial Economy, Cities, Region and International Trade, , MIT Press.

[40] Gonzalez. J., Guasch, J. L. Serebrisky, T. (2007), Latin America: High Logistics Costs and Poor Infrastructure for Merchandise Transportation, World Bank, mimeo.

[41] Green, J. (1974), Vertical Integration and Assurance of Markets, in J.E. Stiglitz and F. Mathewson (eds.), New Developments in the Theory of Market Structure, MIT Press, Cambridge, 177-210.

[42] Greene, W.H. (2000), Econometric Analysis, Fourth Edition, New York University, Prentice Hall International.

[43] Grossman, S. and Hart, O. (1986), The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration, The Journal of Political Economy, Vol. 94, No.4, 691-719.

[44] Guasch, J.L., Laffont, J-J and Straub, S. (2003), Renegotiation of Concession Contracts in Latin America, Policy Research Working Paper Series 3011, The World Bank.

[45] Gulati, A. and Narayanan, S. (2003), The Subsidy Syndrome in Indian Agriculture, Oxford University Press.

[46] Harriss J. (1989), Knowing About Rural Economic Change, in Bardhan P. Conversations Between Economists and Anthropologists: Methodological Issues in Measuring Economic Change in Rural India. Oxford: Oxford University Press.

[47] Jalan, J. and Ravallion, M. (2002), Geographic Poverty Traps? A micro model of consumption growth in rural China, Journal of Applied Econometrics, Vol. 17, pp 329-346.

[48] Johnson, D.G. (1997), Agriculture and the Wealth of Nations, American Economic Review, 87(2), 1-12.

[49] Joskow, P. (2004), Vertical Integration, in C. Ménard and M. Shirley (eds.), Handbook of New Institutional Economics, Springer.

[50] Kaur M. and Sharma, M.L. (1989), Impact of Green Revolution on Status of Rural Women, in Sharma & Dak (eds.), Green Revolution and Social Change, Ajanta Publications, Delhi.

[51] Klein, P. (2004), The Make or Buy Decision: Lessons from Empirical Studies, in C. Ménard and M. Shirley (eds.), Handbook of New Institutional Economics, Springer.

[52] Kohli, D. S. and Singh, N. (1997), The Green Revolution in Punjab, India: The Economics of Technological Change, mimeo.

[53] Laffont, J-J (2005), Regulation and Development, Federico Caffe Lectures, Cambridge University Press.

[54] Laffont, J-J and Tirole, J. (1986). Using Cost Observation to Regulate Firms, The Journal of Political Economy, Vol. 94(3), 614-641.

[55] Laffont, J-J and Tirole, J. (1993), A Theory of Incentives in Procurement and Regulation, MIT Press, Cambridge.

[56] Lieberman, M. (1991), Determinants of Vertical Integration: an Empirical Test, The Journal of Industrial Economics, Vol.39(5), 451-466.

[57] Malmgren, H.B. (1961), Information, Expectations and the Theory of the Firm, The Quarterly Journal of Economics, Vol. 75(3), 399-421.

[58] Mas-Colell, A., Whinston, M. and Green, J.R. (1995), Microeconomic Theory, Oxford University Press.

[59] Maskin, E. and Riley, J. (1984), Monopoly with Incomplete Information, The RAND Journal of Economics, Vol. 15(2), 171-196.

[60] Matsuyama, K. (1992), Agricultural Productivity, Comparative Advantage and Economic Growth, Journal of Economic Theory, Vol. 58, 317-334.

[61] Mazumdar and Sharma (1990), Sexual Division of Labour, in Tinker I. (ed). Persistent Inequalities: Women and World Development. Oxford: Oxford University Press.

[62] Mencher J. (1988), Women's Work and Poverty: Women's Contributions to Household Maintenance in Two Regions of South India, in Bruce J. and Dwyer D. (eds) A Home Divided: Women and Income Control in the Third World. Stanford: Stanford University Press.

[63] Melitz, M. (2003), The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, Econometrica 71 (6), 1695-1725.

[64] McGuirk, A and Mundlak, Y. (1991), Incentives and Constraints in the Transformation of Punjab Agriculture, International Food Policy Research Institute.

[65] Ministry of Agriculture and Irrigation, Government of India (1976), Report of the National Commission on Agriculture, Volume X.

[66] Ministry of Industry, Government of India (1993), Handbook of industrial statistics.

[67] Ministry of Power, Government of India (2001), Blueprint for Power Sector Development, mimeo.

[68] Murphy, K., A. Shleifer and R. Vishny (1989a), Income Distribution, Market Size, and Industrialization, The Quarterly Journal of Economics, 104(3), 537-564.

[69] Murphy, K., A. Shleifer and R. Vishny (1989b), Industrialization and the Big Push, The Journal of Political Economy, 97(5), 1003-1026.

[70] Planning Commission, India (1961), Third Five Year Plan.

[71] Planning Commission, India (1999-2002), Annual Report on the working of State Electricity Boards and Electricity Departments, mimeo.

[72] PriceWaterHouseCoopers (1998), Infrastructure Development Action Plan for Chhattisgarh, mimeo.

[73] Ravallion, M. and Datt, G. (1999), When is Growth pro-poor? Evidence from the Diverse Experience of India's State, Policy Research Working Paper 2263, World Bank.

[74] Rosenzweig M. and Schultz T.P. (1982), Market opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in rural India, American Economic Review, Vol. 72(3), 803-815.

[75] Sachs, J., Bajpai, N and Ramiah, A. (2002), Understanding Regional Economic Growth in India, CID Working Paper 88, Center for International Development at Harvard University.

[76] Sibley, D. (1989), Asymmetric Information, Incentives and Price-Cap Regulation, The RAND Journal of Economics, Vol. 20(3), 392-404.

[77] TERI (1999), Indian Power Sector: change of gear, mimeo.

[78] Sims, H. (1988), Political Regimes, Public Policy and Economic Development: Agricultural Performance and Rural Changes in Two Punjabs, Sage Publications, New Delhi.

[79] Sharma, M.L. and Dak, T.M (1989), Green Revolution and Social Change, Introduction, edited by Sharma and Dak,Ajanta Publications, Delhi.

[80] Sridhar, D. (2004), The Green Revolution in India: The Changing Nature of Work and its effect on child nutrition, University of Oxford, mimeo.

[81] TERI (1999), Indian Power Sector: change of gear, mimeo.

[82] Tongia, R. (2003), The Political Economy of Indian Power Sector Reforms, WP 4, Program on Energy and Sustainable Development, Stanford University.

[83] Topalova, P. (2004). Trade Liberalization and Firm Productivity: the Case of India,.IMF Working Paper No. 04/28.

[84] United Nations Development Programme (2003), Tamil Nadu Human Development Report. New York City: United Nations Development Programme.

[85] Vanneman, Reeve and Douglas Barnes.(2000), Indian District Data, 1961-1991: machine-readable data file and codebook. College Park, Maryland: Center on Population, Gender, and Social Inequality.

[86] Voigtländer, N. and Voth, H.-J. (2006), Why England? Demographic factors, structural change and physical capital accumulation during the Industrial Revolution, Journal of Economic Groth, 11(4), 319-361.

[87] Williamson, O. (1971), The Vertical Integration of Production: Market Failure Considerations, American Economic Review, Vol. 61(2), 112-123.

[88] Wooldridge, J. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press.

[89] Wooldridge, J. (2003), Cluster-Sample Methods in Applied Econometrics, American Economic Review, XCIII, pp 133-138.

[90] World Bank (2000), India: Policies to Reduce Poverty and Accelerate Sustainable Development, World Bank.

[91] World Bank (2005), Infrastructure and the World Bank: A Progress Report, World Bank.

[92] World Bank (2006), Infrastructure: Lessons From the Last Two Decades of World Bank Engagement, World Bank.

[93] World Bank (2007), World Development Report 2008: Agriculture for Development, World Bank.