FACTORIES AND FARMS: HOW DOES ECONOMIC GROWTH IMPACT RURAL EDUCATIONAL INVESTMENT?

REENA BADIANI[†]

PRELIMINARY: PLEASE DO NOT CITE

ABSTRACT. Do households in high growth regions invest more in education because their incomes are growing or because growth raises their returns to education? What is the relative contribution of these channels to explaining education choices? To answer these questions, this paper exploits variation in the demand for skill and land in two major sectors of rural India - agriculture and manufacturing. To guide my empirical analysis, I build a model of household education choices. The model generates three testable predictions on how changes to the economic environment alter returns to education and household incomes and how changes in these determinants of education decision rules feed through to education choices. I test the predictions of the model using district and household level data from India between 1983 and 1999. My results indicate that household income and educational investment responses to changes in the economic environment vary with the existing distribution of asset and skill endowments in an economy. I examine the decision of rural landless households to send their children to school. Rural educational investment among these households is low - one in three children of primary school age didn't attend school during the 1990s. I find that male education responds positively to increases in income and returns to education, and that the income effect is substantially larger in magnitude than the returns to education effect.

Keywords: Labor Markets, Manufacturing Employment, Agricultural Development
JEL Codes: J30, J43, O25, Q10, Q31

[†]DEPARTMENT OF ECONOMICS, YALE UNIVERSITY, BOX 208268, NEW HAVEN, CT 06520-8268
E-mail address: reena.badiani@yale.edu.
I am indebted to Joe Altonji, Mike Boozer, Mark Rosenzweig and Chris Udry for their guidance and comments. I thank Padmaja Mehta at the Ministry of Agriculture for allowing me to access her data resources. I thank David Atkin, Prashant Bharadwaj, Gharad Bryan, Rahul Deb, James Fenske, Rachel Heath, Stefan Hunt, Katrina Jessoe, Mushfiq Mobarak, Gus Ranis, Steve Redding, Maher Said, Peter Schott, Paul Schultz, Siddharth Sharma, TN Srinivasan and Vis Taraz for suggestions and support. Saba Joshi provided excellent research assistance. All remaining errors remain mine.

1. Introduction

For many people in the developing world, the most valuable asset they have is their labor market time. In areas where migration is low, the wage distribution and mix of occupations faced by individuals in their local labor market will play a large role in determining their education choices, the occupations and sectors they are employed in and the incomes that they earn. Changes to the economic environment, such as sector specific technological change or skill biased growth, are likely to alter the return to productive assets, such as human capital and land. Household demand for education in high growth regions may increase because household incomes are growing and because growth raises their return to education. Economic growth may directly raise the returns to schooling. For example an increase in the complexity of the environment or technological change may require greater educational skills (Nelson and Phelps, 1966; T. W. Schultz, 1975). Growth may also alter the distribution of income through changing the relative return of productive assets in an economy, such as land (Kuznets, 1955). Therefore, both the level and slope of wages enter directly into education decision rules through household incomes and labor market returns to education. To disentangle the effect of returns to education and household incomes on education decisions and estimate the channels through which growth alters educational investment, we need to understand how it alters both the returns to education and the level and distribution of household incomes.

This paper makes two contributions to our understanding of how changes in the aggregate economic environment affect household incomes and education choices. First, I investigate the relationship between growth, returns to education and incomes. I examine how two drivers of growth differentially alter the wages of educated and uneducated workers, changing the returns to education. I test whether the level and distribution of rural incomes varies with the underlying distribution of household assets. To do this, I exploit variation in the demand for skill and land in two major sectors of rural India - agriculture and manufacturing. Since the two sectors vary in their demand for skill and land, shocks which affect the size of these two sectors differ in their impact on wages at different education levels and on household incomes, according to their initial skill and land endowments. Secondly, I detect and quantify the response of investment in education to changes in the returns to education and household incomes. The estimates focus on understanding the determinants of the decision to send children in these households to school among rural landless households.

This paper exploits variation in the demand for skills and land across the two major sectors of rural India, agriculture and manufacturing, to examine how changes in the aggregate economic environment affect incomes and education choices. There is substantial variation across the sectors in the complexity of occupations conducted and their complementarity with education. Hired labor in the agricultural sector conducts activities such as harvesting, weeding and sowing, tasks with predominantly use physical rather than cognitive skills.[1] By contrast, hired labor in the non-agricultural sector conducts a variety of tasks, from white collar occupations such as managerial, decision making and clerical tasks to physical tasks in

[1]Complex decision making tasks are conducted in agriculture, such as deciding what crops to grow and fertilizers to use. These tasks appear however to be conducted by household members and are therefore associated with households land ownership (Foster and Rosenzweig, 1996).

blue collar unskilled occupations. The two sectors therefore vary in their demand for education in the hired labor market, through differences in the use of physical or cognitive skills. Shocks which change the proportion of the population employed in these sectors, such as unanticipated changes to industrial policy that shift manufacturing labor demand, impact the wages of workers at different education levels. Alternatively factors which affect the agricultural sector change the return to land as well as the return to skills. For example, the introduction of new seed varieties increases total factor productivity in agriculture.[2] Correspondingly, the two sources of wage and income growth examined in this paper are increases in total factor productivity in agriculture and changes to manufacturing labor demand, induced by changes to industrial policy.

The focus of this paper is on the decision of rural landless households to send their children to primary school. Rural educational investment among these households is low - one in three children of primary school age in these households didn't attend school during the 1990s.[3] Conditional on having started formal schooling, over four fifths of children get to at least primary school level.[4] The decision as to whether to acquire any education or not is an important margin along which education choices are made (Dreze and Kingdon, 1999). One explanation for low levels of educational participation in these households is low rates of return to education in agricultural (WDR, 2008; Rosenzweig, 1995). Examining the education response of landless households to changes in the skill composition illuminates how changes in the structure of the economy can alter decisions of a group typically found at the bottom tails of rural income, land ownership and education distributions.

The literature on education in developing countries can be broadly split into studies that focus on factors which alter the supply of education, such as enhancing the quality of education or reducing fees, and those which alter the demand for education. Other studies of the demand for rural schooling find that educational choices respond to rising return to education (Kochar, 2004; Rosenzweig and Foster, 1996), household incomes (Edmonds, 2006; Edmonds et al, 2004; Jacoby and Skoufias, 1997), the opportunity costs of schooling (Rosenzweig and Foster, 2004) and changes in the direct cost of schooling (Schultz, 2004; Kremer et al. 2003).[5] Previous studies predominantly study education in a partial equilibrium framework, focusing on one determinant of the education decision rule in a piece-meal approach. I deviate from the literature by nesting household education choices in a general equilibrium framework in which changes to the aggregate economic environment alters rural wages, household incomes and the returns to schooling. The approach taken is motivated by the observation that both the incomes and returns to education enter directly into education decisions rules; to disentangle these two effects we need to generate independent variation in these two determinants of schooling. To do this, I exploit the

[2]An example of this is the "Green Revolution" in which yields of food grains such as rice and wheat doubled during the 1960s and 1970s due to the introduction of High-Yielding Variety seeds (Evenson). A lesser discussed example is the "Yellow Revolution", in which new varieties of oilseeds resulted in a doubling of yields during the 1980s and 1990s in the Semi-Arid Tropic areas in which these crops are best suited (Gulati, 2004).
[3]Primary education in India has been historically neglected by the state, with education expenditures focused on tertiary education (Dreze and Sen, 1995).
[4]83% of children complete primary school, conditional upon ever having entered formal education. This descriptive statistic is based on 15-20 year olds in the 1999 round of the Employment-Unemployment module in the NSSO. Respondents were asked whether they had ever attended school and what level of education they had achieved.
[5]Recent studies on the supply side include, to name but a few, Duflo (2000), Deaton and Case (1997).

observation that the returns to education and incomes across households respond differently to the two drivers of growth examined in this paper.

To guide my empirical analysis of household education and income responses to changes in the aggregate economic environment, I build a model of household education choices. Households incomes and the returns to education facing households are determined within the model. Households are endowed with two assets: their land-holdings and education. Labor market returns to education are defined as the relative wages of skilled (literate) and unskilled (illiterate) workers. Wages and incomes grow over time due to shocks in the aggregate economic environment, notably changes to total factor productivity in the agricultural sector and shifts in skilled and unskilled manufacturing labor demand induced by policy changes.

The model highlights a source of endogeneity bias and generates three testable predictions which frame the empirical analysis. Manufacturing employment and its skill composition are choices that reflect the wages of skilled and unskilled workers, as well as determinants of education and income.[6] The model suggests exclusion restrictions which I use to frame an instrumental variables strategy: if the labor market is the only local input market in which the manufacturing and agricultural sector overlap, then shifts in manufacturing labor demand, induced by local responses to industry level policy, will only alter education, incomes and wages through competition in the local labor market.

The model generates three testable predictions concerning wage, income and education responses to changes in the aggregate economic environment. Firstly, educational investment is predicted to increase in the returns to schooling, household incomes and to decrease in the opportunity cost of school attendance. Secondly, the wages of both skilled and unskilled workers are predicted to grow as agricultural productivity and manufacturing labor demand rises. The magnitude of the effect on skilled and unskilled wages is predicted to differ in the relative shifts in the demand for skilled and unskilled workers. Thirdly, household income responses to changes in the aggregate economic environment vary according to the household's skill and land endowment. Income in all households is predicted to increase in agricultural productivity, which raises the return to both labor and land. Since landed households earn rents from both their labor market time and their landownings, their income response is predicted to be greater than that of landless households. Manufacturing growth is predicted to raise the return to labor and reduce the return to land; therefore the model predicts that the incomes of large landed households falls as labor demand in manufacturing rises.

The predictions of the model are tested using four rounds of district and household level data from rural India between 1983 and 1999. I use a novel instrumental variables strategy to identify regional variation in manufacturing labor demand that is uncorrelated with unobserved correlates of education, wages and income. I use variation in an industry's technologies, in terms of the raw materials they use, and in industry level import tariff and licensing policies, to generate region and time varying variation in total and skilled manufacturing employment. The intuition behind the strategy can be seen through

[6]For example, raw correlations between manufacturing employment, wages and education are likely to reflect manufacturing input choices that are correlated with the attributes of the rural population, such as the quality of education, as well as the direct relationship between the explanatory and outcome variables.

an example - Figure 1 provides a graphical depiction of this example. Limestone is a key and bulky input into cement production - it is expensive to transport and easiest to process at source. Employment in the cement industry in India is predominantly found where cement grade limestone is found, while employment in the structural wood products industry is greatest in areas which are densely wooded, although the relationship is not monotonic. Employment growth between 1990 and 2000 is greater in the cement industry in districts in which cement grade limestone deposits are larger, while in the structural wood products industry it is greater in densely wooded areas. Since industries vary in the raw materials that they use, their employment responses to changes in industry level policy vary across regions according to differences in raw material endowments across regions. I use variation across industries in the broad types of raw materials they use and over time in their policies to identify the causal impact of increases in manufacturing employment on the outcomes variables of interest.[7]

The agricultural productivity frontier is measured using a proxy variables strategy. I combine information on the technological frontier of crops across India with variation in the type of crops that can be grown within India. The maximum of the India-wide yield in a particular crop is used to measure its frontier. I use a district-level time-invariant measure of the physical suitability of crops to local agro-climatic conditions to create a Lespeyres volume weighted index which captures how districts that are suited to different crops differ in their technological frontier.

The IV and proxy variables strategies are used to estimate how changes in agricultural productivity and in the demand for labor in the manufacturing sector, broken down by skill, alter wages, household incomes and education. The IV estimates indicate that growth in rural manufacturing employment increases unskilled wages, increases consumption among landless and small landowning households and reduces poverty. The type of manufacturing growth matters: a 10% increase in manufacturing unskilled employment raises unskilled wages by approximately 3% and reduces the head-count rate of poverty by 7%, while skilled employment raises unskilled wages by 0.6% and reduces poverty by less than 1.4%. Agricultural productivity raises unskilled wags by 40% but has little impact on skilled wages. Therefore the two sources of growth vary in their impact on the wages of skilled and unskilled workers, altering the relative return to education in the labor market.

The estimates suggest that the source of growth is important for understanding the distributional impact of economic change. Agricultural technical change is found to increase consumption inequality in rural areas while manufacturing employment reduces it - a 10% increase in unskilled employment reduces the gini coefficient of inequality by 0.007 points, or 2.8% o the 1987 mean while a 10% increase in agricultural productivity increases it by 0.006 points. The intuition behind this lies in the relative impact of the two different sources of economic growth on the returns to labor and land in local markets. The consumption response of non-cultivator households to the two sources of growth varies with their initial skill endowments. Illiterate landless households exhibit strong consumption responses to unskilled manufacturing employment - a 10% rise in manufacturing employment raises consumption in this group by 2.4% - while literate landless households exhibit consumption responses to both skilled and unskilled

[7]Changes in industrial policy over this period were largely driven by a series of political crises. They are therefore unlikely to reflect the decision of policy makers to use industrial policy as a tool to, for example, reduce poverty through focusing on industries which demand unskilled workers.

manufacturing growth. These results indicate that, in the presence of skilled biased manufacturing growth, the initial distribution of education determines which households experience a rise in household welfare. The results provide empirical evidence supporting the hypothesis that the existing distribution of asset and skill endowments determines how household income responds to periods of economic growth.

The relationship between educational investment and the two drivers of wage and income growth varies both within and across household skill and landholding groups. Raising agricultural productivity by 10% increases the probability a boy aged 5 to 9 attends school by 0.1 in illiterate landless households while it reduces the probability of attendance by 0.05 among the sons of large landed households. The education response to changes in agricultural productivity for these households reflects a combination of the income, returns to education and opportunity cost effects. I estimate the structural parameters of the education decision rule using Minimum Distance. I find that boys in rural landless households exhibit large positive educational investment responses to rising household incomes and labor market returns to education, but reduce enrollment in response to factors which increase the opportunity cost of schooling. A 10% increase in household income and in the labor market returns to education raises the probability of attending school of boys aged between 5 and 9 by 0.3 and 0.1, while increasing the opportunity cost has no effect. Females, in contrast, do not exhibit positive responses to rising labor market returns to education - a finding which fits with the observation of low female participation in complex occupations in the rural areas.

The results suggest that sources of growth which raise the unskilled wage but have little impact on the skilled wage, such as agricultural technological change and unskilled manufacturing growth, both increase educational investment at the same time as raising consumption among households at the bottom end of income and asset distribution in India - rural illiterate landless households. In comparison, the estimates suggest that skill biased manufacturing growth raises the educational investment of these households whilst having little impact on their current welfare. These results indicate that the skilled biased focus of Indian manufacturing has had little effect on the incomes or education choices of the rural poor.

This paper relates to the literature of the demand for education. Foster and Rosenzweig show that, during a period of technical change in Indian agriculture between 1971 and 1982, the educational enrollment of children in cultivator households was responsive to changes in the returns to education facing these households. They show that the introduction of HYV seeds, raised the return to literacy among cultivator households through.[8] This paper contributes to the literature by examining whether *different drivers of* economic growth alters the relationship between growth and educational investment. If growth induces changes in incomes and returns to education that are complementary to assets traded in incomplete or thin markets, then the returns to education facing households and their education responses vary according to the initial distribution of assets. For example technical change is complementary to land, whose markets are thin in India; as found by Foster and Rosenzweig (1996), education responses to

[8]Foster and Rosenzweig (1996) use changes in agricultural technology during the green revolution in India to estimate the response of educational investment to changes in the returns to schooling. They find that technical change in agriculture increased the returns to schooling and educational investment of land-owning households, who conduct complex decision making tasks on their own land-holdings. In contrast, educational investment in landless households, declines in the face of technical change, which is attributed to increases in child labor market wages that raise the opportunity cost of schooling (Foster and Rosenzweig, 2004)

technical change in agriculture vary across landed and landless households. In contrast, changes in the economic environment which raise the return to education in the hired labor market, such as the expansion of skilled labor demand in the manufacturing sector, will induce education responses from a different segment of the rural population. Therefore, to understand how economic growth alters both the levels and the distributions of educational investment we need to first understand the relationship between the drivers of growth and the existing distribution of endowments in the economy.

Finally, this paper also suggests an answer to an overarching question of interest to policy makers: did agricultural productivity growth or employment growth in manufacturing have a greater impact on increasing agrarian wages and reducing poverty between 1983 and 1999? My estimates suggest that growth in agricultural potential resulted in a 21.6% increase in real wages over the period, accounting for just under half of the total wage growth of 52% over the same period. Total manufacturing employment expanded by 40%, although much of this growth consisted of skilled employment. Growth in manufacturing employment over this period raised the wage of rural unskilled and skilled workers by approximately 7% and 15% respectively. If manufacturing employment growth had been purely unskilled, the wages of rural unskilled workers would have increased by 16%. The small impact of the manufacturing sector on unskilled wages and poverty is an observation that has been repeatedly asserted in the policy literature. To my knowledge, the estimates in this paper are the first to empirically validate this observation and to provide an explanation of why this is the case.

The paper is structured as follows. The next section describes the setting of rural India and provides descriptive evidence which is called upon throughout the paper. In section 3, I put forward a model of educational investment by rural households in the context of a two sector rural economy. The model makes 4 sets of testable predictions. In sections 4 through 8 I test the predictions of the model. Section 9 concludes.

2. Setting: The Rural Indian Economy

In this section, I provide descriptive evidence supporting the central arguments of the paper.

The first hypothesis is that, within rural areas, the two major sectors of interest are the agricultural and manufacturing sectors. In section 2.1 I present descriptive statistics on the relative size of these sectors. I show that there is substantial variation over time and across the states of India in the proportion of the workforce employed in these sectors. The second hypothesis is that occupations vary substantially across the agricultural and manufacturing sectors. In section 2.2 I present evidence indicating that the vast majority of hired workers in agriculture conduct unskilled (physical, manual labor) tasks. Workers in the manufacturing sector conduct a mix of skilled (complex tasks, decision making) and unskilled tasks. I show that a wide variety of tasks in the manufacturing sector use mathematical knowledge and literacy and that the skill composition of manufacturing varies substantially by industry. Thirdly, I hypothesize that education is higher amongst individuals conducting complex tasks in both sectors. The data suggest a strong relationship between the complexity of the task conducted and literacy: agricultural laborers are disproportionately illiterate, while manufacturing white collar workers have the highest level of literacy amongst workers in the manufacturing sector. In section 2.3 I show that, in the wage labor market, literate individuals in the manufacturing sector receive a premium that is largely explained by the more complex tasks they're performing. Finally, in section 2.4 I show that there is a strong positive correlation between educational enrollment and the proportion of the population employed in manufacturing, which is driven by skilled manufacturing employment.

The descriptive statistics on India's 15 major states are based on four rounds of employment data collected in 1983, 1987, 1993 and 1999 by the National Sample Survey Organization (NSSO) of India. In addition, I make use of the 1982 and 1999 rounds of data from the Rural Economic and Development Surveys (REDS) collected by the National Council of Applied Economics Research (NCAER). Both of these data sets are described in greater detail in the Data Appendix (Appendix B).

2.1. **The Sectoral Composition of Labor Markets.** Table 1a presents descriptive statistics which divide the rural workforce into agricultural and non-agricultural sectors in 1983 and 1999, the end points of the period examined in this paper. During this period, the majority of the working age population in rural India was employed in agricultural activities: 65% of males and 34% of females declared agriculture to be their primary sector of employment in 1999. The agricultural sector can broadly be separated into landed households, who cultivate their own land and may also work on the wage labor market, and landless households, who work on the wage labor market. There is substantial variation in the size of the agricultural workforce across the major states of India: the agricultural sector employed 85% of the workforce in Madhya Pradesh in 1999 compared to 41% in Kerala and 61% in Tamil Nadu. The manufacturing sector is the largest employer in the non-agricultural workforce in rural areas. In 1999, approximately 31% of non-agricultural workers were employed in this sector (NSS Report 460).

2.2. **Occupations Tasks and Education by Sector.** In this section, I provide evidence that the complexity of tasks conducted in rural labor markets varies by sector - hired workers in agriculture conduct mostly unskilled tasks while those in manufacturing conduct a mix of skilled and unskilled tasks.

Table 1a provides evidence that the hired labor force in agriculture conducts mostly manual, physical tasks and that these workers are disproportionately illiterate. A substantial fraction of those employed in the agricultural sector work in the wage labor market - 28.5% of males and 18% of females worked for wages in 1983. The vast majority of these workers conducted manual labor tasks. The NSSO survey collects information on the task conducted during the previous week. Agricultural work is divided into manual tasks such as weeding, sowing and transplanting and non-manual tasks in cultivation. 99.1% and 99.5% of hired male and female labor time is devoted to manual tasks. Agricultural laborers are disproportionately illiterate relative to the population as a whole - in 1983, 71% of working aged males who work as agricultural laborers are illiterate, compared to 53% of the rural working age population as a whole. Among females, 94% of agricultural laborers were illiterate, compared to 83% of working age females in the population as a whole in 1983.

Non-manual work is more prevalent amongst individuals who cultivate their own land. Slightly over a half of males and females working in the agricultural sector reported working on their own household's lands as their primary activity. Unlike in the hired labor market, there appears to be some scope for non-manual tasks in agriculture - approximately 13% of family time in agriculture is devoted to supervisory activities, according to the 1999 round of REDS. Males working on their household's land spent 2.9% of their time conducting non-manual work in agriculture according to the 55th round of the NSS. Household heads engaged in cultivation devoted 4.3% of their time to these tasks. Tables 1a and 1b indicate that individuals who work on their own household lands have much higher levels of education than hired labors in the agricultural sector. Household heads, who are likely to be making decisions about input usage, are found to be slightly more educated than individuals of the same age who are not household heads.[9]

Occupations within the manufacturing sector can broadly be divided into production and non-production work and, within these, into white-collar and blue-collar. Table 2 breaks down manufacturing employment into white-collar, blue-collar skilled and blue-collar unskilled occupations. Blue-collar unskilled work is defined as production work in which purely physical tasks were reported.[10] Approximately 35% of male and 55% of female occupations in manufacturing fall into this category.

Table 2a indicates that education is decreasing in the skill category of the occupation conducted. White-collar and blue-collar skilled workers are comparable in their levels of education - 78% of male white collar workers are literate compared to 72% of blue-collar skilled workers. Blue-collar unskilled workers have much lower levels of literacy relative to the more skilled manufacturing groups - 49% were literate in 1987 and their education levels are similar to those of the rural working age population as a whole.[11]

Table 2b provides an indication of why education may vary according to the skill category of the occupation conducted. The table describes the different types of tasks conducted by occupation group in manufacturing. The task definitions come from the Dictionary of Occupational Titles, 1977.[12] Each task variable is assigned a value ranging from zero to ten, with higher values representing greater intensity of a skill used in an occupation. The five broad measures of tasks are routine manual activity, non-routine manual activity, routine cognitive tasks, non-routine interactive tasks (direction, control and planning) and quantitative and analytical tasks (mathematical reasoning).[13] White collar occupations appear to use

[9]Table 1b conditions on age polynomials and region fixed effects to examine the education levels of individuals in different occupations.

[10]The breakdown of occupations into categories was done by reading and classifying occupations on the basis of the description of occupations given at a 5-digit level in the National Classification of Occupation, 1968. The blue-collar categories were separated according to the task described. Those which were described as purely manual work (where the words "work doing purely physical activity" cropped up) were classified as blue collar manual work and those which are more "technical" - working with machines, mixing paint using the right proportions of different chemicals - were defined as blue collar skilled. White collar occupations include managers, clerks, secretaries, production-floor supervisors and quality control workers. These codes and categorizations are available from the author on request.

[11]This is shown more clearly in table 1b, which conditions on village fixed effects to take out any concerns that differences in education levels amongst agrarian and manufacturing workers may arise simply due to the disproportionate presence of manufacturing employment amongst more highly educated regions. The literacy of blue-collar manual manufacturing workers lies just below the population average while skilled manufacturing workers lies approximately 20 percentage points above the average.

[12]A crosswalk was created between the 1968 Indian National Classification of Occupations and the 1960 US Census of Occupation Titles. This crosswalk is available from the author on request. Autor, Levy and Murmane aggregate the data from this publication to capture five measures of occupational skills by 1960 occupations. I thank David Autor for sharing this data with me.

[13]The US Department of Labor periodically evaluates the tasks required for more than 12,000 occupations. Tasks are assigned values ranging from zero to ten, with higher values representing greater intensity of a given skill used in an occupation.

quantitative and non-routine cognitive skills more than both sets of blue-collar occupations. The blue-collar occupations appear to be quite similar in all attributes, with the exception of routine cognitive tasks which are conducted to a far higher degree in the skilled blue-collar occupation bracket.[14]

2.3. **Wages by Sector, Occupation and Complexity of Task.** In the previous section, I showed that individuals conducting more skilled, complex tasks in both manufacturing and cultivation have higher levels of education than those conducting unskilled, manual tasks. In this section, I present descriptive statistics which indicate that wages vary according to an individual's education level, the industry worked in and the complexity of the task performed.

Table 1c displays descriptive statistics on wage earners in rural areas. The table reveals that literate males and females working in the wage labor market earn approximately 20% more than illiterate males and females. Illiterate workers in the manufacturing sector earn a 10% wage premium over illiterate workers in other sectors (column c, table 1c). Illiterate workers in the agricultural sector earn no premium over other sectors, although literate workers in this sector do earn 7% more than illiterate workers. Splitting agriculture and manufacturing sector workers by task rather than by education level shows that the premium is driven by complex tasks - an illiterate person conducting an unskilled task in manufacturing earns the same wage as an illiterate person in the other non-agricultural sectors.[15]

2.4. **Educational Enrollment and the Non-Agricultural Sector.** In this section, I present evidence that educational enrollment varies substantially across India. There is a strong positive correlation between attendance at primary and middle school levels (6-11 and 12-16 year olds respectively), and in the proportion of the population employed in the manufacturing and non-agricultural sector more generally.

Table 4a reveals that approximately two thirds of boys and just under a half of girls attended primary school between the ages of 6 and 11 in 1987. Both male and female attendance in education drops substantially after primary school. Educational enrollment varies substantially by state in India. In 1987, enrollment among boys aged 12-16 was highest in Kerala, followed by Maharashtra, Punjab, West Bengal and Tamil Nadu. Interestingly, Kerala, West Bengal and Tamil Nadu, as discussed above, have high proportions of the non-agricultural workers, while Punjab has the strongest agricultural sector. Bihar and Madhya Pradesh were in the bottom three for enrollment as well as for the size of the non-agricultural sector.

Table 4b indicates that there is a strong positive correlation in the raw data between the proportion of children enrolled in school and the proportion of the population employed in the non-agricultural sector, in particular in the skilled manufacturing sector.

[14]An example of routine cognitive tasks taken from the Handbook for Analyzing Jobs, quoted from Autor, Levy and Murnane, is: "measures dimensions of bottle, using gauges and micrometers to verify that setup of bottle-making conforms to manufacturing specifications".

[15]Non-manual workers in agriculture earn a premium over agricultural laborers. This should however be taken with a note of caution since the number of workers falling into this industry-occupation cell is extremely small - 0.7% of hired labor in agriculture reported conducting these occupations.

3. Theoretical Model

In this section, I present a model of household education choices. The small country general equilibrium model builds upon Foster and Rosenzweig (1996, 2003) and Ellison and Glaeser (1999). The purpose of the model is two fold. Firstly, the model motivates my empirical strategy by highlighting a source of endogeneity bias and puts forward exclusion restrictions to overcome it. Secondly, it delivers 3 sets of predictions about the response of equilibrium wages, household incomes and education choices to shocks in the aggregate economic environment, notably changes to total factor productivity in the agricultural sector and shifts in skilled and unskilled manufacturing labor demand induced by policy changes. The predictions of the model are tested in section 4.

3.1. **Model Environment.** A country consists of D regions whose borders define a closed labor market. Households in each regional economy have two generations in them who live for two periods; education of the household child is chosen in period one. There are two sectors in each economy - agriculture and manufacturing. I use a specific factors framework (Ricardo-Viner approach) to model the interaction of these sectors. The agricultural and manufacturing sectors overlap in two locally traded inputs - skilled and unskilled labor. Equilibrium wages, household incomes and education choices are determined within the model, and vary with shocks to the aggregate economic environment. Throughout the model, I drop economy and time subscripts unless necessary.

3.1.1. *Economy Endowments and Environment.* Each economy, denoted with a d subscript, is modeled as a small open economy endowed with four immobile factors of production - a population (P), land (A) and two non-renewable raw materials (T_1 and T_2). Final goods are freely traded across regions within the country. There is no migration across countries, therefore the adult population constitutes the labor force. The population is divided into four groups, land owning and landless households and, within these, educated and uneducated denoted P_{As}, P_{Au}, P_{LLs} and P_{LLu} respectively. Land is distributed evenly over landed households. The productivity of agriculture, θ, varies by location and over time according to the region's agro-climatic conditions.[16] There are no land or credit markets but labor markets are active.

Raw materials are immobile across countries and the marginal cost of extracting a raw material is decreasing in the resource endowment. These assumptions imply that there is a mechanical relationship between the price of raw materials and their stock, and that the price of raw materials de facto varies across countries. Evidence on the relationship between extraction costs and the stock of a non-renewable resource is derived from the economic geology literature (Kessler, 1996).[17]

Let the marginal cost of extracting raw materials 1 and 2 be denoted by $c^{r1} = c(r_1, T_1)$ and $c^{r2} = c(r_2, T_2)$ respectively, where r_1 and r_2 denotes the amount of the resources extracted and T_1 and T_2 denotes the total endowment of the resources in the region. The raw material is priced at marginal cost, $p^{r1} = c^{r1}$

[16]The introduction of High Yielding Variety (HYV) seeds in India is an example of a technology induced change in agricultural productivity which varied across districts in India. (Rosenzweig and Foster, 1996).

[17]The immobility of raw materials is an extreme assumption which is made for analytical convenience, an alternative assumption, which would generate similar testable predictions, is that the prices of raw materials are increasing in the distance from their location due to transportation and trade costs across countries.

and $p^{r2} = c^{r2}$. The cost of extracting the raw material is decreasing in the total endowment of the region - $p_2^{r1} < 0$ and $p_2^{r2} < 0$. For simplicity, the marginal cost of extracting the resource is the same no matter how many units are extracted: $p_{12}^{r1} = 0$ and $p_{12}^{r2} = 0$.[18] The rents from raw materials flow outside the region.

3.2. **Households: Endowments, Preferences and Maximization Problem.** Households are modeled over two periods and consist of two generations. In the first period, each household includes an adult a and a child y. In the second period, the child becomes an adult. Household subscripts are suppressed in this section. Households have three endowments in both periods: land (A), adult education (s^a) and a unit of adult labor market time. Adult education takes the value of 0 if the adult is uneducated or 1 if the adult is educated. In the first period households are additionally endowed with a unit of child time, which can be devoted to either education or to the production of a non-marketed consumption good. Household adult education in period 1 is exogenously determined, education in period 2 is the only endowment which is endogenously determined within the model. Adults choose the education of their child in period 1, the choice is discrete - the child either attends school or does not.

Households have time separable preferences over material good consumption in period 1 and 2 - c_1 and c_2. Utility is concave in consumption - $0 < \rho < 1$.[19]

$$V(c_1, c_2) = u(c_1) + \beta u(c_2) = c_1^\rho + c_2^\rho \tag{1}$$

In the first period, households earn income, choose how much time their child spends in school and consume. Consumption goods are of two perfectly substitutable varieties: home produced by the child using a unit of child time l_y or purchased at price p_c:

$$\begin{aligned} c_1 &= c_1^d + c_1^p = \alpha l_y + c_1^p \\ c_2 &= c_2^p = c_2^p \end{aligned}$$

where $0 < \alpha < 1$. Households are unable to borrow or save implying that their budget constraints clear every period. In the first period, adults divide household income between consumption and the schooling of their child. In the second period, the first period children become adults and the adults from the initial household no longer work. For tractability, there is no leisure or child wage labor in the theoretical model; in the empirical framework I discuss how these channels alter the predictions of the model.[20]

Adults split their labor market time between four activities: unskilled work in agriculture, unskilled work in manufacturing, skilled work in manufacturing and skilled work in agriculture. Unskilled work in both agriculture and manufacturing are activities which both educated and uneducated individuals

[18]For simplicity and ease of exposition, the simplest structure for the raw materials market is put forward. The assumptions are modifiable with minimal repercussions to the main testable predictions of the model. A more complicated market structure would allow for a fixed total quantity of raw materials to be consumed every period, and for the cost of extracting the materials to be convex in total consumption.

[19]There is no symbolic consumption of education - i.e. there is no "warm glow" of giving to one's child (Andreoni, 1989).

[20]Jayachandran (2004) presents evidence consistent with the pattern that labor supply responses to a productivity shock are more inelastic in districts in which credit constraints are greater, in which a greater proportion of workers close to subsistence and in which migration costs are high. Foster and Rosenzweig (2004) study child labor markets in India. For tractability, both child labor markets and the leisure-work trade off have been excluded from the model. In the empirical model, I incorporate child labor and allow labor supply to vary according to local labor market characteristics.

can conduct; skilled work is defined as activities that only literate individuals can do. Skilled work in agriculture consists of supervisorial activities, which can only be conducted by household members. Since there are no land markets, this implies that only landed households conduct skilled work in agriculture. Children either go to school or produce the domestic consumption good. The time constraints of adults and children in period 1 and working adults in period 2 are given by:

$$1 = (1 - s_1^a)(l_{u1}^A + l_{u1}^M) + s_1^a(1(A_h > 0)l_{s1}^A + l_{s1}^M) \tag{2}$$

$$1 = (1 - s_1^y)l_{y1} + s_1^y l_{s1} \tag{3}$$

$$1 = (1 - s_2^a)(l_{u2}^A + l_{u2}^M) + s_2^a(1(A_h > 0)l_{s2}^A + l_{s2}^M) \tag{4}$$

where l_{ut}^A and l_{ut}^M is unskilled time in agriculture and manufacturing in period $t = 1, 2$, l_{st}^A is skilled time in agriculture and manufacturing. l_{y1} denotes a unit of child home production time in period 1, while l_{s1} denotes a unit of time spent at school.

Households earn income from supplying labor to the wage labor market and, if landed, earn profits from cultivation. In addition, they earn income m from an exogenous source. The household budget constraint, which clears every period, is given by:[21]

$$\begin{aligned} y_1 &= (1 - s_1^a)w_{u1}(l_{u1}^A + l_{u1}^M) + s_{h1}^a w_{s1}(l_{s1}^M + \mathbf{1}(A > 0)l_{s1}^A) + \mathbf{1}(A > 0)\Pi^A(s_1^a) + m_1 \\ y_2 &= (1 - s_2^a)w_{u2}(l_{u2}^A + l_{u2}^M) + s_{h2}^a w_{s2}(l_{s2}^M + \mathbf{1}(A > 0)l_{s2}^A) + \mathbf{1}(A > 0)\Pi^A(s_2^a) + m_2 \end{aligned}$$

The household problem is given by:

$$\begin{aligned} \max_{s_1^y} \quad & c_1^\rho + c_2^\rho \\ \text{s.t.} \quad 1 \quad & p_c c_1 + s_1^y C^s \leq y_1(s_1^a) \\ 2 \quad & p_c c_2 \leq y_2(s_2^a) \end{aligned} \tag{5}$$

where $s_1^y = s_2^a$ - the schooling of the youth in the first period is the educational endowment of the household adult in the second period. Since education choices are discrete, households educate their children if the indirect utility from obtaining schooling is greater than that from not doing so:

$$\begin{aligned} V(s_1^y = 1) &> V(s_1^y = 0) \\ V(y_1, w_{s2}, \Pi_s^A(s_2^a = 1), C^s, p_c, \beta) &> V(y_1, w_{u2}, \Pi_s^A(s_2^1 = 0), \alpha, p_c, \beta) \\ \left(\frac{1}{p_c}\right)^\rho [(y_1 - C_s)^\rho + (w_{s2} + \Pi^A(s_2^a = 1))^\rho] &> \left(\frac{1}{p_c}\right)^\rho [(y_1 + \alpha)^\rho + (w_{u2} + \Pi^A(s_2^a = 0))^\rho] \end{aligned} \tag{6}$$

Therefore the probability that a child is enrolled in school can therefore be written as:

$$P(s_1^y = 1) = P(y_1, w_{u2}, w_{s2}, \Pi_s^A(s_2^a = 1), \Pi_s^A(s_2^a = 0), C^s, \alpha, p_c, \beta) \tag{7}$$

where y_1 is current period income, $w_{s2} - w_{u2}$ is the labor market return to obtaining education in the second period, $\Pi_s^A(s_2^a = 1) - \Pi_s^A(s_2^a = 0)$ is the return to education in cultivation activities in the second

[21] There are limited possibilities for households to borrow in order to invest in the education of their children in developing countries (Banerjee, 2004). Since this is a model of household education choices, the assumption that budget constraints clear every period is more reasonable.

period, C^s is the direct cost of schooling, α is the opportunity cost of schooling and β are household preferences.

3.3. **Agricultural Production.** Agricultural production combines labor with land to produce a single agricultural output according to a concave, CRS technology. There is no market for land, therefore landless households do not cultivate land.[22] Two types of labor are demanded in agriculture: unskilled and skilled. Unskilled manual labor undertakes physical tasks such as weeding and harvesting. These types of labor can be conducted by both educated and uneducated workers. Hired and family labor are assumed to be perfect substitutes in this type of work.[23] Supervisorial work, such as deciding which varietals of seeds to sow, can only be conducted by household members.[24] Supervisorial work falls into the skilled category if the household adult conducting it educated, otherwise it is defined as unskilled. Agricultural households choose labor inputs to maximize profits, given their land and education endowments and the agricultural productivity that they face:

$$\begin{aligned} \Pi^A &= \max_{d,g} \quad p^A F(d,g;\theta,A,s^a) - w_u d \\ F(d,g;\theta,A,s^a) &= \theta(g(\delta_1+\delta_2 s^a))^{\gamma_1} A^{\gamma_2} d^{\gamma_3} \end{aligned} \tag{8}$$

where: Π^A denotes profits from cultivating land, net of hired labor expenses; d denotes unskilled labor inputs into production and g denotes supervisorial inputs; p^A denotes the (externally set) national price of agrarian output; θ is the level of agricultural productivity, which varies by region and over time; A denotes household land; s^a denotes the adult's education and w_u is the region level wage for unskilled labor.

Education of the household adult raises the marginal product of supervisorial time - $F_{gs} > 0$. The education levels of manual workers, who conduct tasks such as weeding and harvesting, do not alter their productivity in agriculture.[25] This implies that an educated and uneducated individual working as an unskilled hired laborer in agriculture would be paid the same unskilled wage, i.e. there is no return to education in the hired labor market in agriculture.

I simplify the problem faced by households by assuming an interior solution for supervisorial time.[26] A household's demand for unskilled and skilled labor is given by:

[22] This assumption is supported in the NSSO and REDS data: only 5.5% of landless households cultivate land in 1999. In addition, only 5% of landless households in 1971 owned land in 1982 in the REDS data while 8% of the adult sons of landless fathers owned land in 1999.

[23] This assumption is made for tractability, Bharadwaj (2009) shows that hired and family labor can be considered to be perfect substitutes in some, but not all, agrarian tasks.

[24] This assumption is motivated by the data - less than 0.5% of hired workers in agriculture conduct managerial work.

[25] Foster and Rosenzweig (1993) find this to be the case using piece rates data from the Philippines.

[26] Households in which the marginal product of supervisorial time is greater than the agrarian wage will devote all their time to on-farm supervisorial activity and will not engage in the agricultural labor market. This is more likely to occur in landed households with larger land holdings and endowments of schooling and as the agrarian technological frontier shifts out.

$$
\begin{aligned}
d^u &= d(w_u, w_s; p^A, \theta, A, s^a) \\
&= A\left(\frac{\gamma_1^{\gamma_1}\gamma_2^{(1-\gamma_1)} p^A \theta(\delta_1 + \delta_2 s^a)^{\gamma_1}}{w_u^{1-s^a\gamma_1} w_s^{s^a\gamma_1}}\right)^{\frac{1}{\gamma_3}} \\
g^u &= g(w_u, w_s; p^A, \theta, A, s^a = 0) \\
&= A\left(\frac{\gamma_1^{(1-\gamma_2)}\gamma_2^{\gamma_2} p^A \theta(\delta_1)^{\gamma_1}}{w_u}\right)^{\frac{1}{\gamma_3}} \\
g^s &= g(w_u, w_s; p^A, \theta, A, s^a = 1) \\
&= A\left(\frac{\gamma_1^{(1-\gamma_2)}\gamma_2^{\gamma_2} p^A \theta(\delta_1 + \delta_2 s^a)^{\gamma_1}}{w_u^{\gamma_2} w_s^{(1-\gamma_2)}}\right)^{\frac{1}{\gamma_3}}
\end{aligned}
$$

where d^u is demand for unskilled labor conducting manual work, g^u and g^s is demand for supervisorial labor, conducted by uneducated and educated individuals respectively. w_s is the wage of a skilled worker in the manufacturing sector. In the absence of a manufacturing sector, this is the shadow wage of educated supervisors in the agricultural sector. The agrarian profit function can therefore be written as:

$$
\Pi^A(w_u, w_s; p^A, \theta, A, s^a) = p^A F(w_u, w_s; p^A, \theta, A_h, s^a) - w_u d^u(w_u, w_s; p^A, \theta, A, s^a)
$$

Household profits in agriculture increase in agricultural productivity, the adult schooling endowment, the endowment of land, the price of agricultural products and decrease in wages.

Total unskilled and skilled labor demand in agriculture is given by summing household labor demand across households:

$$
\begin{aligned}
D_u^A &= \sum_{h \in P_A} d_h^u(w_u, w_s; p^A, \theta, A_h, s_h^a) + \sum_{h \in P_A} g_h^u(w_u; p^A, \theta, A_h, s_h^a = 0) \\
G_s^A &= \sum_{h \in P_A} g(w_u, w_s; p^A, \theta, A_h, s_h^a = 1)
\end{aligned}
$$

Under the assumption of a CRS technology, the distribution of land will not affect the demand for labor unless education is unevenly distributed across landed households.

3.4. **Manufacturing Production.** The manufacturing sector uses four inputs in production: unskilled labor, skilled labor and two raw materials. Manufacturing production is increasing and concave in all inputs; the technology is CRS. Labor and raw materials are complements in production. There are two manufacturing industries which vary in the parameters of their production functions, such that the elasticity of substitution between inputs varies across industries. For example, the elasticity of substitution between skilled and unskilled labor inputs or wood and cement varies by industry. Therefore two industries faced with the same vector of input prices will vary in their optimal input choices.

The manufacturing sector competes with the agricultural sector for labor. Unskilled and skilled labor in both manufacturing and agriculture is remunerated at the same rates - w_u and w_s. Only literate individuals can conduct skilled manufacturing jobs. Both industries choose labor and raw material inputs to maximize profits, given the prices and wages that they face:

$$\Pi_j = \max_{N_{uj},N_{sj},r_{1j},r_{2j}} \quad p^M_{Dj}F_j(N_{uj},N_{sj},r_{1j},r_{2j}) - w_uN_{uj} - w_sN_{sj} - p^{r1}r_{1j} - p^{r2}r_{2j} \tag{9}$$
$$F_j(N_{uj},N_{sj},r_{1j},r_{2j}) = N_{uj}^{a_{1j}}N_{sj}^{a_{2j}}r_{1j}^{a_{3j}}r_{2j}^{a_{4j}}$$

where Π_j denotes profits in industry j, N^M_{uj} and N_{sj} are unskilled and skilled labor in industry j, r_{1j} denotes raw material one's inputs in manufacturing; r_{2j} denotes raw material one's inputs in manufacturing; p^M_{Dj} is the price of industry j's output; w_u is the unskilled labor wage; w_s is the skilled labor wage and p^1_r and p^2_r are the price of raw materials 1 and 2 in district d. The prices of the two raw materials do not change over time.[27]

Total demand for labor and raw materials in the manufacturing sector is given by:

$$\begin{aligned}
N^M_u &= \textstyle\sum_{j=1}^2 N_{uj}(w_u,w_s,p^{r1},p^{r2},p^M_{Dj}) && = \textstyle\sum_{j=1}^2 a_{1j}w_u^{-1}y_j\prod_{s=1}^4 p_s^{a_{sj}}\\
N^M_s &= \textstyle\sum_{j=1}^2 N_{sj}(w_u,w_s,p^{r1},p^{r2},p^M_{Dj}) && = \textstyle\sum_{j=1}^2 a_{2j}w_s^{-1}y_j\prod_{s=1}^4 p_s^{a_{sj}}\\
r^M_1 &= \textstyle\sum_{j=1}^2 r_{1j}(w_u,w_s,p^{r1},p^{r2},p^M_{Dj}) && = \textstyle\sum_{j=1}^2 a_{3j}p_{r1}^{-1}y_j\prod_{s=1}^4 p_s^{a_{sj}}\\
r^M_2 &= \textstyle\sum_{j=1}^2 r_{2j}(w_u,w_s,p^{r1},p^{r2},p^M_{Dj}) && = \textstyle\sum_{j=1}^2 a_{4j}p_{r2}^{-1}y_j\prod_{s=1}^4 p_s^{a_{sj}}
\end{aligned}$$

Manufactured goods have a perfect substitute produced in the world market. I place myself in a small open economy setting in which the world-price of the perfect substitute ties down the price of the domestically produced good for industry j, p^M_{Dj}. τ_{jt} denotes the time and industry varying transactions cost which represents tariff barriers and transportation costs between countries. C_{Ljt} represents the time and industry varying per-unit (fixed) cost of complying to labor and industrial regulations. Regulations and tariff barriers are the same in all industries across all economies.

$$p^M_{Djt} = p^M_{Wjt} + \tau_{jt} - C^M_{Ljt} \tag{10}$$

Zero profits conditions imply that industries choose inputs until the price of output is equal to the marginal cost of production.

3.5. **Equilibrium Conditions.** Using the two labor market clearing conditions for skilled and unskilled labor, the two zero profits condition in manufacturing, the ten first order conditions from production and the utility maximization condition, it is possible to solve for the equilibrium quantities and prices of interest, notably the wages of skilled and unskilled workers, the allocation of labor across sectors, raw material inputs and the schooling choices of households.

Equilibrium wages are determined by the labor market clearing conditions:

$$P_u = D^A_u(w^*_u,w^*_s) + N^M_u(w^*_u,w^*_s) \tag{11}$$
$$P_s = G^A_s(w^*_u,w^*_s) + N^M_s(w^*_u,w^*_s) \tag{12}$$

[27]This assumption assumes that the decrease in the total stock of raw materials due to raw material consumption during the period is small.

There are two regions of interest for industries in the manufacturing sector: the corner solution where production is zero and the interior solution. The corner solution occurs in areas where the existing wage of illiterate workers, the marginal product of literate farmers and the prices of the raw materials are sufficiently high that the marginal cost of the first unit of production is greater than the price of the produced, or where the supply of literate workers is zero. In the interior solution, manufacturing output and inputs are chosen such that price is equal to marginal cost. From an initial situation in which one unit of good is produced at a price greater than marginal cost, output in manufacturing expands. pulling labor out of agriculture and into manufacturing. This raises the marginal product of the existing labor in manufacturing. This process continues until the zero profit conditions are fulfilled.

The wages of skilled and unskilled workers are identical if the supply of educated workers at the skilled wage is greater than or equal to the demand for skilled workers at the unskilled wage. Intuitively, this is because literate individuals can conduct the tasks of illiterate individuals. If the demand for skilled workers is greater than the supply of workers at this wage, the skilled wage rises above the unskilled wage. At a skilled wage greater than the unskilled wage, educated workers no longer conduct unskilled tasks. I focus on the more interesting case in which the wages are different and literate workers only conduct work in the skilled labor market.

3.6. **Testable Predictions from the Model.**

3.6.1. *Prediction 1 - Manufacturing Employment.*

1.a.1 An increase in the price of raw materials decreases output in manufacturing, ceteris paribus. The demand for unskilled and skilled labor in manufacturing changes as the price of a raw material increases. There are two effects operating. First, an increase in raw material prices implies that, at the labor market wage, the zero profit condition no longer holds prompting a decrease in manufacturing output. Secondly, a change in the relative price of the raw material and labor induces substitution amongst inputs. If the former effect dominates, an increase in the price of raw materials reduces the demand for skilled and unskilled labor.[28]

$$\frac{\partial y_j}{\partial p_r^n} < 0, \qquad \frac{\partial N_{uj}}{\partial p_r^n} = \frac{\partial N_{uj}}{\partial y_j}\frac{\partial y_j}{\partial p_r^n} + \left.\frac{\partial N_{uj}}{\partial p_r^n}\right|_{y_j=\overline{y_j}} \neq 0, \qquad \frac{\partial N_{sj}}{\partial p_r^n} = \frac{\partial N_{sj}}{\partial y_j^M}\frac{\partial y_j}{\partial p_r^n} + \left.\frac{\partial N_{sj}}{\partial p_r^n}\right|_{y_j=\overline{y_j}} | \neq 0$$

1.a.2 The response of manufacturing labor demand to a change in raw material prices increases in the raw material's share in total output:

$$\left(\frac{\partial N_{uj}}{\partial p_r^n}\right)\frac{\partial}{\partial a_{nj}} > 0, \qquad \left(\frac{\partial N_{sj}}{\partial p_r^n}\right)\frac{\partial}{a_{nj}} > 0$$

Labor demand responses to an increase in raw material prices vary across industries according to the parameters of their production functions. In the extreme case where the share of a raw material in total output is zero the labor demand response to a raw material price change will be

[28] This occurs if the increase in the price of raw materials reduces output sufficiently to overwhelm the increase in labor demand due to an increase in output. This is the case for both industries if the second order effect of a change in equilibrium wages due to a shift of labor out of the agricultural sector does not overwhelm the direct price effect.

zero, ceteris paribus.

Prediction 1b: Response of Manufacturing Employment to Policy Changes

1.b.1 Industrial output responds positively to an increase in the domestic price of the good, induced by a change in industry level policy, ceteris paribus.

$$\frac{\partial y_j}{\partial p_{Dj}^M}\frac{\partial p_{Dj}^M}{\partial \tau_j} > 0, \qquad \frac{\partial y_j}{\partial p_{Dj}^M}\frac{\partial p_{Dj}^M}{\partial C_{Lj}} > 0, \qquad \frac{\partial N_{uj}}{\partial \tau_j} > 0, \qquad \frac{\partial N_{sj}}{\partial \tau_j} > 0$$

1.b.2 Output and labor demand responses to industry policy vary across districts with different raw material prices:

$$\frac{\partial^2 y_j^M}{\partial \tau_j^M \partial p_r^n} < 0, \qquad \frac{\partial^2 N_{uj}^M}{\partial \tau_j^M \partial p_r^n} \neq 0, \qquad \frac{\partial^2 N_{sj}^M}{\partial \tau_j^M \partial p_r^n} \neq 0$$

1.b.3 Industry level policy changes induce own-industry output and labor demand responses that vary according to the region's raw material price and the raw material's share in total output. Within a region with a given raw material endowment, labor demand responses across industries to changes in policy vary according to the industries' technologies.

$$\left(\frac{\partial^2 y_j}{\partial \tau_j \partial p_r^n}\right)\frac{\partial}{\partial a_{nj}} > 0, \qquad \left(\frac{\partial^2 N_{uj}}{\partial \tau_j \partial p_r^n}\right)\frac{\partial}{\partial a_{nj}} \neq 0, \qquad \left(\frac{\partial^2 N_{sj}}{\partial \tau_j \partial p_r^n}\right)\frac{\partial}{\partial a_{nj}} \neq 0$$

3.6.2. *Prediction 2: Wages.*

2.a Agricultural Productivity: An increase in agricultural productivity raises skilled and unskilled wages. Intuitively, an increase in agricultural total factor productivity raises the marginal product of all labor in agriculture.

$$\frac{\partial w_u}{\partial \theta} > 0, \qquad \frac{\partial w_s}{\partial \theta} > 0$$

2.b Increases in skilled and unskilled manufacturing labor demand due to a change in industry level policy raises the wages of skilled and unskilled workers, ceteris paribus.

$$\frac{\partial w_u}{\partial N_u}\frac{\partial N_u}{\partial p_{Dj}^M} > 0, \qquad \frac{\partial w_u}{\partial N_s}\frac{\partial N_u}{\partial p_{Dj}^M} > 0, \qquad \frac{\partial w_s}{\partial N_u}\frac{\partial N_u}{\partial p_{Dj}^M} > 0, \qquad \frac{\partial w_s}{\partial N_s}\frac{\partial N_u}{\partial p_{Dj}^M} > 0$$

Intuitively since labor is the only overlapping factor of production at a local level between agriculture and manufacturing, as labor demand in manufacturing increases labor is pulled out of agriculture, increasing the marginal product of remaining workers in agriculture and raising rural wages. A change in industry policy therefore only affects agriculture through the labor market.

2.c An increase in the demand for unskilled labor in manufacturing raises unskilled wages more than an increase in the demand for skilled manufacturing labor, ceteris paribus. And vice versa for skilled wages.

$$\left.\frac{\partial w_u}{\partial N_{uj}^M}\right|_{N^M=\overline{N}^M} > \left.\frac{\partial w_u}{\partial N_{sj}^M}\right|_{N^M=\overline{N}^M}, \qquad \left.\frac{\partial w_s}{\partial N_{sj}^M}\right|_{N^M=\overline{N}^M} > \left.\frac{\partial w_s}{\partial N_{mj}^M}\right|_{N^M=\overline{N}^M}$$

Intuitively, an increase in skilled manufacturing demand draws educated supervisorial labor from agriculture, reducing the average product of the remaining labor in agriculture. The intuition behind (b) is that an increase in skilled manufacturing demand draws skilled labor from agriculture, raising the marginal product of skilled labor, while an increase in unskilled manufacturing demand draws unskilled labor from agriculture reducing the marginal product of the remaining skilled labor.

3.6.3. *Prediction 3: Incomes.*

3.a Profits from cultivation activities, net of hired labor expenses, are predicted to (a) increase in agricultural productivity and (b) decrease as manufacturing employment expands.

$$\frac{\partial \Pi_h^A}{\theta} > 0, \qquad \frac{\partial \Pi_A}{\partial w_u}\frac{\partial w_u}{\partial N_u^M}\frac{\partial N_u^M}{\partial p_{Dj}^M} < 0, \qquad \frac{\partial \Pi_A}{\partial w_s}\frac{\partial w_s}{\partial N_s^M}\frac{\partial N_s^M}{\partial p_{Dj}^M} < 0$$

The intuition behind prediction (a) is that increases in agricultural total factor productivity raises the return to land, as well as the return to labor. Since there are no land markets, profits from cultivation reflect returns to land and are therefore increasing in agricultural productivity. The intuition behind prediction (b) is that the manufacturing sector competes with the agrarian sector for labor, pushing wages up and decreasing profits in agriculture.

3.b The incomes of all households increase with agricultural productivity: since agricultural productivity raises both unskilled and skilled wages, as well as the returns to land, the incomes of both landed and landless households are increasing in agricultural productivity.

$$\frac{\partial y_h}{\partial \theta} > 0$$

3.c The incomes of landless households increase with shifts out in skilled and unskilled manufacturing labor demand induced by changes in industrial policy. The incomes responses of landed households vary with land endowments.

$$\frac{\partial y_{LL}}{\partial w_u}\frac{\partial w_u}{\partial N_u}\frac{\partial N_u}{\partial p_{Dj}^M} > 0, \qquad \frac{\partial y_{L,(A>A_X)}}{\partial w_u}\frac{\partial w_u}{\partial N_u}\frac{\partial N_u}{\partial p_{Dj}^M} < 0,, \qquad \frac{\partial y_{L,(A<A_X)}}{\partial w_u}\frac{\partial w_u}{\partial N_u}\frac{\partial N_u}{\partial p_{Dj}^M} \neq 0$$

The incomes of net importers of labor, households with land holdings above the import threshold - $A > A_X$ - decrease in manufacturing production, which increases the return to labor but decreases the return to land. The net effect is negative for net importers of labor, while it is likely to be positive for small landowners and decreasing in the size of landholdings.

3.d The difference in incomes between landless and landed households is decreasing in manufacturing labor demand and increasing in agricultural productivity. Inequality decreases more with a shift out in unskilled manufacturing labor demand than with a corresponding increase in skilled manufacturing demand.

$$\frac{\partial (y_A - y_{LL})}{\partial \theta} > 0, \qquad \frac{\partial (y_A - y_{LL})}{\partial N} < 0$$

$$\left.\frac{\partial (y_A - y_{LL})}{\partial N_u}\right|_{N=\overline{N}} < \left.\frac{\partial (y_A - y_{LL})}{\partial N_s}\right|_{N=\overline{N}}$$

3.6.4. *Prediction 4: Education.* A child is enrolled in school if the household indirect utility from having acquired schooling is greater than from having not done so:

$$S_h^y = 1 \quad \text{if} \quad \left(\frac{1}{p_c}\right)^{\rho} [(y_h^1 - C_s)^{\rho} - (y_h^1 - \alpha)^{\rho} + (w_s^2 + \Pi^A(s_h^{a2} = 1))^{\rho} - (w_u^2 + \Pi^A(s_h^{a2} = 0))] > 0$$

4.a *Returns to education in agriculture:* The probability that a landed household invests in the education of the child is increasing in agricultural productivity, in the land wealth of the household and their cross partial.

$$\frac{\partial P(s_h^y = 1)}{\partial \theta} > 0, \quad \frac{\partial P(s_h^y = 1)}{\partial A_h} > 0, \quad \frac{\partial P(s_h^y = 1)}{\partial \theta \partial A_h} > 0$$

Intuitively, an increase in the level of the agrarian technology, such as the introduction to India of High-Yielding Variety seeds in the mid 1960s (Evenson, Foster and Rosenzweig 1996), raises the return to education of landed households. Since landless households do not cultivate land and have no expectations of doing so in the future, an increase in agricultural productivity does not change their returns to education. Landed households respond by increasing their investment in education.[29]

4.b *Returns to education in the labor market:* In this case, an increase in the relative wage of skilled and unskilled workers, holding constant the unskilled wage, raises the probability that a household invests in the education of their child:

$$\frac{\partial P(s_h^y = 1)}{\partial (\frac{w_s^2}{w_u^2})} = \left(\frac{1}{p_c}\right)^{\rho} \rho w_u (\frac{w_s^2}{w_u^2} + \Pi_A(s_h^{a2} = 1))^{\rho - 1} > 0$$

4.c *Household income:* An increase in household income increases the probability that a household invests in the education of their child, ceteris paribus. This prediction is driven by the assumption that the budget constraint clears every period.

$$\frac{\partial P(s_h^y = 1)}{\partial m_h} = \left(\frac{1}{p_c}\right)^{\rho} \rho((y_{h1} - C_s)^{\rho - 1} - (y_{h1} + p\alpha)^{\rho}) > 0$$

4.d *Opportunity Cost:* The probability that a household invests in schooling is decreasing in the opportunity cost of schooling, lost domestic production.

$$\frac{\partial P(s_h^y = 1)}{\partial \alpha} = -\rho(y_h + p_c \alpha)^{\rho - 1} < 0$$

3.7. **From Theory to Empirics.** The 4 sets of predictions put forward in the theoretical model trace out an intuitive order for approaching the empirical analysis. Since education choices vary with wages and incomes, these specifications appear last in the empirical section. In step 1 of the empirical strategy, I use prediction 1 to address the issue of causality - raw correlations between manufacturing employment and wages are likely to reflect labor input choices by the manufacturing sector that are correlated with the attributes of the rural population as well as the direct relationship between the explanatory and outcome variables. I draw upon prediction 1 to pursue an instrumental variable strategy to mitigate concerns of

[29]These prediction has been previously been tested by Rosenzweig and Foster (1996), who find that the returns to schooling increased during the green revolution in India. Areas with the greatest agrarian technical change witnessed the greatest average increases in school enrollment, conditional on the availability of a school.

bias due to the correlation between manufacturing employment and local population and labor market characteristics.

In step 2, I use the estimates from my first stage regression to test prediction 2 - I evaluate whether growth in agricultural productivity and predicted manufacturing employment, broken down by skill, increases unskilled and skilled wages. To capture the relationship between agricultural productivity and rural wages, I use a proxy variables approach in which I put forward a novel empirical measure of agricultural productivity. I test the robustness of my identification strategy by examining alternative explanations for my results. I test prediction 3 in step 3 by examining how the level and distribution of rural consumption is affected according to the source of change; income data is not available to test this set of predictions directly. In step 4 I examine the response of education to changes in wages and incomes. I use the results from the education regressions to disentangle the income, returns to education and opportunity cost effects for educational enrollment. Throughout the analysis, I break manufacturing employment apart into employment of literate and illiterate individuals to test whether the relationship between manufacturing growth, wages and incomes varies with the skill composition of the manufacturing sector.

The data used in this paper can be divided into five categories: industry level data on employment, wages, raw material endowments, factor intensities and industrial policies. The data will be presented as it is used. A more detailed description of the data can be found in the appendix 2.

4. Empirical Strategy

This section describes the empirical methodology used to examine the relationship between rural schooling, wage and income growth using four rounds of district level panel data between 1983 and 1999. The framework presented in section 3 provides a platform for empirically assessing the relationship between rural wages, incomes, rates of return to schooling and education choices. A focal point of the empirical analysis is the differential impact of sector specific changes to labor demand on wages, incomes and rural education choices according to the initial distribution of land and skills in the population.

4.1. **Empirical Specification and Issues.** The model predicts that a shift outwards in skilled and unskilled labor demand induced by changes in industry level policy draws labor out of agriculture and into the manufacturing sector, raising equilibrium skilled and unskilled wages. In this section, I derive the empirical specification used to test this prediction and discuss why running an OLS specification of rural wages on rural manufacturing employment will result in biased estimates of the parameters of interest. In section 5 I put forward an instrumental and proxy variable strategy which I use throughout the paper to correct for this potential bias.

The model applies to a locally clearing labor market. The appropriate empirical analogue in the Indian context is the district. Migration across districts and state boundaries within India is low both in absolute terms and relative to other comparable developing countries. In the ARIS/REDS data collected by the NCAER, 8% of adult males had migrated from their villages of birth, and of these over 80% had migrated within district.

The equilibrium unskilled and skilled rural wages are determined by setting labor demand of skilled and unskilled workers equal to their labor supply at a district level. The linearized supply and demand equations in the agricultural and manufacturing sectors are given by the following system of equations:

$$
\begin{array}{lcllllll}
\text{Labor Demand}^{M}_{Unskilled} & = & a_0 & +a_1 w_{udt} & +a_2 w_{sdt} & +a_3 X_{dt} & +a_4 Z^{M}_{dt} & +u^{a}_{dt} \\
\text{Labor Demand}^{M}_{Skilled} & = & b_0 & +b_1 w_{udt} & +b_2 w_{sdt} & +b_3 X_{dt} & +b_4 Z^{M}_{dt} & +u^{b}_{dt} \\
 & & & & & & & \\
\text{Labor Demand}^{A}_{Unskilled} & = & c_0 & +c_1 w_{udt} & +c_2 w_{sdt} & +c_3 X_{dt} & +c_4 Z^{A}_{dt} & +u^{c}_{dt} \\
\text{Labor Demand}^{A}_{Skilled} & = & d_0 & +d_1 w_{udt} & +d_2 w_{sdt} & +d_3 X_{dt} & +d_4 Z^{A}_{dt} & +u^{d}_{dt} \\
 & & & & & & & \\
\text{Labor Supply}_{Unskilled} & = & P_u & & & & & \\
\text{Labor Supply}_{Skilled} & = & P_s & & & & &
\end{array}
$$

w_{udt} and w_{sdt} denote unskilled and skilled wages in district d at time t, X_{dt} denotes all common determinants of supply and demand, Z^{M}_{dt} are all variables which only enter into the manufacturing demand equation, conditional on equilibrium wages and X_{dt}, such as the price of raw materials and the price of industrial output. Z^{A}_{dt} refers to all variables which only enter into agricultural demand such as the price of agrarian output or agrarian productivity.

Total demand for unskilled and skilled labor is given by the sum of labor demand in the two sectors. Setting demand for skilled and unskilled workers equal to supply, we find the equilibrium wages which clear the market. Rather than solving out for wages as a function of all the exogenous variables, I evaluate labor demand for manufacturing at the equilibrium wages and solve for equilibrium wages as a function of manufacturing employment:

$$
\begin{aligned}
\text{(13)} \qquad w^{*}_{udt} &= \alpha_0 + \alpha_1 E^{M}_{unskilled,dt}(w^{*}, Z^{S}_{dt}) + \alpha_2 E^{M}_{skilled,dt}(w^{*}, Z^{S}_{dt}) + \alpha_3 \theta_{dt} + \\
&\quad + X_{dt}\gamma + Z^{A}_{dt}\eta + Z^{S}_{dt}\kappa + \delta_d + \delta_t + v_{dt} + \\
\text{(14)} \qquad w^{*}_{sdt} &= \beta_0 + \beta_1 E^{M}_{unskilled,dt}(w^{*}, Z^{S}_{dt}) + \beta_2 E^{M}_{skilled,dt}(w^{*}, Z^{S}_{dt}) + \alpha_3 \theta_{dt} + \\
&\quad + X_{dt}\xi + Z^{A}_{dt}\lambda + Z^{S}_{dt}\psi + \delta_d + \delta_t + v_{dt} +
\end{aligned}
$$

where w^{*}_{udt} and w^{*}_{sdt} denote equilibrium unskilled and skilled wages in district d at time t, $E^{M}_{unskilled,dt}$ and $E^{M}_{skilled,dt}$ denote equilibrium employment of unskilled and skilled workers in the manufacturing sector, θ_{dt} denotes the agricultural productivity frontier. X_{dt}, Z^{S}_{dt} and Z^{A}_{dt} capture all other determinants of wages that enter through agricultural labor demand and labor supply, such as the total land cultivated, a measure of the climatic shock experienced by district during that year, the number of households by quantile of the land distribution and the size of the working age population.

The specification above keeps manufacturing employment in the regression, rather than directly examining reduced form wage responses to exogenous determinants of manufacturing. There are two reasons for doing this. Firstly, since skilled and unskilled manufactured employment have the same reduced form determinants, it is not possible to examine how the response of wages to changes in manufacturing employment varies by its skill composition using the reduced form specification. Secondly, examining the specification in this way answers a policy relevant question. Both manufacturing employment and agricultural productivity are major channels through which development policy can and does attempt to alter rural outcomes. For example, policies aimed at establishing special-economic zones in "backward" regions of India have tried to attract particular industries to that region through marketing, investment credits and, most recently, tax exemptions Government of India invests extensive resources into the research and development of new seed varietals and outreach programs which promote the adoption of new technologies [30]

Adding and subtracting $\alpha_2 E^M_{skilled,dt}$ and $\beta_2 E^M_{skilled,dt}$ from these equations and suppressing the notation indicating equilibrium outcomes, we obtain the final estimated equations:

$$
\begin{aligned}
(15) \qquad w_{udt} &= \alpha_0 + \alpha_1 E^M_{total,dt}(w, Z^M_{dt}) + (\alpha_2 - \alpha_1) E^M_{skilled,dt}(w, Z^M_{dt}) \\
&\quad + \alpha_3 \theta_{dt} + X_{dt}\gamma + Z^A_{dt}\eta + Z^S_{dt}\kappa + \delta_d + \delta_t + v_{dt} + \\
(16) \qquad w_{sdt} &= \beta_0 + \beta_1 E^M_{total,dt}(w, Z^M_{dt}) + (\beta_2 - \beta_1) E^M_{skilled,dt}(w, Z^M_{dt}) \\
&\quad + \beta_3 \theta_{dt} + X_{dt}\xi + Z^A_{dt}\lambda + Z^S_{dt}\psi + \delta_d + \delta_t + v_{dt} +
\end{aligned}
$$

Appendix C.1 shows the derivation of equations (15) and (16) in greater detail and defines the α and β terms as a combination of agricultural labor demand wage parameters.[31] The reduced form labor demand and supply parameters indicate that $\alpha_1 > 0$, $\alpha_1 > \alpha_2$ and that $\beta_2 > 0$, $\beta_2 > \beta_1$, if own price labor demand responses are greater than cross-price responses. Therefore we expect the estimated coefficients on skilled manufacturing employment to be negative when the unskilled wage is the dependent variable - equation (15) - and to be positive when the skilled wage is the dependent variable - equation (16).

Regressing these equations directly is likely to lead to biased estimates of the parameters of interest. The simultaneous determination of manufacturing employment and wages implies that any locally unobserved common determinants of agricultural labor demand, labor supply and manufacturing labor demand - will be correlated with both wages and manufacturing employment. For example, changes in the quality of electricity services may increase the demand for labor in the agricultural and manufacturing sectors. Therefore the OLS estimates of equations 15 and 16 will yield biased estimates of the parameters of interest. I use an instrumental variables approach to obtain unbiased estimates of these parameters. I use a proxy variables approach to measure agricultural technological change which avoids using the actual yields in a location. The measure is discussed in section 6.2.

[30] An alternative approach would be to directly examine how wages respond to a change in agricultural employment relative to a change in manufacturing employment. Under the assumption of labor market clearing, the coefficients on agricultural and manufacturing employment will capture a measure of the labor supply elasticity which should be the same for both variables. This approach however doesn't directly get at the mechanisms through which a change in policy alters wages.

[31] The derivation of the equations in the appendix is in a more general form, in which labor supply is responsive to changes in equilibrium wages.

5. Step 1: IV Strategy

This section describes the methodology used to estimate the causal effect of a shift out in total and skilled manufacturing labor demand on rural manual and skilled rural wages. As discussed in the section above, a concern is that manufacturing employment is likely to be correlated with unobserved rural labor market and population attributes which are themselves determinants of wages. I use a Two-State Least Squares (TSLS) strategy to correct for the potential bias on manufacturing employment in the wage regressions.[32] The model predicts that, within a district, labor demand responses across industries to changes in policy vary according to the industries' technologies. Changes in industry level policy affect the marginal product of labor in agriculture only through the employment channel in the local labor market. The TSLS strategy builds upon this prediction. I motivate why the generated variation is uncorrelated with the error term in section 5.2 and explain why this strategy is likely to have predictive power in section 5.1. I present and interpret the results from my first stage estimations in section 5.4.

5.1. **Why are these instruments likely to be correlated with manufacturing employment?** The Two-Stage Least Squares (TSLS) strategy is motivated by my framework and by two different literatures. The first literature examines the relevance of "natural advantages" for determining (the share) of employment in a given location by industry (Krugman, 1991; Ellison and Glaeser, 1997; Romalis, 2004). The second literature focuses on the effects of the elimination of industrial regulations and trade barriers on manufacturing output, employment, investment and productivity in India (Aghion, Burgess, Redding and Zillibotti, 2009; Chari, 2007; Topalova, 2004; Mehta and someone, 2002). The empirical model from the first stage is built on the overlap between these two literatures. If "natural advantages" contribute to the explanation of the static distribution of employment within an industry across districts of India, the interaction of natural advantages and policy changes should capture differential responses of industrial employment across districts with different initial concentrations of an industry.

5.2. **Validity of Instrumental Variables.** In this section, I discuss the validity of the TSLS strategy. The framework predicts that, within a district endowed with a given vector of raw materials, employment responses across industries to changes in policy vary according to the industries' technologies. This implies the following empirical specification:

$$
\begin{aligned}
(17)\qquad E^M_{total/skilled,idt} &= \beta_0 + \beta_1\tau_{it} + \beta_2\tau_{it} * s_i + \beta_3\tau_{it} * s_i * n_d + \beta_4\tau_{it} * n_d + \beta_5 s_i * n_d + \\
&\quad + \beta_6 s_i + \beta_7 n_d + \beta_8 X_{dt} + \delta_d + \delta_t + u_{idt}
\end{aligned}
$$

[32] In using a two-stage least squares specification, I am explicitly not estimating a structural equation for industry level employment. To identify the parameters that describe an empirical labor demand equation, I would have to include the price of all inputs including the wage, the price of labor. The first stage regression I use omits all terms related to local labor market wages. I use two-stage least squares to identify the coefficients of interest by generating variation in the endogenous variable, manufacturing employment, that is uncorrelated with unobserved determinants of wages. Conditional on the validity of my instruments, this approach will identify the coefficients on manufacturing employment in equations (15) and (16).

where $E^M_{total/skilled,idt}$ is total or skilled manufacturing employment in industry i, district d at time t, τ_{it} are industry level tariffs at time t, s_i is a measure of an industry's use of a given input, and n_d is a measure of the district level endowment of that input.[33][34]

The first stage strategy can be thought of as a difference-in-difference-in-difference strategy. To consistently identify the coefficients on total and skilled manufacturing employment in equations (15) and (16), the excluded variables should be uncorrelated with all unobserved determinants of district level wages. To separately identify the coefficients on total and skilled manufacturing employment, the instruments must also generate sufficient independent variation in these two variables; I discuss the condition required to do so in section 5.3. In this section I focus on why, *conditional on lower order interactions*, the triple interaction between industrial policy, industrial resource use and regional resources, can be treated as excludable.

The framework implies that the only channel through which changes in industry policy affects the agricultural sector is through the rural labor market. This would imply that all policy interactions can be treated as excludable.[35] In my empirical strategy only the triple interaction term are treated as excludable. All lower order terms are included in the second stage regression; intuitively this allows industrial policy to alter agricultural productivity through *nationally* traded manufacturing inputs that enter into the agricultural production function, such as fertilizers. This effect is allowed to vary according to the region's raw material endowments and the region's distance from trading hubs (major ports/state capitals). The identifying assumption needed is that labor is the only *locally* traded overlapping input between the agricultural and manufacturing sectors, and that output markets for both agricultural and manufacturing commodities do not clear at a local level. I examine the robustness of my results to these assumptions in section 6.5.

---

[33] A conceptually equivalent approach would be to aggregate industry-district level employment across industries to arrive at a district level strategy:

$$\begin{aligned} E^M_{total/skilled,dt} &= \gamma_0 + \gamma_1 \overline{\tau}_t + \gamma_2 \overline{\tau s}_t + \gamma_3 \overline{\tau s}_t * n_d + \gamma_4 \overline{\tau}_t * n_d + \\ &\quad + \gamma_5 \overline{s} * n_d + \gamma_6 \overline{s} + \gamma_7 n_d + \gamma_8 X_{dt} + \delta_d + \delta_t + v_{dt} \end{aligned}$$

where $\overline{\tau}_t = \frac{1}{J}\sum_{i=1}^{J} \tau_{it}$ captures the average level of the policy of interest in the economy at time t, $\overline{\tau s}_t = \frac{1}{J}\sum_{i=1}^{J} \tau_{it} * s_i$ is a weighted average of industry tariffs at time t, and $\overline{\tau s}_t * n_d = \frac{1}{J}\sum_{i=1}^{J} \tau_{it} * s_i * n_d$ is the interaction of resource-use weighted policy with the raw material resource endowment of the district. While the conceptual source of variation is the same, by aggregating tariffs across industries the variation in timing and magnitude is compressed. For this reason, the district-industry regressions are my preferred specification but I report results from both sets of regressions for clarity.

[34] In the first stage specification above, I am assuming that industry employment responds to changes in their own policies but not in response to changes in the policies of all other industries. In practice, this is a strong restriction on the specification. Changes in policy in industry i are likely to affect industry j through the price of overlapping inputs, such as labor. For these omissions to invalidate my identification strategy, the variables capturing how industry i's tariff changes affects j's employment need to be correlated with the triple interaction term *as well as* entering directly into the wage regression. Since the timing and magnitude of policy shifts was random across industries during the import tariff reform (Topalova, 2005), it is unlikely that this is the case. However, in order to ensure that my results are robust to this story, I run two specification tests. In the first, I allow industry i to be affected by industry j's tariff change through the labor market. Appendix C.3 shows how this changes the district and district-industry specifications. In the second, I take a more general approach and allow industry i's employment to vary with the policy changes across all industries. For reasons of parsimony, I use 2-digit industry categories in this specification.

[35] If this were the case, both the triple order interaction terms and the lower order interaction terms could be treated as excludable from the wage regression. However, this is a strong assumption since it requires that changes in the price, quality, quantity or variety of products induced by variations in industrial policies will have no direct impact on agricultural productivity.

The next paragraphs explain why the triple interaction terms are excludable from the wage regression while the lower order terms are not. The lower order terms are considered in three groups: those that vary across districts but not over time, $s_i * n_d$, s_i and n_d; those that vary over time but not over districts, $\tau_{it} * s_i$, τ_{it}; and those that vary both over time and across districts, $\tau_{it} * n_d$. For each of these three groups, there are reasons to suspect that they may be capturing local agricultural responses to changes in industrial policy or local variation in agricultural productivity. Conditional on these lower order interactions, the triple interaction term captures variation in employment that is driven only by differences in technologies across industries and in industry policy across time.

District level time-invariant resource endowments (n_d and $\overline{m} * n_d$) are likely to be correlated with a number of time invariant determinants of agrarian wages.[36] For example, the mineral endowments of a region are correlated with its soil quality, an unobserved input in the agricultural production function. Chemical and mechanical weathering at the Earth's surface decomposes rocks and the minerals within those rocks to form soil (Kessler, 1996). The soil's nutrient content, profile and composition will depend on the texture, structure, chemical and mineralogical composition of the parent material, as well as the climate and topography of the region. Therefore, the underlying quality of the soil for agricultural purposes, as well as the type of crops that can be cultivated and the type of fertilizers that should be used are likely to be highly correlated with the mineral composition of ores in a district. Variation in raw material endowments across districts cannot therefore be considered to be excludable to the wage regression, since it is highly correlated with unobserved differences in an input into agricultural production. These time invariant district level terms are therefore captured by district fixed effects.

Explanatory variables which vary over time but not across districts capture aggregate effects of changes in policy. For example, the removal of import tariff barriers has been found to alter both the price and number of imported intermediate goods sold, as well as the number of final goods produced by domestic firms within India (Goldberg et al, 2009). Expansions in the number and price of intermediate and final goods due to these changes in policy is likely to lead to farmers altering their optimal input mix. If the market for final manufacturing products is at a national rather than regional level, I can capture these responses using year fixed effects. To capture variations in a region's exposure to imported goods due to the distance from the port of entry, I include the interaction of average tariff changes with the distance from the nearest major port in all specifications. If manufacturing products are locally distributed, for example fertilizer products are sold to cultivators in the immediate vicinity of the plant, then there may be additional local price effects. I address this concern by examining firstly whether the employment effect is driven by industrial sectors which are upwardly linked to the agricultural sector, and secondly examine whether the instruments are partially correlated with regional prices of inputs such as crop seeds and fertilizers.[37]

[36]Since I aggregate over industries to get my measure of predicted employment at a district level, the district level measure of the interaction of raw material endowments with industry level resource usage ($m_i * n_d$) is a constant proportion of the district-level time-invariant resources, $\overline{m} * n_d$.

[37]These effects are, in any case, likely to result in a downward rather than upward bias on my estimate, and are therefore unlikely to be driving my results.

The interactions between time varying policy and natural resource endowments capture how the response to changes in policy varies with regional resource endowments. A concern is that aggregate policy changes may affect regions differentially according to their raw material endowments. Extending the soil quality example above, if changes in aggregate policy reduce the price of rice seeds, then we would expect agricultural profits in areas in which rice is cultivable and in areas in which rice is not cultivable to be differentially affected by aggregate policy changes. A further concern is that time trends in wages may vary with observed raw material endowments. For example, regions which are densely wooded are likely to have fewer urban settlements. The baseline size of the urban population is likely to be associated with differential time trends in rural wages. The policy and endowment interactions are therefore included in the second stage specification. To address the concern that district with different initial underlying characteristics may be experiencing differential trends in wages and industrial growth, I allow for regional time trends in my specifications.

The interaction between industrial policy, industrial resource use and regional resources, $\tau_{it} * s_i * n_d$, captures how employment responses to national policy vary across industries according to their factor use and within districts endowed with a given bundle of raw materials. The variation used to identify the coefficient on this interaction arises from technological differences across industries in the raw materials used and from differences in the timing and depth of policy changes across industries. By definition, the variables inducing this identifying variation are the same across all locations and are therefore in themselves uncorrelated to the attributes of the local rural population or local labor markets.

Since the lower order terms capture employment responses that are likely to be correlated with local attributes, the triple-interaction captures variation in employment that is driven only by variation across industries in technologies and policies, conditional upon the lower order terms. Therefore, these terms are unlikely to reflect any local unobserved variation in baseline conditions that may be correlated with agrarian wages nor are they likely to be correlated with time trends/aggregate policy responses that vary according to district level baseline conditions. In table A.2, I check whether this is the case by examining whether growth in the triple interaction terms is correlated with baseline characteristics, conditional upon changes in the lower order terms. The second order terms are not systematically and jointly correlated with pre existing district level characteristics, such as the level of wages in 1980, male or female literacy or the size of the population. To check the second concern, I examine the robustness of my results to including region level growth trends and time trends interacted with baseline characteristics.

The coefficients on manufacturing employment in the second stage in equations (15) and (16) are identified under the assumption that unobserved district-specific time varying shocks that affect agricultural wages are uncorrelated with changes in raw material weighted policies over time. However, the triple interaction will also capture all other industry-region level responses to a change in tariff changes, such as changes in output, input demand for all inputs other than labor and prices. A key identifying assumption of my approach is that the agricultural and manufacturing sectors only overlap through the labor market. If agriculture and manufacturing are linked through other input markets or through vertical or horizontal channels, the triple-interaction term may be correlated with region-time varying input and/or output prices that affect agricultural wages. I test for the robustness of my results to other potential channels in my robustness section.

5.3. **Identification of the Effects of Skill Composition.** By exploiting variation in the timing and degree of tariff movements across industries with different proportions of literate workers, I am able to separately identify the impact of shifts in demand for literate versus illiterate workers. To see this, imagine that there are two industries, industry 1 and 2. Industry 1 employs 50% literate workers and 50% illiterate workers. Industry 1 experiences a larger drop in tariff barriers 3 years before industry 2. Industry 2 employs only illiterate workers. The direct effect of changes in industry 1's tariffs will shift both the demand for literate and illiterate labor in locations in which industry 1 is predicted to be present 3 years before industry 2's literate labor demand curves are shifted. Both industries will of course be affected by both sets of tariff reforms through general equilibrium effects (the most simple of which works through equilibrium wages at a district level), this will not violate the exclusion restrictions on my set of instruments as long as labor is the only district level interaction between agriculture and manufacturing.

In order for this strategy to separately identify variation in skilled versus unskilled employment, industries must vary in their technological requirements for skilled labor. The education and skill composition of the manufacturing workforce in India does appear to vary substantially by industry. While part of the variation among industries in the proportion of skilled or literate workers in the workforce may be attributed to variations in tasks by industry, it is also likely to be a function of the relative wages of literates and illiterates in a given location. To examine how much of the variation in the proportion of illiterates in a district-industry cell is industry specific and how much is due to regional variation in relative wages, I conduct an Analysis of Variance of region-industry level skill and education proportions using region and industry fixed effects. If industry level factors (e.g. technology, type of work) contribute more to the explanation of the literate proportion than location specific factors, then industry level dummies should be able to explain a greater fraction of the variance in the skilled/literate proportion than location specific dummies. If district level factors provide a better explanation of the skilled/literate proportion (e.g. due to differences across regions in the relative wages of skilled and unskilled workers), then these factors should be common among all industries employing labor within that district. For example, in areas in which unskilled wages are low, both relative to the labor market wage for literate individuals in that district and in absolute terms, all firms within that district will be likely to employ a greater proportion of illiterates.

Table 3 reports the results from the variance decomposition for 1983 and 1999. The industry level dummies capture 21.5% of the variance in the proportion of literates and skilled workers employed; they are jointly statistically significant at a 1% level with an F-Statistic of 17.32 for literatesvariation in the proportion of skilled workers across industries may be due to variations in the relative wages of individuals by education, industry level factors do also play a role and are able to explain a large and statistically significant fraction of the variance.

5.4. **First Stage Regressions: Results.** In section 5.4.1, I conduct a basic validation of the framework by examining whether the interaction between raw material endowments and the intensity with which an industry uses a raw material helps to predict industry level employment across and within NSS regions.[38] In section 5.4.2 I discuss the results from the first stage regressions. Predicted manufacturing employment from these regressions is then used to empirically evaluate wages responses to exogenous variation in the size of the manufacturing labor force in section 6.

5.4.1. *Do resource endowments help to explain the static geographic distribution of employment?* The results from the static specification are presented in column a of table 5. All specifications include lower order interaction terms and region fixed effects. Standard errors are clustered on a region level. The dependent variable is the log of total region industry employment; factor intensity is measured using a dummy indicator.[39] The variables are presented in descending order of their degree of transportability. Appendix table A.2.3raw materials. The cost to the consumer of minerals with a low unit value increases with distance to the place of use and the difficulty of transporting the material (the lower the density). Consequently, low unit value commodities are normally of no or little value unless they are processed close to their location. I focus on these relatively immobile inputs to explain the geographic concentration of industries.

The interaction between an industry's intensity of raw material use and the district level endowment of the raw material captures the prediction that employment in industries that more intensively use a raw material are more responsive to variations in the endowments of factor k than industries which use that raw material less intensively. The results support the hypothesis that raw material endowments contribute to the explanation of the geographic distribution of employment. Employment is greater in those industries which use the resources available in a given district; in a given industry, employment is greater within the districts which are endowed with the resources used by that industry.

The first two variables reflect agricultural inputs. In areas well endowed in groundwater resources (at the 75th percentile of endowments), employment in industries that use agricultural inputs is predicted to be 67% greater than in industries which don't use agricultural inputs. Similarly, employment in industries which use agricultural inputs is 16% greater in regions well endowed with groundwater resources than in regions that are not (regions at the 25th percentile of groundwater endowments). The next six variables reflect the availability of industrial raw materials in a district: wood, commonly used minerals (ceramics, construction materials, chemical and metals) and energy - coal deposits and electricity prices. The results indicate that the availability of bulky materials such as wood, construction minerals and ceramics minerals are positively and significantly correlated with employment. For example, in regions within the top quartile of ceramics materials, employment in industries using these materials is 19.4% greater than in those industries which do not use ceramics materials. The coefficients are largest for the commodities which are bulkiest and least cost effective to transport. By contrast, the availability of materials whose price and density combination make them more worthwhile to transport do not appear to have a positive impact on the location of industries which use those materials most intensively.

[38]Badiani (2009) conducts a more thorough empirical analysis of the static distribution of industrial employment.
[39]A more complete variety of specifications nab discussion of results can be found in Badiani (2009).

5.4.2. *Does the interaction capture differential employment responses to policy changes?* Tables 5 present Fixed-Effect specifications of changes in district-industry level employment over time. Columns (b) through (e) present estimates of the employment response to changes in industrial regulations while columns (f) through (i) presents results capturing the response to changes in import tariffs. All specifications include lower order interaction terms, region fixed effects and time variant district characteristics including a region and time varying measure of agricultural productivity, the proportion of illiterate males and females in the working age population, and the logarithm of the population, split by landless and landownership quartiles.

The coefficients on the triple interactions confirm for the most part the hypothesis that employment responses to changes in industry time varying import tariffs and licensing reforms vary by industry resource use and a region's resource endowments. The lower order interaction terms are not shown for reasons of parsimony, therefore the heterogeneous employment response of industries across regions and industries is best illustrated by examples. Within wood intensive industries the employment response to a 65% drop in tariffs (the average decrease between 1983 and 1999) results in a greater drop in employment in areas better endowed with wood resources: employment is predicted to decrease by 78.2% more in the top quartile of wood areas than in the bottom quartile of wooded areas. The corresponding drop across wooded areas within non-wood intensive industries (those in the bottom quartile of wood usage) is 7.3%. Similarly, within the top quartile of wooded areas, the wood intensive industries saw employment drop by 34.4% more than non-wood intensive industries.

6. Step 2: Wage Regressions

In this section I test the wage predictions of the model, prediction 2. In the first step of the empirical analysis, I discussed the difficulties of estimating the causal impact of an increase in manufacturing employment on wages, incomes and education and put forward an instrumental variables strategy. In this section, I use the IV strategy from step 1 to estimate the response of wages to changes in agricultural productivity and predicted manufacturing labor demand, broken down by skill. In section 6.4, I quantify what proportion of wage growth between 1983 and 1999 can be attributed to these two drivers of wage growth; I further evaluate the relative impact of agricultural and industrial policy in monetary terms. In section 6.5, I examine the robustness of my results to alternative explanations.

6.1. **Empirical Specification.** Prediction 2 suggests that rural wages increase in both agricultural productivity and manufacturing employment. An increase in unskilled manufacturing employment is predicted to raise the unskilled wage, and to do so by more than an increase in skilled manufacturing employment. The skilled wage is predicted to increase as skilled manual employment increases and to do so by more than in response to an increase in unskilled manufacturing employment.

The empirical strategy for testing these predications was derived in section 5. For convenience, equations (15) and (16) are restated below:

$$(18) \qquad w_{udt} = \alpha_0 + \alpha_1 \hat{E}^M_{Total,dt} + (\alpha_2 - \alpha_1)\hat{E}^M_{skilled,dt} + \alpha_3\theta_{dt} + A_{dt}\kappa + \delta_d + \delta_t + \epsilon_{dt}$$

$$(19) \qquad w_{mdt} = \beta_0 + \beta_1 \hat{E}^M_{Total,dt} + (\beta_2 - \beta_1)\hat{E}^M_{skilled,dt} + \beta_3\theta_{dt} + A_{dt}\psi + \delta_d + \delta_t + \varepsilon_{dt}$$

where w_{udt} and w_{sdt} denotes equilibrium unskilled and skilled wages in district d at time t, $\hat{E}^M_{Total,dt}$ and $\hat{E}^M_{skilled,dt}$ denotes total and skilled employment in manufacturing, predicted using the estimated coefficients of the first stage regression. denotes skilled employment, which is measured as literate workers in manufacturing. Alternative measures include workers with at least primary education working in manufacturing, and those conducting white collar and blue-collar skilled occupations. θ denotes a measure of the agricultural productivity frontier, discussed below. A_{dt} captures all other determinants of wages that enter through agricultural labor demand and labor supply, such as the total land cultivated, a measure of the climatic shock experienced by district during that year, the number of households by quantile of the land distribution and the size of the district level population. District dummies absorb all location invariant factors while year dummies capture all common year effects.

The α and β terms are combinations of agricultural labor demand and labor supply parameters. As shown in appendix xx, we predict $\alpha_1 > 0$, $\alpha_1 > \alpha_2$ and $\alpha_3 > 0$; and $\beta_2 > 0$, $\beta_2 > \beta_1$ and $\beta_3 > 0$. Therefore we anticipate that the estimated coefficient on skilled manufacturing employment is negative (positive) in the unskilled (skilled) wage regression.

6.2. **Data.** To test prediction 2, I use measures of wages, agricultural productivity, characteristics of the rural population, and the predicted measures of manufacturing employment from the first stage regressions. Further information on the data used in this paper can be found in appendix 2.

Agrarian labor market wages, collected from the Agricultural Wages of India (Government of India, various years), are the primary measure of unskilled wages. Wage workers in the agricultural sector are disproportionately illiterate relative to the population average and conduct manual physical tasks, as discussed in section 2.2. This measure therefore constitutes a good proxy for the wage obtained by individuals working in unskilled occupations which don't require any educational qualifications. As a robustness check, I impute wages of illiterate wage earners in the district using data collected by the NSSO.[40]

I put forward a new approach for measuring agricultural technological change. Changes in productivity in agriculture can be captured using growth in agricultural yields over time (Foster and Rosenzweig (2004), Jayachandran (2006), Lanjouw and Murgai (2009)). This measure however will reflect a combination of technical change as well as endogenous responses of inputs to that change (such as labor, irrigation and fertilizer). Technical change in agriculture constitutes a moving out of the agricultural productivity frontier. Foster and Rosenzweig (1996) measure technical change by estimating the agricultural technology before and after the onset of the green revolution. This approach to measuring technical change

[40]This measures excludes manufacturing sector workers since the excluded variables will be directly correlated the marginal product of workers in the manufacturing sector, for example through changes to the optimal mix of inputs.

requires detailed cultivator level data on agricultural production over time, which is unavailable for the time period I examine.

Agricultural technological change is measured by combining information on the technological frontier of crops across India with variation in the type of crops that can be grown within India. The maximum of the India-wide yield in a particular crop as a measure of it's frontier. For example, districts in the Punjab are regularly at the yield frontier in rice. I combine this with a district level time invariant measure of the physical suitability of crops to local agro-climatic conditions, computed by the Food and Agriculture Organization (FAO). The Global Agro-Ecological Zones (GAEZ) project run by the FAO imputes an index of the suitability of a particular region to growing rice, maize, wheat, cotton, sugar, pulses and sorghum. The index combines information on climatic conditions, soil quality and terrain to make these predictions. For example sorghum is not physically suitable for cultivation to parts of Rajasthan but is well suited to most of Andhra Pradesh. I combine the physical suitability of land to the cultivation of different crops with the crop level yield frontier in India to capture the maximum potential productivity of a district at a point in time, where c represents crops, and p is the price of the agricultural product:

$$\begin{aligned} (20) \qquad \theta_{dt} \;=\; & \sum_{c}^{C} \text{Maximum National Yield in Crop}_{c,t} * \\ & *\text{Proportion of District Most Suitable to Crop}_{c,d} * \text{Crop Price}_{c,1980} \end{aligned}$$

Finally, three measures of weather fluctuations in an agricultural year are computed using weekly rainfall data collected by the Indian Meteorological Department. The measures of rainfall are total monsoon rainfall, the square of total monsoon rainfall and a weather "shock" variable taking the value of -1 if total rainfall is 50% below it's long term mean, and 1 if it is 50% above. Data on the demographic characteristics of the rural population, such as the size and age structure of the working age population, were put together from various rounds of the NSSO.

6.3. **Results.** Table 6 and 7 present the estimates of equation (15), where panel A of table 6 puts aside the distinction of skilled and unskilled manufacturing employment. The results for (16) are presented in table 8. In all tables, the results from the Fixed-Effect (FE) specification are presented in column (a), column (b) presents the Fixed-Effect Instrumental Variables (FE-IV) specification. Columns (c) through (g) present specification and data checks using the FE-IV specification. The FE coefficient captures two effects. Firstly as industrial employment grows, labor is drawn out of agriculture raising the marginal product of the remaining agrarian workforce. Secondly, industries are attracted to areas with lower wages and wage growth profiles. The FE-IV strategy isolates the first effect; therefore the FE-IV coefficient is expected to be greater than the FE coefficient. For example, in the Indian context, cross subsidization between agricultural and industrial electricity prices implies that the two rates are negatively correlated. Manufacturing is deterred from places with low agricultural electricity prices, which are in turn likely to raise productivity in agriculture. Another example is that manufacturing employment may be higher in locations with a large and growing working age populations; for a given quantity of land, this is likely to be associated with reductions in wages in the agricultural sector. The first effect is predicted to be positive and the second to be negative.

The FE coefficient on manufacturing employment reported in column (a) of table 6.A shows no statistically significant relationship between manufacturing employment and unskilled wages. The coefficient estimated using the IV strategy in column (b) is greater than the FE coefficient, confirming the hypothesis that the coefficient estimate on manufacturing employment in the FE specification was downward biased. A shift in manufacturing employment is predicted to have a positive impact on male unskilled wages - $\beta_1 > 0$ - as predicted by the model. A 10% increase in employment in manufacturing is estimated to raise wages by between 1.1% and 1.7%. To verify that the result is robust to the aggregation approach, I estimate the coefficients using both the region-industry column (b) and the district level approach column (c).[41] A result of a similar magnitude is found by instrumenting employment at a district level.[42] Conditioning on region time trends in column (d) increases the estimated wage impact slightly - this suggests that growth paths in manufacturing employment and wages are negatively correlated. The estimated coefficients are of a similar magnitude to the coefficient of 0.9% found by Rosenzweig and Foster (2004) using village level data and a different instrument strategy.

An increase in agricultural productivity is predicted to have a positive impact on agrarian unskilled wages - $\alpha_3 > 0$. The increase in the coefficient on agricultural productivity between columns (a) and (b) suggests that manufacturing employment appears to orientate itself towards areas where, ceteris paribus, the unskilled labor wage is already relatively low. The coefficient estimates suggest that an increase in the agricultural productivity of a region exerts a positive impact on agrarian wages - a 10% increase in the technological frontier increases wages by approximately 4%.

Table 6.B reports estimates which break manufacturing employment down by skill. The model predicts that the coefficient on total manufacturing employment should be positive, $\alpha_1 > 0$, while that on skilled employment should be negative - $(\alpha_2 - \alpha_1) < 0$. The intuition for this prediction is that drawing unskilled labor from agricultural raises the marginal product of the remaining unskilled agricultural workers more than drawing out skilled agricultural laborers. The reduction of educated labor in agriculture dampens this effect through a decline in the average productivity of labor in agriculture.

The FE-IV coefficient estimates reported in column (b) confirm the predictions of the model: the wage effect of an exogenous change in unskilled employment is estimated to be greater than that of a change in skilled employment. The coefficients on total and skilled manufacturing employment are of opposite signs and are both significantly different from zero. A 10% increase in unskilled (purely illiterate) manufacturing employment raises agrarian wages by 3% while an increase in purely skilled (literate) manufacturing employment raises wages by 0.4 to 0.6%. Similarly, a 10% increase in manufacturing employment of individuals with less than primary education raises wages by 3% whereas increasing the employment of those with primary education and above raises wages by 0.5%.

[41]Column (b) reports the results from the district level specification, in which the log of district level manufacturing employment is instrumented using the interaction of local resource endowments with factor use weighted tariff changes as instruments. Columns (c) presents the results from instrumenting the log of industry-region level employment using the interaction of local resource endowments, industry level tariffs and industry resource usage as instruments. Predicted industrial employment is then aggregated to a region level.

[42]I do a Hausman test to verify whether the results are within sampling error of each other, where the null-hypothesis is that the district-industry exogeneity assumption is correct and that this approach is more efficient. Under the null hypothesis the test statistics is distributed as a chi-squared statistics with 20 degrees of freedom; the critical value at a 5% level is therefore 31.4. The Hausman test statistic is 14.09, therefore I do not reject the null hypothesis.

The opposite direction of the movements of the coefficients between the FE and FE-IV specifications indicate that the coefficient on total employment was downward biased, while the estimated effect of a shift out in purely manufacturing employment remains broadly unchanged. This indicates that selection on the basis of the unobserved determinants of unskilled wages occurs most predominantly in the unskilled manufacturing sector.

Table 7 reports estimates of equation (16) in which the dependent variable is skilled wages. The model predicts that an increase in skilled manufacturing employment should have a greater impact on skilled wages than an increase in unskilled manufacturing employment, therefore the coefficient on skilled manufacturing employment should be greater than that on total manufacturing employment. Skilled manufacturing employment has a positive and statistically significant impact on skilled wages - a 10% increase in skilled employment raises skilled wages by 1.2%, while an increase in unskilled employment has no statistically significant effect. Removing the potential bias in the coefficients by using a FE-IV specification, we see that the impact of a 10% increase in skilled employment rises to 2% while that of unskilled employment remains unchanged in column (b). The direction of movement of coefficients between the FE and FE-TSLS specification indicates that purely unskilled manufacturing employment does not select into locations on the basis of the unobserved determinants of skilled wages, mirroring the observation for skilled employment and unskilled wages. Agricultural productivity is positively correlated with skilled wages, although the estimated coefficient is not statistically significant. A 10% rise in productivity is estimated to raise skilled wages by 1.4%.

6.4. **Interpretation of Results from Wage Regressions.** Between 1983 and 1999, agricultural potential increased by 54% or at an annualized rate of 2.51%. Growth in agricultural potential resulted in a 21.6% increase in real wages over the period, accounting for just under half of the total wage growth of 52% over the same period. Over the same period, total manufacturing employment (both skilled and unskilled) has increased by 40%, although the proportion of individuals employed in the manufacturing sector has increased by approximately 14%. Of this, the majority of employment growth consists of skilled employment, which grew by 70% over the period compared to an increase of 12% in unskilled employment. The changes in manufacturing employment seen between 1983 and 1999 raised the wage of rural unskilled workers by approximately 7%, but raised the wages of skilled workers by approximately 15%. Had the increase in manufacturing workers seen over the period consisted of only unskilled workers, the coefficient estimates suggest that the wages of rural unskilled workers would have increased by 16%. The skill-biased employment growth witnessed in India (Kochhar, 2006) has implied that there has been relatively little effect of changes in the size of this sector on the wages of unskilled individuals. This is an observation that has been repeatedly asserted in the policy literature. To my knowledge, the estimates in this paper are the first to empirically validate this observation.

The relative impact of agriculture and manufacturing are opposite to those found in the existing literature. Using household panel data between 1982 and 1999, Foster and Rosenzweig (2004) find that a 74% increase in agricultural productivity over the period resulted in a 47.7% increase in village agricultural wages. A 900% increase in factory workers resulted in a 93% increase in the village wage. Their results therefore suggest that manufacturing growth contributed more to wage growth over the period than agricultural productivity, while the results in this paper suggest the opposite. The point estimates in the two papers are of similar magnitudes; the primary explanation for the differences estimated is that the trends in manufacturing employment seen in the REDS data do not match those seen in any other nationally representative data-sets.

6.5. **Robustness Checks.** In this section, I verify that the coefficients estimated in the wage specifications are indeed driven by movements in labor between the agricultural and manufacturing sector and are not driven by alternative explanations. The coefficients may capture overlaps between the manufacturing and agricultural sector in other locally clearing input and output markets, such as local markets for agricultural or manufacturing products (Adhvaryu et al, 2009.) or groundwater inputs (Keskin, 2009). In this section, I discuss the robustness of my results to these alternative explanations.[43]

6.5.1. *Agricultural Outputs.* If supply and demand for agricultural products clear at a district level, a reduction in industrial production and employment may additionally impact the marginal revenue product of labor in agriculture through a change in the demand for agricultural products and agricultural prices in a district. Atkin (2008) argues that movement restrictions for agricultural products imply that a district can be considered to be the relevant market for certain agricultural products.

Demand for agricultural products may be affected through two primary channels. Firstly, there may be a direct demand effect from the agro-processing industry, since all input demand functions are likely to be simultaneously affected by trade reform. Secondly, demand for products may be affected by a household level income effect. Shifts in industrial employment will induce changes in equilibrium wages and household nominal income. Real income effects will of course depend on the total household price vector, which is a function of agricultural prices as well as the prices of manufacturing goods. Assume at first that the prices of all products other than agricultural products are determined at a national level and are held constant. An increase in agricultural wages at a local level will increase local nominal income, demand and prices of agricultural products, ceteris parables. This will bias the coefficient on industrial employment upwards.

Even under the assumption that all input and output markets other than agriculture and labor clear at a national level, the prices of all goods purchased by households are likely to be directly affected by trade reform. Trade reform, particularly of consumption goods, will affect household consumption bundles directly through the prices of finished manufacturing goods. Even in the absence of direct trade reform to consumption goods, this effect would be expected in a more attenuated form since changes in the price of intermediate inputs into production will feed through to final goods prices of consumer products. In the presence of non-homogeneous preferences for agricultural products across locations (Atkin, 2008) or non-homothetic Engel Curves, trade reform is likely to induce agricultural prices over time to vary across locations if agricultural product markets clear regionally. The direction of the bias on the coefficient on industrial employment is therefore unclear.

[43]A more detailed discussion of the robustness checks is available from the author upon request.

This concern is addressed by examining whether the excluded set of instruments are conditionally correlated with farm-harvest prices at a district level. Appendix table A.1.2 reports the results from a regression of district level nominal rice, wheat, maize, cotton and sugar prices on the interaction of average and use-weighted average tariffs with district level endowments. A F-Test over the excluded set of instruments for each dependent variable confirms that they are jointly unable to explain a significant fraction of the variation in farm-harvest prices.

6.5.2. *Other Overlapping Input or Output Markets.* The identification strategy relies upon the assumption that labor is the only locally clearing input over which industry and agriculture compete. If industry and agriculture compete for inputs other than labor, trade reform induced shifts in the industrial demand function for all other inputs will lead to changes in the prices of all other locally clearing inputs. For example, if industry competes with agriculture for groundwater (Keskin, 2008), land or capital, the interaction between use weighted tariffs and local resources will be correlated with local land and groundwater prices.[44]

I require at least three rounds of data to estimate whether my excluded set of instruments, conditional on lower order interaction terms, is correlated with local groundwater prices or measures of local groundwater depletion. The rationale behind this test is that in areas in which trade reform is most likely to decrease demand for water from the manufacturing sector, groundwater depletion should be lower (Keskin, 2009). Since groundwater measures for the period in hand are hard to come by, I use two waves of the Minor Irrigation Census of India (1993 and 2000) to examine whether my excluded instruments are correlated with changes in groundwater levels *unconditional* on my lower order interaction terms which is a stronger test than the exclusion restricts would suggest. The groundwater measure I use is the proportion of villages within a district whose water table is below 5 different threshold depths. An increase (reduction) in the proportion of villages whose water table lies, for example, below 10 meters measure captures whether groundwater is being extracted as a faster (slower) than it is being replenished. Trends in groundwater depletion are likely to vary over time according to the type of rock in the region (Jessoe, 2009; Badiani and Jessoe, 2009), which is strongly correlated with mineral composition. I therefore include region time trends to capture variations with districts in groundwater usage over time in areas with common rock types. This would normally be captured by the lower order interaction terms and brings me closer to my original specification. The results of this specification are presented in columns (f) through (j) of appendix table A.1.2. A F-Test over the excluded set of instruments confirms that they are jointly unable to explain a significant fraction of the variation in groundwater depletion.

If industries vary in their intensity of water usage, an alternative test is whether the ratio of employment in water intensive industries is able to explain away the effect of predicted employment on wages. I examine the robustness of my results to separating industries into water and non-water intensive industries

[44] Note the bias in wages generated by other competing inputs runs is likely to run in the opposite direction to my hypothesis - intuitively, a policy shock that shifts out industrial output and input demands in manufacturing is likely to (but does not necessarily) result in an increase the equilibrium price of all locally clearing inputs. A decrease in the amount of land/groundwater employed in agriculture will pull down the marginal product of labor, while a decrease in labor employed will increase the marginal product of labor. Therefore, the finding of a positive effect of predicted labor employment and agricultural wages indicates that the competing inputs story does not drive away the result. Since it is of course possible to construct scenarios in which this is not the case, I verify whether my results are robust to these stories.

following Keskin (2008).[45] I find that my results are robust to this specification: including the ratio of predicted employment in water intensive industries to total employment has little impact on the magnitude or significance of the coefficients on total manufacturing employment.

If the market for goods produced by the manufacturing sector clears at a local level, productivity in the agricultural sector will be directly affected by the policy reforms due to changes in the price of inputs (Adhvaryu et al, 2009). To test whether the exclusion restrictions continue to hold, I would ideally examine the partial correlation between the excluded instruments and the price of inputs into manufacturing. In the absence of district level price data covering the key inputs into agricultural production, I assess the importance of this channel by using the Input-Output matrix of India to divide manufacturing into industries which are backwardly linked to the agricultural sector, and those which are not. Including the ratio of predicted employment in industries supplying to the agricultural sector fertilizers slightly raises the magnitude of the coefficient on total manufacturing employment from 0.30 to 0.34.[46].

6.5.3. *Alternative Channels.* Industrial change may also impact the agrarian sector through non-market channels, for example through negative externalities such as air and water pollution. If the volume of the pollutant is positively correlated with output and agricultural production is an increasing function of groundwater quality, an increase in industrial production is likely to decrease the marginal product of existing labor in agriculture and decrease the unskilled wages. Therefore, the presence of negative externalities would make it less likely that I would find an effect of changes in industrial labor on agrarian wages. It is harder to think of examples of positive externalities of industries which could be driving my results.

Import tariff reforms are also likely to change the price of agrarian inputs produced by industry (e.g. fertilizer). Even if the price of inputs are set at a national level, if the optimal quantity of fertilizer varies by agro-climatic zone, the local input demand response in agriculture will vary by locality. My exclusion restrictions are unlikely to be violated by this channel however since it is unlikely that the localized input demand response is correlated with the instruments used.

If government interventions into agriculture, such as public investment in irrigation or the distribution of seeds, are targeted at areas on the basis of their growth in agricultural productivity, this may contaminate the estimated impact of an increase in agricultural productivity on the marginal product of labor in agriculture. The direction of the bias is unclear, since it will depend on whether government spending in a district is compensatory or reinforcing.

Finally, I conduct a series of specification checks to ensure the results aren't driven by the measures of wages and regions employed. NSSO regions are made up of 4 to 7 districts similar in their agro-climatic and geographic conditions, there is also a high degree of correlation in their mineral and natural resource

[45]Cotton, wool, silk, jute textile production and paper and paper production industries are defined as water-intensive; all other industries are classified as non-water intensive. The classification is based upon water intensity measures published by the Central Water Commission of India.

[46]These results are available from the author upon request

endowments. Therefore regional growth in manufacturing employment and its skill composition are likely to be highly correlated with their district level counterparts. I verify whether the results are sensitive to using region level predicted employment rather than district level employment by using district level industry-employment data from the Economic Census. The magnitude and statistical significance of the estimated coefficient remains robust to the use of district rather than region level data - 10% increase in manufacturing employment is predicted to increase male agrarian wages by 1.4%.[47]

In columns (g) and (h) of table 6, two different measures of wages are used to examine the sensitivity of the results to alternative measures of wages. In column (g) the dependent variable is the log of the median wage of agricultural laborers in the district; in column (h) the wage measure is the log of the median wage of all illiterate wage earners in the district, with the exception of those working in the manufacturing sector.[48] This second measure, by including all agricultural and non-agricultural occupations, provides an indication that both non-agricultural and agricultural wages for unskilled workers respond in similar ways to changes in the structure of production.[49]

Finally, to show that the mechanisms through which shifts in manufacturing labor demand affect wages correspond to my model, I additionally examine the time devoted by households to agricultural production.[50] I find that landless illiterate households are the most responsive to shifts in low-education employment, while landless literate households are the most responsive to shifts in high-education employment. The total days worked in agriculture decreases among landless households in response to an increase in manufacturing employment, while days worked increase for large landed households. The results indicate that large landed households are increasing own-household labor supply as manufacturing employment expands.

7. Step 3: Income Regressions

The results in step 2 of the empirical analysis indicate that both skilled and unskilled wages respond positively to shifts in manufacturing employment driven by changes in industrial policy and agricultural productivity. In particular, agricultural productivity and manufacturing growth differentially alter the return to skilled and unskilled labor in an economy. I now examine how the level and distribution of rural incomes are affected by the two drivers of wage growth. Agricultural productivity raises the return of both labor and land, while increases in manufacturing labor demand raise the return to labor and reduce the return to land. The impact of the two aggregate shocks on household income is predicted to vary across the population according to initial endowments of skills and assets. Following this line of thought, they will also vary in their impact on poverty and inequality. Any source of economic growth

[47] Since the Economic Census does not collect data on the educational attainment of a firm's workforce, it is not possible to verify whether the results are upheld for a shift in the size of the literate population.

[48] Manufacturing sector wages were excluded because the excluded variables will be directly correlated the marginal product of workers in the manufacturing sector, for example through changes to the optimal mix of inputs.

[49] An alternative way of examining this is to look at the ratio of non-agricultural and agricultural wages of unskilled workers, to examine whether they co-move in response to these shocks. I find that the coefficient on the shocks is statistically insignificantly different from 1, confirming this hypothesis. These results are available on request from the author.

[50] These results are available from the author upon request.

that raises the wages of unskilled labor, who constitute the majority of the rural poor in India (Eswaran et al, 2009), is likely to reduce poverty.

7.1. **Empirical Specification.** Prediction 3 implies that household incomes are affected by changes in the aggregate economic environment through two components of income: (a) labor market earnings and (b) profits from cultivation. Household income is restated below:

$$y_h^1 = (1 - s_h^{a1})w_u^1(l_u^{A1} + l_u^{M1}) + s_h^{a1} w_s^1(l_s^{M1} + \mathbf{1}(A_h > 0)l_s^{A1}) + \mathbf{1}(A_h > 0)\Pi_h^A(s_h^{a1})$$

This formulation assumes that households conduct unskilled work in agriculture and manufacturing if they are illiterate and skilled work in manufacturing if they are literate and skilled work in agriculture if they are literate and own land. The labor market earnings of all households are predicted to increase in both agricultural productivity and manufacturing growth. The labor market earnings of households in which the household head is literate are predicted to be more responsive to shifts in skilled rather than unskilled manufacturing growth, which raise the wages of literate individuals more than they do those of uneducated individuals, and vice versa for households with illiterate household heads. Profits from cultivation activities, net of hired labor expenses, are predicted to increase in agricultural productivity and to decrease in manufacturing employment. Profits from cultivation are likely to be particularly affected by shifts outwards in unskilled manufacturing employment, which raise the price of an important input into cultivation - labor. These predictions are tested using the following specification, which is derived in section C.2:

$$y_{hdt} = \gamma_0 + \gamma_1 \hat{\mathrm{E}}^{Manu}_{total,dt} + \gamma_2 \hat{\mathrm{E}}^{Manu}_{skilled,dt} + \gamma_3 \theta_{dt} + \gamma_4 A_{dt} + \gamma_5 HH_{dt} + \mu_d + \delta_t + \epsilon_{hdt}$$

where $\hat{E}^M_{total,dt}$ and $\hat{E}^M_{skilled,dt}$ denotes total and skilled employment in manufacturing, which is predicted using the estimated coefficients of the first stage regression. θ_{dt} is a measure of the agricultural productivity frontier, measured using (20). Household level factors, such as the household's land endowment and skill endowment, are captured by HH_{dt}. A_{dt} captures all other determinants of wages that enter through agricultural labor demand and labor supply, such as the total land cultivated, a measure of the climatic shock experienced in the district during that year, the number of households by quantile of the land distribution and the size of the district level population. District dummies absorb all location invariant factors while year dummies capture all common year effects.

I estimate this equation separately for landed and landless households, where I separate households into groups according to their landholding status and the education of the household head. As such, I divide households into groups according to the likelihood that they are working in the skilled and unskilled labor market, as well as whether they are earning profits from agricultural activities. The coefficients on agricultural productivity and predicted manufacturing employment capture the average effect of changes in these terms on households within that endowment category. In the absence of a measure of household income, I use household consumption as a proxy for income. This measure is likely to be a better measure of income for individuals at the bottom than at the top tail of the income distribution.

7.2. **Data.** In this empirical step, I use data on consumption per capita, poverty, inequality, agricultural productivity, characteristics of the rural population, and the predicted measures of manufacturing employment from the first stage regressions. Consumption per capita and the measures of district level poverty and inequality, are computed using the consumption modules of the NSSO.[51] I follow Deaton (2003a, 2003b) in correcting the estimates for changes in the recall period used and use an alternative set of poverty lines to those put forward by the Indian Planning Commission. Poverty is measured as the headcount ratio and poverty gap, inequality is measured using the gini coefficient of inequality.[52]

7.3. **Results.**

7.3.1. *Consumption.* Since measures of household income are hard to come by, I examine how the two drivers of income growth alter household per capita consumption by household endowment groups. In table 8 I separate households according to their land-holdings and the educational status of their household head. Landless households are defined as those households who own less than 0.1 acres of land; these households are split according to the literacy of the household head. Landed households are separated into those which are net importers and net exporters of labor.[53]

The results from the consumption specifications indicate that an increase in agricultural productivity raises the consumption of all households, but has the greatest impact on net importers of labor - those who own the most land. A 10% increase in agricultural productivity raises the consumption of large landed households by approximately 3%, while it raises the consumption of all other households by approximately 2%.

The effect of manufacturing growth varies substantially by the skill and asset group of the household. A 10% increase in unskilled manufacturing employment raises the consumption of illiterate households by 3.6% and has a smaller impact on the incomes of literate households, raising them by 2.5%. In section 6.3, we saw that shifts in unskilled manufacturing employment have little impact on the wages of rural skilled workers. The income response of literate landless households to changes in unskilled manufacturing employment indicates that these households earn labor income in both the skilled and unskilled labor market. In contrast, skilled manufacturing employment only exerts a positive consumption effect on literate landless households - a 10% in purely skilled manufacturing employment raises the incomes of literate landless households by 4%. These results indicate that illiterate households do not appear to earn any income from skilled labor market activities, while literate landless households earn income from both the skilled and unskilled labor markets.

Landed households are likely to see a reduction in income from cultivation activities as unskilled manufacturing employment rises. This will unambiguously lower the incomes of net importers of labor. The results corroborate this prediction - in contrast to other household groups, these households see no positive

[51] I thank Petia Topalova for sharing her district level poverty data with me.

[52] District level identifiers were not entered into the 1983 data therefore poverty and inequality measures for this year are measured at the NSS region level.

[53] Since there is no information on inputs into cultivation activities in the NSS surveys, I use the 1982 wave of the REDS data to identify the land cut-off at which labor is imported. The REDS data indicate that 75% of households who own less than 2 acres of land don't hired labor, while 90% of household who own more than 2 acres of land do.

consumption effect from an increase in manufacturing employment. By contract small landed households see a positive consumption response to unskilled manufacturing growth; as predicted, the estimated effect is smaller than that of landless households who don't experience reductions in cultivation income.

7.3.2. *Poverty.* Households declaring unskilled, manual labor as their primary source of income constitute a substantial fraction of the rural poor, therefore determinants of unskilled wage growth are expected to decrease poverty. In step 1, we saw that agricultural productivity growth had the greatest positive impact on unskilled wages, therefore this driver of wage growth is likely to have the greatest impact on poverty.

In panels A and B of table 9 I examine how the two drives of wage and income growth differently impact the two measures of poverty, the headcount ratio and depth of poverty. Columns (a) and (b) present estimates using a FE specification, columns (c) through (f) present results from the FE-IV specification. Since the results are similar for the two measures of poverty, I only discuss the results for the poverty rate.

The estimated coefficients from the FE-IV specification indicate that a 10% increase in total manufacturing employment results in a 0.08 point decrease in the poverty rate, or a 2.2% decrease in the mean proportion of households living in poverty in 1987. The composition of employment in the manufacturing is found to have a large effect on the magnitude of the poverty impact: a 10% increase in unskilled employment reduces the proportion of the population living in poverty by 0.13 points, or 3.3% of the mean poverty rate in 1987, whereas the same increase in literate employment has no impact on poverty. A 10% increase in agricultural productivity is found to have a slightly larger impact on rural poverty as the same increase in manual employment - poverty is reduced by 0.14 points or 3.8% of the mean poverty rate in 1987.

7.3.3. *Inequality.* In table 9 I examine the impact of a change in manufacturing employment on the gini coefficient of inequality. The model predicts that agricultural productivity growth will raise inequality since it raises both district level wages and profits from cultivation. A shift out in unskilled manufacturing labor demand growth will decrease inequality since it raises unskilled wages and reduces profits from cultivation. A shift out in skilled manufacturing labor demand will decrease inequality by less than a shift out in unskilled labor demand since unskilled wages increase by less and skilled jobs are more likely to be conducted by wealthier skilled landless households.

The estimated coefficients from the FE-TSLS specification support the predictions of the model. The increase in agricultural productivity raises the Gini coefficient of inequality, although the estimated coefficients are not precisely estimated. A 10% increase in total manufacturing employment results in a 0.003 point decrease in the gini coefficient of inequality, or a 1.2% decrease in mean inequality in 1987. This effect varies according to the skill composition of the manufacturing sector: a 10% increase in unskilled employment reduces the gini coefficient of inequality by 0.007 points, or 2.8% of the mean coefficient in 1987. The same increase in skilled employment has little effect on inequality. Agricultural productivity is found to be positively correlated with inequality, but the estimated effect is statistically insignificantly different from zero.

7.4. **Robustness Checks.**

7.4.1. *Alternative Explanations.* If household's earn income from manufacturing employment in both the wage labor market as well as profits from own-household manufacturing activities, there may be concern that the coefficient on predicted manufacturing employment is partially capturing the response of own-household manufacturing employment to national level industrial policy. An additional concern is that the coefficient on both manufacturing employment and agricultural productivity maybe capturing income responses of households working in the service sector. To ascertain whether these alternative mechanisms may be driving the results, I separate households according to the occupation of the household head. Households are separated into two groups: those who work on the wage labor market and those who conduct their own non-farm business activities.

7.4.2. *Household Level Heterogeneity in Omitted Variables.* I use variation in observed household endowments to examine how household responses to changes in their aggregate economic environment varies according to their observed endowments. Of concern is that households are also likely to vary in their unobserved endowments, for example in their prowess at farming or in their ability to absorb and process new information. In this case, we may be concerned that the relationship between, for example, agricultural productivity and income for large landowners may not just reflect the observation that the returns to land are increasing in agricultural productivity. It may also reflect the observation that large landowners are better poised to take advantage of changes in agricultural technology since they are better farmers. Therefore, the coefficient on agricultural productivity is likely to be larger for landowners than for small landowners, since they are simply better farmers.

8. Step 4: Education Regressions

In the two preceding sections, we have seen that the two correlates of wage and income growth - technical change in agriculture and shifts in manufacturing labor demand - vary in their distributional effects as well as in their impact on the returns to education, as captured by the difference in their effects on unskilled and skilled wages. Secondly, the different sources of growth differ in their effect on the labor market return to education, since they differentially impact the wages of the two education groups. Secondly, the different sources of growth induce consumption responses that vary according to the initial land and skill endowments of a household. I use these observations to disentangle the effect of an increase in income from changes in the return to education by examining variations in enrollment responses across households with different land and skill endowments.

8.1. **Empirical Specification.** Prediction 4 states that the probability that an individual is educated increases in current household income, the anticipated return to education within the agricultural sector, the anticipated return to schooling in the labor market (the relative wages of skilled and unskilled workers in period 2), and decreases in the opportunity cost of schooling. I restate the condition under which households educate their children here for convenience (equation (6)):

$$S_h^y = 1 \quad \text{if} \quad \left(\frac{1}{p_c}\right)^{\rho} [(y_h^1 - C_s)^{\rho} - (y_h^1 - \alpha)^{\rho} + (w_s^2 + \Pi^A(s_h^{a2} = 1))^{\rho} - (w_u^2 + \Pi^A(s_h^{a2} = 0))] > 0$$

Schooling decisions reflect expected returns to education. To empirically estimate the determinants of schooling choices, we require assumptions about the formation of expectations. I assume that parents have myopic expectations about the returns to education faced by their children. Parents therefore believe that the future returns of their children, throughout their working lives, are the same as the returns the parents face today:

$$\frac{w_{sd,t+1}}{w_{ud,t+1}} = \frac{w_{sdt}}{w_{udt}} + \varsigma_{hdt}$$
$$\theta_{d,t+1} = \theta_{dt} + \sigma_{hdt}$$

where $E[\varsigma_{hdt}] = 0$ and $E[\sigma_{hdt}] = 0$. I therefore use the wage of workers in different education brackets as a proxy for the wages that the children will receive upon entering the labor market. For older children about to enter the labor market (those aged between 10 and 14) this assumption may be more reasonable than for the younger children (aged 5 to 9). I test the robustness of the coefficient estimates to empirically relaxing this assumption. [54]

The model captures education choices in discrete terms which implies that there are two education choices and labor market wages: literate or illiterate. While this is a convenient but not necessary assumption made for tractability, an alternative approach would be to consider households evaluating future labor market and cultivation earnings and their cost at multiple schooling levels. Guided by the descriptive statistics, I examine three margins along which households make their education choices: whether to acquire any schooling or not, whether to complete primary education or not and whether to continue beyond primary school.

The data indicate that the decision as to whether to acquire any education is a relevant margin among which rural education choices are made. In 1987, 29% of males between the ages of 10 to 14 hadn't completed any schooling. Among landless households this figure rises to 34%. Among the 70% of males aged 15 to 19 who had started primary school, approximately 10% had not completed it; among landless households the corresponding figures are 66% and 54%. Female school attendance and primary school completion rates are even lower - 51% of females between the ages of 10 and 14 hadn't attended any schooling; approximately 10% of the girls aged 13-18 who had started primary school did not complete it. For the two groups who are deciding whether to invest in education - attending or completing primary school - the appropriate rate of return to consider is the return to obtaining literacy relative to the return to remaining illiterate.

[54]First, I assume that households expect the returns to education to grow at a trend growth rate correlated with initial returns. Therefore, I include a linear time trend interacted with the initial return to education in the education specification. As an additional test, I examine whether education choices of the younger cohort respond to future observed wages.

I capture education outcomes using ex-post education choices of groups of children. I separate children by ages, according to the education choice they were making in the previous period. Children in India start primary school at 6 and middle school at 12, although late commencement of primary school is not uncommon. To capture the choice of whether to attend school or not, I examine whether children aged between 10 and 14 have ever attended school. Since the NSSO survey is collected every five years, children aged between 10 and 14 in 1987 would have been aged between 5 and 9 in 1983.

I specify the following linear approximation to equation (6):

$$\text{(21)} \qquad \mathrm{S}_{ihdt} = 1 \quad if \quad \kappa_0 + \kappa_1 y_{hdt} + \kappa_2 (w_{sdt} - w_{udt}) + \kappa_3 \theta_{hdt} * A_h + \kappa_4 OC_{dt} + \kappa_5 \mathrm{C}^s_{dt} + \mu_h + \mu_d + \mu_t + \epsilon_{hdt} > 0$$

where S_{ihdt} is a dummy variable capturing whether the individual started primary school or continued to middle school and OC_{dt} represents the opportunity cost of schooling. Educational investment is predicted to increase in household income ($\kappa_1 > 0$) and in the labor market return to an extra year of education $\kappa_2 > 0$. For cultivator households, enrollment is predicted to increase directly in agricultural technical change, due to the complementarity between agricultural technical growth and education ($\kappa_3 > 0$). Education is predicted to decrease with the opportunity cost of schooling ($\kappa_4 < 0$) and the direct monetary cost of schooling $\kappa_5 < 0$. Educational is also additionally likely to depend on household preferences through the marginal utility of income and household time preference parameters μ_h.[55]

Education responses to incomes are predicted to vary across households since they reflect household level credit constraints or symbolic consumption of education. Households whose marginal utility of an additional unit of education is higher will exhibit higher income responses than those for whom it is lower. Since illiterate landless households fall into the lowest income bracket, their education responses are anticipated to be greater than those of literate landless households. I test whether this is the case - i.e. I test whether $\kappa_{1,LL,u} > \kappa_{1,LL,s}$. For similar reasons, the response of illiterate landless households to an increase in the opportunity cost of schooling is expected to be greater than that of literate landless households. However, as argued below, it is likely that the opportunity cost of schooling is zero for the young children examined in this paper; therefore

The opportunity cost of schooling reflects other productive uses of child time, which could be conducted in lieu of investing in education.[56] These includes domestic production or engaging in income generating activities, in cultivation for example. The data presented in section 2.4 suggest that, prior to age 10, few children work in the wage labor market or at home - in 1987, 2.5% of boys and 8% of girls aged between 6 and 10 reported conducting domestic or income generating activities. Among children who do not attend school between the age of 6 and 9, only 3.4% of boys and 5% of girls report conducting domestic

[55]The literature also considers the impact on parental inputs, variation in individual education human capital production functions and individual/household level time preferences and various dimensions of education quality. While these are likely to be important for the intensive margin of improving test scores, it isn't clear that they are as important on the extensive margin.

[56]For tractability, a child labor market was not introduced into the model and the opportunity cost of schooling was modeled in terms of domestic consumption. In the presence of a market for child labor, the opportunity cost of attending school will depend on factors which raise child wages.

or income generating activities. This indicates that the opportunity cost of children to attending school in the youngest age group are likely to be small. I test whether this is the case by estimating whether $\kappa_4 = 0$.[57]

As noted in section (??), landless households earn income in the wage labor market while landed households earn income from both cultivation and the labor market. Substituting household income into (21) and combining parameters, we get an estimating equation which is a function of the wages of skilled and unskilled workers, the cost of schooling, and cultivation income:

$$\begin{aligned}(22) \qquad \mathrm{S}{=}1_{ihdt} \quad if \quad & \kappa_0 + (\kappa_1\lambda - \kappa_2 + \kappa_4)w_{udt} + (\kappa_1\xi + \kappa_2)w_s + \kappa_1 1(A > 0)\Pi^A + \kappa_3\theta_{dt} * A_h \\ & + k_5\mathrm{C}^s_{dt} + \mu_h + \mu_d + \mu_t + \epsilon_{hdt} > 0\end{aligned}$$

where λ denotes the proportion of household time spent working in the unskilled labor market and ξ denotes the proportion of household time spent in the skilled labor market. Π^A is the household cultivation profit function, which is decreasing in labor market wages, household land endowments, the agricultural technology θ_{dt} and the education of the household head.

In equation (22), the coefficients on wages and incomes capture combinations of the reduced form parameters. I separate the income, opportunity cost and returns to education effects by noting that the combinations of the parameters varies across groups of households according to their skill and land holdings, and across children according to their age. Firstly the parameters on wages vary across households according to their educational endowments, which determine the time they devote to skilled and unskilled labor market activities. Secondly cultivator households additionally earn income from cultivation activities, which are themselves decreasing functions of wages; in the absence of data on cultivation incomes, the coefficients on wages will capture education responses to changes in cultivation incomes driven by variation in wages. Thirdly, children aged six to ten have little to no opportunity cost of schooling in India, as indicated by the low proportion of children working in this age group in either the wage labor market or in household activities.

Equation (22) cannot be estimated directly using reported measures of household incomes and wages since the coefficients may reflect a variety of different explanations. For example, the ratio of skilled and unskilled wages are likely to be correlated with local preferences for education. Similarly, labor market incomes may reflect the quality of schooling in a given location. I use two different approaches to back out the parameters of the education specification which both overcome the potential bias from using an OLS specifications: (1) a "semi-reduced form" strategy and (2) an instrumental variables strategy. The first approach requires more structure to back out the estimated parameters.

In the first approach, I substitute empirical specifications for the profit function and wages into equation 22 to obtain the estimating equation. The profit specifications are discussed in greater detail in appendix C.2.

$$(23) \qquad \mathrm{S}_{hdt} \quad = \quad a_0 + a_1\hat{\mathrm{E}}_{total,dt} + a_2\hat{\mathrm{E}}_{skilled,dt} + a_3\theta_{dt} + a_4\mathrm{C}^s_{dt} + a_5 HH_{ht} + a_6 A_{d,t} + u^a$$

[57] Both types of work increase substantially above the age of 10 - among children not enrolled in school between 10 and 14, 50% of males and 65% of females conduct income generating or domestic activities.

The estimated coefficients are combinations of the parameters of the α and β terms in the wage regressions (equations 15 and 16), and the structural parameters - the κ terms - from the education specification (equation 21). I include district fixed effects in all specifications to capture variation across districts in unobserved time invariant determinants of education, such as the quality of education infrastructure, which may be correlated with the average level of manufacturing employment and agricultural productivity in a district. In addition, in the absence of data on the costs of schooling facing households, I include state-year interactions which capture variations in the provision and cost of education at a state level.

The parameter combinations vary across groups of households according to their skill and land holdings. I use a linear probability model to estimate the reduced form coefficients. I estimate the three structural parameters, κ_1, κ_2 and κ_4 using a minimum distance estimator. I use the optimal minimum distance estimator, where the weight matrix used is the inverse of the variance covariance matrix of the reduced form coefficients.

In the IV strategy, I instrument wages using the interactions of the effective tariff and district level natural resource endowment and agricultural productivity. In the absence of data on net cultivation revenues, this approach can only be used for landless households who do not earn income from cultivation activities. I separate landless households into two groups: those with illiterate household heads, and those with literate ones. I assume that illiterate households work only in the unskilled labor market, an assumption consistent with the results presented in section 7.3.1. By contrast, skilled households work in both the skilled and unskilled labor markets. The two estimating equations are therefore:

Illiterate Landless:

$$\text{S=1}_{ihdt} \quad if \quad \kappa_0 + (\kappa_1 - \kappa_2 + \kappa_4)w_{udt} + \kappa_2 w_{sdt} + k_5 \text{C}^s_{dt} + \mu_d + \mu_t + \epsilon_{hdt} > 0 \tag{24}$$

Literate Landless:

$$\text{S=1}_{ihdt} \quad if \quad \kappa_0 + (\kappa_1\lambda - \kappa_2 + \kappa_4)w_{udt} + (\kappa_1\xi + \kappa_2)w_{sdt} + k_5 \text{C}^s_{dt} + \mu_d + \mu_t + \epsilon_{hdt} > 0 \tag{25}$$

The three structural parameters, κ_1, κ_2 and κ_4 are estimated using a minimum distance estimator. The models are estimated separately by sex. This allows me to examine whether the estimated coefficients vary across males and females.[58]

8.1.1. *Data.* To test prediction 4, I use data on education outcomes, agricultural productivity, characteristics of the rural population, and the predicted measures of manufacturing employment from the first stage regressions. Households are asked the level of educational attainment that their child has achieved. The NSSO uses a self-reported measure of educational attainment to classify individuals into literate and illiterate (Government of India, 2001). A literate is defined as someone who can both read and write. Children who have never attended school are those who report being illiterate or are literate without ever

[58]To test variation in the parameters of the educational specification across males and females, ideally I would use male and female wages for both skilled and unskilled workers. I conduct this test in the absence of such data.

having conducted any formal schooling.[59] Individuals who continue beyond primary school are defined as those whose highest education standard is middle school or beyond.

8.1.2. *Results.* The estimates from the education specifications are presented in tables 10, 11 and 12. Table 10 presents the estimates from equation (23), while table presents the estimates from the IV regression. The structural parameters estimated using a minimum distance estimator are presented in table 12.

The results from the specification estimating education responses to the two drivers of wage and income growth, equation (23), are displayed in Table 10. The relationship between educational outcomes and the two drivers of growth varies across landholding and skill. In illiterate landless households, raising agricultural productivity by 10% raises the probability that a boy aged 5 to 9 in these households will attend school by 0.1. Since growth in agricultural productivity raises wages but reduces the returns to schooling, the sign of the coefficient estimate for landless illiterate households indicates that the positive income effect from rising unskilled wages outweighs the negative effect from the reduction in the returns to schooling and, potentially, from the opportunity cost effect. Landed households also exhibit positive education responses - a 10% increase in agricultural productivity raises the probability that a boy from a large landed households (a net importer of land who owns more than 2 acres of land) by 0.05. This is likely to reflect both income and returns to education effects - agricultural productivity additionally raises the return to primary school through managerial activities in cultivation (Foster and Rosenzweig, 1996).

A 10% increase in unskilled manufacturing employment raises the probability that a child in a literate household will acquire schooling by 0.1, while in large landless households it reduces it by 0.05. Since both landed and landless households experience similar reductions in the return to literacy, the difference in education responses is likely to reflect the heterogeneous effect of rising manufacturing labor demand on household incomes according to their initial land endowments. The coefficient estimates indicate that both literate and illiterate landless households display a positive response to increases in predicted skilled manufacturing employment. Since illiterate landless households are unlikely to be working in the literate labor market and shifts out of skilled manufacturing employment have a very small impact on unskilled wages, this is likely to reflect a positive returns to education effect.

The reduced form parameters capture combinations of structural parameters, I estimate the structural parameters using optimal minimum distance. The results are presented in table 12. In columns (a) and (b), I estimate the income, returns to education and opportunity cost effect for landless households. I allow the income effect to differ across literate and illiterate households, the income effect is predicted to be larger in illiterate households. The coefficient estimates suggest that the probability that a child

[59]Children may be reported as illiterate if they have attended school but have not learned how to read or write their names. The NSSO classifies the education of an individual according to the highest standard that they have obtained. Therefore in the case of an individual who has attended school but cannot read or write, they should be classified as having some formal education (Government of India, 2001). To verify whether this is the case, the 55th round of the NSSO survey conducted in 1999 asks children to report whether they have ever attended school. Of the 33% of illiterate 9 to 14 year olds, 92% had never attended school while 3% attended school but dropped out. This provides some reassurance that this measure is capturing individuals who decided not to enroll in school.

commences school is increasing in households incomes. As predicted, the magnitude of the effect is larger for illiterate households than for literate households - a 10% increase in income raises the probability that children in illiterate households start school by 0.3, compared to 0.2 in literate households. As predicted, the magnitude of the effect is larger for illiterate households. The positive income effect may represent that households are credit constrained or that education enters directly into household utility. I am unable to distinguish between these two channels. Intuitively, the larger effect for illiterate landless households indicates that their choices were “more” constrained - their average marginal utility of sending their child to school was higher.

The coefficient estimates also indicate that the returns to education effects are large and statistically significant. The relative magnitude of the coefficients indicates that the household response to increases in incomes, at this level of education choice, are substantially larger than the education responses to returns to education. The parameter estimates suggest that the size of the effect is slightly larger in literate landless households than in illiterate landless households - a 10% rise in returns to education increases the probability of attending school by 0.09 in illiterate households, versus by 0.13 in literate households - although the difference is not statistically significant. In comparison, the female response for girls is small and insignificant. This is likely to reflect the absence of female labor demand in skilled occupations in rural areas: fewer females work in skilled occupations in manufacturing than men and females conduct less than 5% of supervisory labor in agriculture (REDS, 1999). Females are also more likely to marry outside of the district (Rosenzweig and Stark, 1989); the estimated effect may also be because female education responds to a geographically wider return to education than that found in district level labor markets.

The estimated opportunity cost effect for boys is not statistically different from zero. This is likely to reflect the observation that boys not attending school do not appear to be working, either in income generating or home production activities. For girls, the opportunity cost is substantially larger in magnitude, although not statistically different from zero. This may reflect the observation that females are more likely to be assisting in domestic work between the ages of 6 and 10.

9. Conclusion

One third of children in rural households did not attend school during the 1990s. This paper has shown that growth that raises household incomes and the returns to schooling raises the probability that a child commences primary school. The response of education to increases in incomes is larger in magnitude than to rising returns to schooling. The estimates suggest that policies promoting growth that raise the incomes of unskilled workers, those at the bottom tail of the income distribution, are likely to raise both the probability that their children attend any schooling as well as raise households incomes and reduce poverty.

The paper provides empirical evidence supporting the hypothesis that growth alters the distribution of income through changing the returns to unevenly distributed productive assets in an economy. The source of growth is important for understanding the distributional impact of economic change: in this paper, I find that agricultural technical change increases consumption inequality in rural areas while manufacturing employment reduces it.

The estimates from my regressions also suggest an answer to an important overarching question: did agricultural productivity growth or employment growth in manufacturing have a greater impact on increasing agrarian wages and reducing poverty between 1983 and 1999? My estimates suggest growth in agricultural technology resulted in a 40% increase in real wages between 1983 and 1999, accounting for 33% of total unskilled wage growth. Over the same period, growth in manufacturing employment resulted in a 10.8% increase in unskilled wages, accounting for 14.6% of wage growth. The majority of employment growth consists of skilled employment. Had the increase in manufacturing workers seen over the period consisted of only unskilled workers, the coefficient estimates suggest that the wages of rural unskilled workers would have increased by nearly twice the amount. The skill-biased employment growth witnessed in India (Kochhar, 2006) has implied that there has been relatively little effect of changes in the size of this sector on the wages of unskilled individuals and poverty. This is an observation that has been repeatedly asserted in the policy literature. To my knowledge, the estimates in this paper are the first to empirically validate this observation.

10. Results

Table 2a: Descriptive Statistics on Individuals Working in Manufacuturing, by Occupation Type						
Occupation Bracket	**Proportion of**		**Literate**		**At Least Primary**	
	Males	**Females**	**Males**	**Females**	**Males**	**Females**
White Collar	0.17	0.13	0.78	0.29	0.69	0.19
Blue-Collar, Skilled	0.48	0.32	0.72	0.33	0.56	0.26
Blue-Collar, Manual	0.35	0.54	0.49	0.23	0.35	0.15
Average In NSS	-	-	0.52	0.20	0.41	0.14

Table 2b: Tasks Done by Occupational Bracket in the Manufacturing Sector					
Occupation Bracket	**Non-Routine Analytic Tasks**	**Non-Routine Physical Tasks**	**Routine Physical Tasks**	**Non-Routine Interactive Tasks**	**Routine Cognitive Tasks**
Panel A: Males					
White Collar	4.19	0.56	3.35	3.65	3.32
	(1.69)	(0.37)	(0.76)	(3.05)	(3.08)
Blue-Collar, Skilled	2.68	1.64	4.58	0.54	8.49
	(1.99)	(1.59)	(1.22)	(1.43)	(2.01)
Blue-Collar, Manual	2.10	1.63	4.23	0.39	6.81
	(1.26)	(1.26)	(0.99)	(0.39)	(2.56)
Average in DOT (1977)	3.84	1.45	4.02	2.01	4.94
Panel B: Females					
White Collar	4.50	0.16	3.31	3.75	1.56
	(1.08)	(0.21)	(0.49)	(3.45)	(1.68)
Blue-Collar, Skilled	1.48	0.82	4.77	0.04	8.49
	(1.86)	(0.75)	(1.21)	(0.22)	(1.83)
Blue-Collar, Manual	0.88	0.89	3.92	0.07	4.85
	(1.05)	(0.90)	(0.89)	(0.25)	(3.81)
Average in DOT (1977)	3.63	1.31	4.03	1.86	5.04

A crosswalk was created between Indian Occupational Codes (NCO-68) and the US Census Occupation Codes (1960). The Dictionay of Occupational Titles dataset was assembled by Autor, Levy, and Murnane (2003); they collected data on job task requirements from the US Department of Labor's Dictionary of Occupational Titles (DOT) and merged them with census occupation classifications. The classification of jobs into White Collar, Blue-Skilled and Blue Manual was conducted using the NCO-68. The average characteristics in the DOT is an unweighted mean across all occupations. Definitions: Non-Routine Analytical Tasks -General Educational Development, Maths; Non-Routine Physical Tasks - Eye-Hand Coordination; Routine Physical Tasks - Finger Dexterity; Non-Routine Interactive Tasks - Direction, Control and Planning; Routine Cognitive Tasks - Set Limits, Standards

TABLE 3: Decomposition of Variance in the Proportion of Educated Workers in a Region-Industry				
Dependent Variable: Proportion of Region-Industry Workforce by education				
	Literate	**< Primary**	**< Middle**	**< Secondary**
Panel A: 1987	**(a)**	**(b)**	**(c)**	**(d)**
Region Dummy	10.11	10.87	12.02	14.33
Industry Dummy	17.32	21.11	24.84	26.36
R-Squared	0.27	0.31	0.34	0.36
Panel B: 1993	**(a)**	**(b)**	**(c)**	**(d)**
Region Dummy	5.09	5.8	7.41	9.2
Industry Dummy	17.52	20.26	23.17	24.99
R-Squared	0.26	0.29	0.32	0.34
Panel C: 1993-1987	**(a)**	**(b)**	**(c)**	**(d)**
Region Dummy	1.69	1.93		
Industry Dummy	10.67	11.36	6.24	
R-Squared	0.14	0.17		
Obs	8372	8372	8372	8372
N Region Dummies	52	52	52	52
N Industry Dummies	161	161	161	161

Table 1a: Primary Occupations for Working Individuals of Working Age, 20 to 55

Panel A: Males

	1983				1999			
Occupation	**Percent in Population**	**Proportion Illiterate**	**Proportion Primary +**	**Land Owned (acres)**	**Percent in Population**	**Proportion Illiterate**	**Proportion Primary +**	**Land Owned (acres)**
Agricultural Sector	74.72	0.57	0.29	4.99	69.8	0.44	0.43	2.77
of which...								
Hired Labor Market	30.24	0.69	0.17	1.15	33.04	0.58	0.27	0.64
Farm manager	0.02	0.35	0.47	3.62	0.02	0.49	0.40	1.86
Manual Laborer	30.22	0.70	0.16	1.17	33.02	0.56	0.28	0.64
Working on Household Farm	44.48	0.49	0.37	7.61	36.76	0.33	0.55	4.69
Head	26.91	0.55	0.30	6.34	21.03	0.40	0.46	3.78
Other family member	17.57	0.41	0.47	9.54	15.73	0.23	0.67	5.91
Non-Agricultural Sector (excl gov)	20.81	0.38	0.46	1.45	27.52	0.27	0.59	0.91
of which...								
Manufacturing Worker	7.07	0.37	0.45	1.05	8.00	0.27	0.58	0.89
Other	13.73	0.40	0.46	1.67	19.52	0.28	0.59	0.92
TOTAL - RURAL	**95.53**	**0.5**	**0.36**	**4.3**	**97.32**	**0.37**	**0.5**	**2.25**

Panel B: Females

	1983				1999			
Occupation	**Percent in Population**	**Proportion Illiterate**	**Proportion Primary +**	**Land Owned (acres)**	**Percent in Population**	**Proportion Illiterate**	**Proportion Primary +**	**Land Owned (acres)**
Agricultural Sector	86.08	0.90	0.06	4.27	83.75	0.79	0.13	2.43
of which...								
Hired Labor Market	41.21	0.93	0.04	1.18	47.37	0.83	0.09	0.73
Farm manager	0.04	0.81	0.19	0.51	0.1	0.84	0.09	0.74
Manual Laborer	99.96	0.93	0.04	1.20	99.9	0.76	0.00	2.09
Working on Household Farm	44.87	0.87	0.08	7.10	36.38	0.74	0.19	4.64
Head	8.92	0.88	0.07	7.45	92.06	0.74	0.19	4.84
Other family member	91.08	0.81	0.12	3.54	7.94	0.75	0.16	2.32
Non-Agricultural Sector (excl gov)	6.26	0.82	0.11	0.96	13.64	0.65	0.23	0.61
of which...								
Manufacturing Worker	6.14	0.77	0.14	0.76	7.25	0.61	0.26	0.57
Other	0.12	0.86	0.08	1.16	6.39	0.70	0.20	0.66
TOTAL - RURAL (excl gov)	92.34	0.81	0.13	4.31	97.39	0.66	0.25	2.15

TABLE 1b: Literacy in the Rural Labor Force by Industry and Occupation

Dependent Variable: Takes a value of 1 if an individual is literate

	Panel A: Males						**Panel B: Females**					
	Agriculture and Manufacturing			**Total Rural Labor Force**			**Agriculture and Manufacturing**			**Total Rural Labor Force**		
Manufacturing	0.352***	0.344***	0.412***	0.122***	0.120***	0.185***	0.272***	0.273***	0.321***	0.017	0.017	0.061**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.023)	(0.026)	(0.026)	(0.027)
Manual Manufacturing	-	-	-0.194***	-	-	-0.188***	-	-	-0.086***	-	-	-0.188***
	-	-	(0.011)	-	-	(0.011)	-	-	(0.012)	-	-	(0.011)
Own-Farm Cultivation	0.338***	0.326***	0.326***	0.101***	0.093***	0.093***	0.117***	0.122***	0.121***	-0.136***	-0.136***	-0.136***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.018)	(0.018)	(0.018)
Head*Own-Farm Cultivation	-	0.028***	0.027***	-	0.024***	0.024***	-	0.037***	0.037***	-	0.011	0.011
	-	(0.009)	(0.009)	-	(0.008)	(0.008)	-	(0.011)	(0.011)	-	(0.011)	(0.011)
Hired Agriculutral Laborer	-	-	-	-0.230***	-0.226***	-0.226***	-	-	-	-0.254***	-0.253***	-0.253***
	-	-	-	(0.014)	(0.014)	(0.014)	-	-	-	(0.018)	(0.018)	(0.018)
Observations	69037	69037	69037	91696	91696	91696	28273	28273	28273	33076	33076	33076
R-squared	0.211	0.213	0.216	0.219	0.222	0.224	0.259	0.259	0.26	0.276	0.276	0.277

All columns include district identifiers and age polynomials.

TABLE 1c: Wages by Industry and Occupation

Dependent Variable: Log(Individual's Real Wage)

	Panel A: Males				**Panel B: Females**			
Literate	0.189***	0.171***	0.208***	0.167***	0.188***	0.185***	0.245***	0.154***
	-0.007	-0.007	-0.008	-0.007	(0.007)	(0.007)	(0.016)	(0.012)
Manufacturing Sector	-	0.147**	0.112*	0.048	-	0.189***	0.245***	0.067
	-	(0.062)	(0.065)	(0.062)	-	-0.049	(0.072)	-0.068
Agricultural Sector	-	-0.042***	0.01	-	-	-0.056***	0.031	-
	-	-0.008	(0.010)	-	-	(0.008)	(0.010)	-
Literate*Manufacturing	-	-	0.034	-	-	-	-0.267***	-
	-	-	(0.026)	-	-	-	(0.041)	-
Literate*Agriculture	-	-	-0.139***	-	-	-	-0.189***	-
	-	-	(0.014)	-	-	-	-0.025	-
Skilled*Manufacturing	-	-	-	0.225***	-	-	-	0.106***
	-	-	-	(0.023)	-	-	-	(0.039)
Manual*Agriculture	-	-	-	-0.042***	-	-	-	0.006
	-	-	-	(0.008)	-	-	-	(0.010)
Non-Manual*Agriculture	-	-	-	0.325**				
	-	-	-	(0.134)				
Observations	23598	23598	23598	23598	28273	28273	28273	33076
R-squared	0.513	0.516	0.518	0.518	0.259	0.259	0.26	0.276

All columns include district identifiers and age polynomials.

Table 4: Educational Enrollment and Child Labor, by Age Group

	1987					
	Males			Females		
Age…	Work	Domestic	Enrolled	Work	Domestic	Enrolled
6 to 11 (Primary)	0.08	0.01	0.68	0.08	0.09	0.46
12 to 14 (Middle)	0.26	0.02	0.63	0.25	0.30	0.35
14 to 16 (Lower Secondary)	0.53	0.01	0.44	0.38	0.47	0.18
	1993					
	Males			Females		
Age…	Work	Domestic	Enrolled	Work	Domestic	Enrolled
6 to 11 (Primary)	0.06	0.01	0.78	0.06	0.08	0.59
12 to 14 (Middle)	0.20	0.01	0.72	0.20	0.28	0.49
14 to 16 (Lower Secondary)	0.50	0.01	0.48	0.34	0.45	0.27
	1999					
	Males			Females		
Age…	Work	Domestic	Enrolled	Work	Domestic	Enrolled
6 to 11 (Primary)	0.03	0.01	0.82	0.03	0.04	0.71
12 to 14 (Middle)	0.14	0.01	0.73	0.13	0.19	0.59
14 to 16 (Lower Secondary)	0.39	0.01	0.53	0.27	0.37	0.37

TABLE 5: First Stage Specification

Dependent Varible: Log(Region-Industry Total Employment)	No Policy	Policy: Deregulated, D=1				Policy: Import Tariffs			
	Total	Total	Literate	At Least Primary	Total	Total	Literate	At Least Primary	Total
Measure of Factor Use	Dummy		Dummy		Quantile		Dummy		Quantile
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
Policy	0.050***	-0.3842	-0.9535***	-1.1765***	0.2706	1.0385**	0.8481**	0.7103**	1.4166**
	(0.0150)	(0.3179)	(0.2837)	(0.3028)	(0.3245)	(0.4200)	(0.3331)	(0.2971)	(0.6363)
Policy*Groundwater	-0.335	0.0348**	0.0353***	0.0188	0.0048	0.0275**	0.0092	0.0027	0.0054
*Agro Proportion	(0.2690)	(0.0131)	(0.0122)	(0.0121)	(0.0038)	(0.0119)	(0.0122)	(0.0106)	(0.0039)
Policy*Uncultivable Land	0.102***	-0.7137**	-0.3166	-0.1415	-0.1500**	-0.4691***	-0.3574**	-0.2745**	-0.1994***
*Agro Proportion	(0.0280)	(0.2821)	(0.2398)	(0.2520)	(0.0683)	(0.1589)	(0.1459)	(0.1311)	(0.0603)
Policy*Forest Cover	0.016***	0.4867***	0.3843***	0.3376***	0.1025***	0.3361***	0.2962***	0.2876***	0.1153***
*Wood Proportion	(0.0050)	(0.0424)	(0.0387)	(0.0372)	(0.0133)	(0.0438)	(0.0432)	(0.0443)	(0.0195)
Policy*Ceramics Minerals	0.027***	0.0258***	0.0182*	0.0173*	-0.001	0.0538***	0.0342***	0.0300***	0.0135***
*Ceramics Proportion	(0.0040)	(0.0093)	(0.0103)	(0.0093)	(0.0024)	(0.0079)	(0.0087)	(0.0094)	(0.0019)
Policy*Construction Minerals	-0.030***	0.0303***	0.0314***	0.0301***	0.0077***	0.0162***	0.0131***	0.0109**	0.0160***
*Construction Proportion	(0.0060)	(0.0040)	(0.0043)	(0.0044)	(0.0017)	(0.0044)	(0.0049)	(0.0044)	(0.0021)
Policy*Chemicals Minerals	-0.051***	-0.0417***	-0.0372***	-0.0327***	-0.0081***	-0.0069	-0.0048	-0.0026	-0.0045*
*Chemicals Proportion	(0.0050)	(0.0057)	(0.0050)	(0.0053)	(0.0022)	(0.0051)	(0.0049)	(0.0045)	(0.0027)
Policy*Metals	-0.001	-0.0163***	-0.0085*	-0.0085*	-0.0032*	0.0165***	0.0192***	0.0168***	-0.002
*Metal Proportion	(0.0090)	(0.0048)	(0.0047)	(0.0050)	(0.0017)	(0.0060)	(0.0054)	(0.0048)	(0.0020)
Policy*Coal	-0.315*	-0.0594***	-0.0469***	-0.0332***	-0.0160***	0.0595***	0.0400***	0.0332***	0.0205***
*Energy Proportion	(0.1660)	(0.0162)	(0.0151)	(0.0119)	(0.0047)	(0.0124)	(0.0122)	(0.0091)	(0.0043)
Policy*Electricity	1.978***	-0.2142***	-0.2045***	-0.1754***	-0.0543***	-0.3307***	-0.2708***	-0.2487***	-0.0748***
*Energy Proportion	(0.0710)	(0.0244)	(0.0237)	(0.0210)	(0.0093)	(0.0617)	(0.0487)	(0.0559)	(0.0261)
Log(Agricultural Productivity)		-0.9102	1.2275	0.5079	-1.1394	-0.9286	-0.9247	-1.0722	-0.9156
		(3.4739)	(4.0889)	(3.4532)	(3.6185)	(0.9572)	(0.8500)	(0.8016)	(0.9407)
Constant	1.978***	13.2469	12.7234	14.5823	11.9758	-0.8807	-1.2457	-1.0083	-0.8627
	(0.0710)	(15.6356)	(13.9158)	(14.2868)	(15.7953)	(16.3827)	(15.2170)	(14.8191)	(16.1081)
Observations	5856	23450	23450	23450	23450	22539	22539	22539	22539
F-Statistic - Triple Interactions	-	34.49	32.35	21.98	22,93	23.76	13.16	11.39	28.52

* significant at 10%; ** significant at 5%; *** significant at 1%

All specifications include NSS region fixed effects, standard errors are clustered at the region level. Columns 1-3 measure the 0/1 use of a factor in an industry, where an industry is classified as using the input if the measured factor intensity lies above the median for all industries. In Column 4, the intensity with which an industry uses a given factor is measured using quantiles.

TABLE 6: Unskilled Wage Response to Manufacturing Employment and Agricultural Productivity

Panel A: Dependent Variable: Log Male Agricultural Wages for Manual Tasks

	FE	FE-IV	FE-IVD	FE-IV	FE-IV	FE-IV	FE-IV
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Log(Manufacturing Employment)	0.028	0.144*	0.175**	0.159*	0.146**	0.118*	0.132
	(0.036)	(0.085)	(0.076)	(0.086)	(0.074)	(0.062)	(0.145)
Log(Agricultural Productivity)	0.643***	0.428**	0.443*	0.214	0.439**	0.456	0.365**
	(0.208)	(0.206)	(0.239)	(0.246)	(0.214)	(0.244)	(0.167)
Constant	-10.286***	-11.678***		-8.799***	-2.451***	-8.369***	-7.346***
	(2.715)	(3.056)		(3.694)	(0.879)	(3.107)	(2.566)
Adjusted R-Squared	0.800	0.806		0.842	0.888	0.674	0.747
District Fixed Effects	Y	Y	Y	Y	Y	N	Y
Region Time Trend	N	N	N	Y	N	N	N
First Stage F-Stat	-	32.19***	13.087***	32.19***		32.19***	32.19***
Observations	678	930	678	930	418	930	1100

Panel B: Dependent Variable = Log Male Agricultural Wages, Manufacturing Employment Broken Apart by Skill

	FE	FE-IV	FE-IVD	FE-IV	FE-IV	FE-IV	FE-IV
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Log(Manufacturing Employment)	0.001	0.296**	0.309***	0.288*	0.299**	0.217	0.318*
	(0.002)	(0.135)	(0.138)	(0.226)	(0.120)	(0.142)	(0.165)
Log(Literate Employment)	0.046	-0.239*	-	-	-0.251**	-0.155	-0.379*
	(0.031)	(0.142)	-	-	(0.118)	(0.145)	(0.206)
Log(At Least Primary Employment)	-	-	-0.259**	-	-	-	
	-	-	(0.151)	-	-	-	
Log(Skilled Employment)	-	-	-	-0.226	-	-	
	-	-	-	(0.222)	-	-	
Log(Agricultural Productivity)	0.618***	0.354**	0.371*	0.386*	0.258	0.322	0.321*
	(0.208)	(0.206)	(0.215)	(0.216)	(0.221)	(0.250)	(0.173)
Constant	-10.294***	-9.999***	-10.667***	-10.433***	-7.319**	-6.553**	-6.900***
	(2.823)	(3.188)	(3.217)	(3.253)	(3.462)	(3.179)	(2.668)
Adjusted R-Squared	0.801	0.801	0.801	0.801	0.842	0.683	0.743
District Fixed Effects	Y	Y	Y	Y	Y	N	Y
Region Time Trend	N	N	N	Y	N	N	N
First Stage F-Stat	-	32.19***	13.087***	32.19***		32.19***	32.19***
Observations	678	930	678	930	418	930	1100

The dependent variable in panel A, columns (a) through (f), and panel B is log male agrarian wages, from the Agricultural Wages of India. In column (g) of panel A, the dependent variable is the measure of wages is the wages of illiterates in all sectors other than manufacturing, imputed by the Employment-Unemployment Rounds collected by the NSSO. All specifications include district fixed effects (apart from (f) in both panels which includes region fixed effects), year dummies, rainfall variables (total rainfall between june and september, rainfall squared and a shock measure taking the value 1 if there is a positive rainfall shock, 0 if no shock and -1 for negative shocks), the log of male and female population for the landless and by land quantiles. Columns (b) onwards additionally control for the interaction of district level resources with average tariff and delicensing reforms. The first stage instruments are the average interaction between between district level resources, industry resource usage and industry tariffs and regulations. The F-Stat on the first stage is 32.19. Manufacturing employment data in column (a) (and (c) in panel A) comes from the NSSO surveys conducted in 1987, 1993 and 1999, column (b) onwards additionally includes 1983. The number of observations in columns (a) (and (c)) are lower than the rest because district identifiers are not available for the 1983 NSS. Column (e) in panel A uses district-industry representative data from the Economic Census for 1991 and 1998.

TABLE 7: Skilled Wage Response to Manufacturing Employment and Agricultural Productivity							
Dependent Variable: Log Male Skilled Wages							
	FE	FE-IV	FE-IVD	FE-IV	FE-IV	FE-IV	FE-IV
	(a)	(b)	(c)	(d)	(e)	(f)	(h)
Log(Manufacturing Employment)	0.028	0.020	0.015	0.017	0.082*	0.217	0.318*
	(0.029)	(0.040)	(0.042)	(0.038)	(0.041)	(0.142)	(0.165)
Log(Literate Employment)	0.132**	0.203**	-	-	0.186*	-0.155	-0.379*
	(0.060)	(0.101)	-	-	(0.112)	(0.145)	(0.206)
Log(At Least Primary Employment)	-	-	0.231**	-	-	-	
	-	-	(0.106)	-	-	-	
Log(Skilled Employment)	-	-	-	0.259*	-	-	
	-	-	-	(0.152)	-	-	
Log(Agricultural Productivity)	0.141	0.140	0.162	0.114	0.059	0.322	0.321*
	(0.375)	(0.208)	(0.209)	(0.205)	(0.222)	(0.250)	(0.173)
Constant	-10.294***	-4.598***	-5.3061	-5.0383	-8.612	-6.553**	-6.900***
	(2.823)	(5.241)	(5.224)	(5.199)	(3.732)	(3.179)	(2.668)
Adjusted R-Squared	0.582	0.582	0.553	0.553	0.559	0.683	0.743
District Fixed Effects	Y	Y	Y	Y	Y	N	Y
Region Time Trend	N	N	N	Y	N	N	N
First Stage F-Stat	-	32.19***	13.087***	32.19***		32.19***	32.19***
Observations	703	1034	1034	1034	1034	1034	1100

The dependent variable is log median male wages of literate workers in the non-agricultural sector, excluding manufacturing. This is imputed from the Employment-Unemployment Rounds collected by the NSSO. All specifications include district fixed effects (apart from (f) which includes region dummies), year dummies, rainfall variables (total rainfall between june and september, rainfall squared and a shock measure taking the value 1 if there is a positive rainfall shock, 0 if no shock and -1 for negative shocks), the log of male and female population for the landless and by land quantiles. Columns (b) onwards additionally control for the interaction of district level resources with average tariff and delicensing reforms. The first stage instruments are the average interaction between between district level resources, industry resource usage and industry tariffs and regulations. The F-Stat on the first stage is 32.19. Manufacturing employment data in column (a) comes from the NSSO surveys conducted in 1987, 1993 and 1999, columns (b) onwards additionally includes 1983.

TABLE 8: Log(Consumption per capita)								
Dependent Variable: Log Male Agricultural Wages								
	FE		FE-TSLS					
					Landless		Landed	
	Landless	Landed	Landless	Landed	Illiterate	Literate	<2 acres	> 2 acres
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(e)
Log(Manufacturing Employment)	0.0203**	0.014*	0.222	0.180**	0.362**	0.251**	0.226***	-0.043
	(0.095)	(0.008)	(0.252)	(0.751)	(0.112)	(0.126)	(0.087)	(0.098)
Log(Literate Employment)	0.022	0.019	0.223	-0.211	-0.289	0.169*	-0.262***	0.034
	(0.022)	(0.017)	(0.230)	(0.082)	(0.081)	(0.102)	(0.080)	(0.124)
Log(Agricultural Productivity)	0.195	0.124*	0.204	0.170**	0.207	0.222*	0.201**	0.149**
	(0.158)	(0.067)	(0.164)	(0.078)	(0.134)	(0.121)	(0.078)	(0.073)
Constant	1.167	2.761	0.484	2.545	-1.634	4.302	0.643	4.461
	(2.791)	(1.264)	(3.003)	(1.474)	(1.265)	(2.033)	(1.415)	(1.522)
Adjusted R-Squared	0.377	0.401	0.381	0.359	0.342	0.327	0.347	0.371
District Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	64867	126063	64867	126063	36042	28825	52668	73395

All specifications include district and year fixed effects, rainfall variables (total rainfall between june and september, rainfall squared and a shock measure taking the value 1 if there is a positive rainfall shock, 0 if no shock and -1 for negative shocks), the log of male and female population by land quantiles. Columns (b) onwards additionally control for the interaction of district level resources with average tariff and delicensing reforms. The first stage instruments are the interaction between between district level resources, industry resource usage and industry tariffs and regulations. The F-Stat on the first stage is 32.19. Manufacturing employment in columns (a) and (b) come from the NSSO surveys in 1987, 1993 and 1999, column (c) onwards additionally includes 1983.

TABLE 8: Poverty and Inequality estimates

Panel A: Dependent Variable: Headcount Rate

	FE	FE-TSLS				
		Rural			Urban	
	(a)	(b)	(c)	(d)	(e)	(f)
Log(Manufacturing Employment)	-0.002	-0.002	-0.088*	-0.127**	0.011	-0.012
	(0.008)	(0.003)	(0.045)	0.059	(0.012)	(0.021)
Log(Literate Manufacturing)	-	0.003	-	0.131*	-	0.027
	-	(0.003)	-	(0.071)	-	(0.023)
Log(Agricultural Productivity)	-0.136*	-0.136*	-0.153*	-0.139*	-0.092	-0.087
	(0.082)	(0.081)	(0.080)	(0.082)	(0.061)	(0.061)
Constant	2.360**	2.374***	2.737**	2.337*	1.353	1.203
	(1.259)	(1.259)	(1.252)	(1.301)	(1.058)	(1.041)

Panel B: Dependent Variable: Poverty Gap

	FE	FE-TSLS				
		Rural			Urban	
	(a)	(b)	(c)	(d)	(e)	(f)
Log(Manufacturing Employment)	-0.003	-0.004	-0.0274	-0.069***	0.011	-0.012
	(0.002)	(0.003)	(0.021)	(0.020)	(0.012)	(0.021)
Log(Literate Manufacturing)	-	0.001		0.077***		0.027
	-	(0.005)		(0.026)		(0.023)
Log(Agricultural Productivity)	-0.042*	-0.041*	-0.052**	-0.059**	-0.092	-0.087
	(0.023)	(0.023)	(0.024)	0.025	(0.061)	(0.061)
Constant	0.629	0.608	0.606**	0.715*	0.667	0.609
	(0.396)	(0.408)	(0.285)	(0.395)	(0.720)	(0.718)
Observations	967	967	967	967	967	967

Panel C: Dependent Variable: Gini Measure of Inequality

	FE	FE-TSLS				
	(a)	(b)	(c)	(d)	(e)	(f)
Log(Manufacturing Employment)	-0.038**	-0.036**	-0.026*	-0.071***	-0.049***	-0.049**
	(0.017)	(0.017)	(0.013)	(0.026)	(0.026)	(0.022)
Log(Literate Manufacturing)	-	0.010	-	0.060***	-	
	-	(0.050)	-	(0.027)	-	
Log(Primary and above Manufacturing)	-	-	-	-	0.033	
	-	-	-	-	(0.029)	
Log(Skilled Manufacturing)	-	-	-	-	-	0.047
	-	-	-	-	-	(0.029)
Log(Agricultural Productivity)	0.046	0.036	0.056	0.062	0.055	0.066
	(0.127)	(0.128)	(0.129)	(0.130)	(0.129)	(0.127)
Constant	0.588	0.603	0.554	0.421	0.532	0.493
	(0.814)	(0.820)	(0.802)	(0.808)	(0.799)	(0.796)
District Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	967	967	967	967	967	967

All specifications include district fixed effects, year dummies, rainfall variables (total rainfall between june and september, rainfall squared and a shock measure taking the value 1 if there is a positive rainfall shock, 0 if no shock and -1 for negative shocks), and the log of male and female population for the landless and by land quantiles. Columns (b) through (f) additionally control for the interaction of district level resources with average tariff and delicensing reforms. The first stage instruments are the average interaction between between district level resources, industry resource usage and industry tariffs and regulations. The F-Stat on the first stage is 32.19. * significant at 10%; ** significant at 5%; *** significant at 1%; Huber-White standard errors are reported in parentheses, standard errors are clustered at a district level in columns (a) and at a region-year level in columns (b) through (f).

Table 10: Started School Between the Age of 5 and 9

Panel A: Dependent Variable: 1 if boy reports having started school

	Landless		Landed	
	Illiterate	Literate	Net Exporters	Net Importers
	(a)	(b)	(c - L)	(c - H)
Log(Manufacturing Employment)	1.048***	-0.287	-0.458***	-0.468***
	(0.306)	(0.134)	(0.534)	(0.175)
Log(Manufacturing Literate)	-.393**	1.055**	0.393*	0.943**
	(0.166)	(0.427)	(0.196)	(0.402)
Log(Agricultural Productivity)	0.895	0.518*	0.682**	0.521
	(0.626)	(0.282)	(0.269)	(0.491)
District Fixed Effects	Y	Y	Y	Y
State*Year Dummies	Y	Y	Y	Y
Data Source	NSS	NSS	NSS	NSS
Adjusted R-Squared	0.210	0.132	0.213	0.2063
Observations	27272	17971	40530	73007

Panel B: Dependent Variable: 1 if girl reports having started school

	Landless		Landed	
	Illiterate	Literate	Net Exporters	Net Importers
	(a)	(b)	(c - L)	(c - H)
Log(Manufacturing Employment)	0.600**	-1.253	-0.233	-0.242
	(0.306)	(0.446)	(0.253)	(0.197)
Log(Manufacturing Literate)	-0.627	0.868**	0.424	0.192
	(0.482)	(0.359)	(0.258)	(0.216)
Log(Agricultural Productivity)	-0.214	0.557	0.609**	0.542*
	0.165	(0.572)	(0.284)	(0.310)
District Fixed Effects	Y	Y	Y	Y
State*Year Dummies	Y	Y	Y	Y
Data Source	NSS	NSS	NSS	NSS
Adjusted R-Squared	0.3648	0.2063	0.3441	0.2063
Observations	20166	22487	34991	63187

Table 11: Started School Between the Age of 5 and 9

Dependent Variable: 1 if individual reports having started school

	Boys Aged 5 to 9		**Girls Aged 5 to 9**	
	Illiterate	Literate	Illiterate	Literate
	(a)	(b)	(c)	(d)
Log(Unskilled Wages)	1.437***	0.429*		
	(0.409)	(0.236)		
Log(Skilled Wages)	0.330*	0.287**		
	(0.133)	(0.133)		
District Fixed Effects	Y	Y		
Year Dummies	Y	Y		
Data Source	NSS	NSS		
Adjusted R Squared	0.066	0.236		
Observations	15063	13576		

Table 12: Structural Parameters of Education Decision Rule for children aged 5 to 9

Panel A: Semi-Reduced Form Approach

	Boys				**Girls**			
	Illiterate	**Literate**	**Illiterate**	**Literate**	**Illiterate**	**Literate**	**Illiterate**	**Literate**
Income	3.19***	2.308***	3.501***	2.617***	1.186	1.404	2.514**	1.032
(Kappa 1)	(0.416)	(0.746)	(1.039)	(0.746)	(0.541)	(0.798)	(1.113)	(1.001)
Returns to Education	0.905***	1.322**	0.905***		-0.189	0.203	-0.189	
(Kappa 2)	(0.177)	(0.595)	(0.177)		(0.368)	(1.878)	(0.368)	
Opportunity Cost	-	-	-0.310		-	-	-1.3274	
(Kappa 4)	-	-	(1.072)		-	-	(1.109)	
Xi-Squared	1.1803	2.059	2.8959		3.1142	13.82	7.2719	
Degrees of Freedom	1	1	2	2	1	1	2	2
Tau	-	0.4	-	0.4	-	0.4	0.4	0.4

Panel B: Instrumental Variables Appraoch

	Boys				**Girls**		**Girls**	
	Illiterate	**Literate**	**Illiterate**	**Literate**	**Illiterate**	**Literate**	**Illiterate**	**Literate**
Income	1.767***	0.501**	1.788***	0.528				
(Kappa 1)	(0.539)	(0.255)	(0.438)	(0.038)				
Returns to Education	0.33*	0.279*	0.33*					
(Kappa 2)	(0.182)	(0.143)	(0.182)					
Opportunity Cost	-	-	-0.021					
(Kappa 4)	-	-	(0.402)					
Degrees of Freedom	0	0	0	0	0	0	0	0
Tau	-	0.4	-	0.4	-	0.4	-	0.4

Appendix A. Results

Appendix B. Data

B.0.3. *Employment and Education Data:* The Employment-Unemployment surveys conducted by the National Sample Survey Organization (NSSO) are the main data source for all employment and education data. I use the "thick"employment-rounds (round 10) conducted in 1983, 1987-88, 1993-94 and 1999-2000. The surveys collect information on approximately 75,000 rural and 45,000 urban households and usually cover all states in India. Employment by industry is constructed at a region level, where a regions consist of 4 to 6 neighboring districts within the same agro-climatic zone.

I use two different education thresholds to define a high-education group. These thresholds are determined by the data: in 1983, approximately 40% of manufacturing employees were classified as illiterate, and 59% had less than primary education. The employment surveys ask households and individuals within household to list their primary and subsidiary usual occupations as well as their primary and subsidiary current occupations. I use an individual's primary occupation status to define the sector he is working in. For the all specifications involving agricultural or manufacturing employment, I restrict the sample to working age males and females between the age of 25 and 55.[60] A weekly time-use recall allows me to capture the days of work devoted to agricultural labor market, own farm and manufacturing activities over the course of the preceding week.[61]

In addition, I use employment data from two waves of the Economic Census. The Economic Census is a country-wide census of all economic activities other than those related to crop production and plantation. It contains basic data on the principal activity conducted by the firm and the number of family and hired male and female labor employed.[62] Since it is a census it captures district-industry level data on employment for both the formal and informal sector.[63]

Tables A.1a and A.1b in the appendix present summary statistics of the main variables used in the analysis. Approximately 6% of rural working age males are employed in the manufacturing sector between 1983 and 1999; during the same period, agricultural employment drops from approximately 70% of the population to 64%. There is however substantial variation both within and across states in the proportion

[60]I test the sensitivity of my results to including all working aged individuals over the age of 25, and also to including only males in the sample. The magnitude and sign of the coefficient estimates in the wage equations are largely unchanged by including a broader age sample; when including only male workers the estimated coefficients decrease in magnitude by about 10 to 20% but the sign and level of statistical significance remains largely unchanged. These results have not been presented in the paper but are available upon request.

[61]Surveying takes place over the course of a calendar year, with fieldwork spread out uniformly over all four quarters within each region. Therefore it is not necessary to account for seasonal variation when using these measures.

[62]The Economic Census is conceived primarily as a means of putting together a frame of non-farm firms from which some would be selected to conduct more detailed follow-up surveys on the unorganized sector.

[63]Statistics on firms in India are collected separately for the registered and unregistered sector. India can be broadly split into two components: registered units which fall under the scope of the 1948 Factory Act and unregistered units. Unregistered units can be further divided into Small and Medium Sized Firms which fall into special consideration in the regulatory framework and the unorganized sector. The registered sector consists broadly of all firms using more than 10 individuals and using power, or more than 25 individuals if not using power. The Annual Survey of Industries (ASI), the principal data source on the registered sector, takes the State and 3-digit industry group as the sampling stratum; therefore it may not provide an accurate depiction of district level employment in the registered sector using this data.

Table A.1.1: Robustness Checks

	Agricultural Prices					Groundwater Table Depth				
	Dependent Varible: Farm-Harvest Price of Crop, in 1987 Rs.					Dependent Varible: Proportion of Villages in a District under a water table threshold.				
	Jowar (a)	Cotton (b)	Rice (c)	Wheat (d)	Sugar (d)	<10m (a)	<20m (b)	<30m (c)	<40m (d)	<50m (d)
Deregulated*Groundwater	0.039	-0.527	0.3188	-0.6311	0.079	-0.013**	-0.004	-0.003	-0.002	0
*Agro Proportion	(0.2050)	(0.7130)	(0.5228)	(0.6177)	(0.0650)	(0.0060)	(0.0050)	(0.0040)	(0.0030)	(0.0020)
Deregulated*Uncultivable Land	0.126	0.013	-0.118	-0.019	-0.105	-0.002	-0.002	-0.001	-0.003	-0.003
*Agro Proportion	(0.1320)	(0.3340)	(0.0990)	(0.1020)	(0.1320)	(0.0030)	(0.0030)	(0.0030)	(0.0020)	(0.0020)
Deregulated*Forest Cover	0.62066	-0.16631	0.3242	-0.6311	0.105***	-0.009	-0.005	-0.003	0.004*	-0.002
*Wood Proportion	(0.4205)	(0.1343)	(0.3452)	(0.6177)	(0.2620)	(0.0080)	(0.0060)	(0.0060)	(0.0024)	(0.0020)
Deregulated*Ceramics Minerals	0.044	-0.203	0.006	-0.003	-0.002	0.003	0.004	0.005	0.002	0.004
*Ceramics Proportion	(0.0480)	(0.1360)	(0.0050)	(0.0040)	(0.0020)	(0.0040)	(0.0030)	(0.0030)	(0.0040)	(0.0020)
Deregulated*Construction Minerals	-0.001	0.012	-0.002	0.002	-0.003	0.002	0.001	0.001	0.001	0
*Construction Proportion	(0.0030)	(0.0090)	(0.0060)	(0.0040)	(0.0020)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Deregulated*Coal	-0.002	0.003	0.006	0.004	-0.003	0.212**	0.094	0.058	0.046	0.022
*Energy Proportion	(0.0050)	(0.0140)	(0.0050)	(0.0060)	(0.0040)	(0.0900)	(0.0620)	(0.0460)	(0.0310)	(0.0180)
Deregulated*Electricity	-0.035***	0.018	-0.156	0.352	-0.115	0.004	0.004	0.003	0.002	0.001
*Energy Proportion	(0.0110)	(1.4220)	(0.3790)	(0.3850)	(0.1870)	(0.0030)	(0.0020)	(0.0020)	(0.0020)	(0.0010)
Constant	32.38**	1.777	11.8	-15.937	-4.747	-1.859	-1.778	-3.501	-5.109**	-0.847
	(13.6900)	(33.8630)	(11.1180)	(12.7640)	(7.7890)	(5.6280)	(4.4650)	(3.8390)	(2.4410)	(1.3910)
Observations	721	614	844	652	586	313	313	313	313	313
F-Statistic over Triple Interactions	2.35	1.29	0.95	0.88	2.73	1.7	1.19	1.1	1.79	0.88

* significant at 10%; ** significant at 5%; *** significant at 1%

All specifications include district fixed effects, year dummies, rainfall variables (total rainfall between june and september, rainfall squared and a shock measure taking the value 1 if there is a positive rainfall shock, 0 if no shock and -1 for negative shocks), the log of male and female population for the landless and by land quantiles. In addition, they include the interaction of district level resources with average tariff and delicensing reforms.

of the workforce employed in agriculture and manufacturing. Figure A.1 displays the proportion of the working age population working in manufacturing and agriculture between 1983 and 1999. There is substantial variation across states in both the proportion of the population employed in manufacturing and in the pattern of employment over time. Figure A.2 depicts variation in the proportion of individuals employed in manufacturing within states. I exploit the substantial variation in industrial employment across districts and time to identify the effects of a change in the size and composition of the industrial workforce on wages in the agrarian sector.

B.0.4. *Wages.* Wage data on agricultural wages come from the Agricultural Wages in India (AWI) collected by the Directorate of Economics and Statist at the Ministry of Agriculture. Data are collected on a monthly basis for various agricultural operations, including ploughing, sowing, weeding, transplanting, reaping, harvesting and field labor. Where possible, separate wages are collected for males, females and children. Data are not collected for all districts within a state and data reporting is uneven across time and districts. The missing observations for male wages however do not appear to be related to any characteristic of local labor markets but rather to failures in the data collection system.[64] Wages are deflated using the State level Rural Agricultural Labor Price Index published by the Indian Labor Bureau. I additionally impute literate non-farm wages from weekly wage labor income in the NSSO Employment surveys. This data is still being processed and will be used in the next draft of the paper.

B.0.5. *Natural Resources.* Raw material endowments are grouped according to the main categories of industrial usage following the Mineral Atlas of India and various NCAER Economic Plans. A more detailed description of the endowments is given in the data appendix, appendix 2.3. To measure the mineral and metal endowment, I have built a geological database of India which allows me to calculate stocks of recoverable mineral and metal resources at any location in India. Data on ore deposits were obtained from the National Mineral Atlas and Geological Map of India published by the Geological Survey of India (GSI). Figure A.4 displays the data in its raw form - the map details mineral deposits by location. These were manually geo-coded.

I use two different measures of mineral and metal endowments to capture district level endowments. The first is to define the endowment as all resources found within the district boundaries. The second is to define the endowment as all resources found within 100 miles of the region boundaries.[65] Region level data on soil conditions and agricultural yields comes from the India Agriculture and Climate Dataset compiled by Robert E. Evenson and James W. McKinsey, Jr., using data from the Directorate of Economics and Statistics within the Indian Ministry of Agriculture. Groundwater data comes from the Central Groundwater Board; I use the natural recharge rate as a measure of the physical groundwater potential within a district.[66] I use the area covered by forest within a district from the Forest Survey of India (1993) as my measure of wood endowments.

[64]The actual data collection is left to individual states, who assign village level officials to collect the locally common current wage on a monthly basis (Himanshu, 2004). Since there appears to be no check or enforcement of data collection at a state level by the Directorate of Economics and Statistics, data is regularly missing at a district or even state level.

[65]This second method of computing regional variation in mineral resources is a work progress.

[66]The recharge rate is the rate at which a groundwater aquifer is replenished. It is determined by precipitation, soil and rock type, rates of evapotranspiration, infiltration and runoff.

B.0.6. *Factor Intensity.* Data from the Input-Output Matrix is used to build measures of the intensity of use of different inputs by industry. The measures reflect the share of costs accounted for by a given category of inputs (energy - coal and electricity; raw materials - unprocessed metals, forestry products, construction minerals, ceramic and refractory materials, industrial minerals (for the fertilizer, paint and chemical industries) and labour. The definition of factor intensity used is given by:

$$\text{Factor Intensity}_{k,l} = \frac{\text{Cost Share of Input k}}{\text{Cost Share of Labor}}$$

Since the continuous factor intensity measure will at any moment in time reflect both the industry production function as well as the distribution of input prices within the economy, I use discrete measures which are designed to capture the broader technological requirements of an industry. I use two measures of industry technological requirements to capture the intensity with which an industry uses a given resource. The quantile measure separates all non-zero factor intensities into quantiles, while the dummy median measure codes an industry as using a factor if it lies above the median factor intensity for that raw material among all non-zero factor intensities. Table xx lists the industries which lie in the top quintile of the factor intensity measure for wood, ceramics, chemicals and metal inputs. While there are a few surprises, the list largely conforms to expectations: the iron producing and metal extracting industries are metal intensive; the glass, cement and refractory industries are ceramics minerals intensive while the aluminium processing, chemicals and coke oven industries are energy intensive.

B.0.7. *Industrial Policy and Regulations.* I use tariff measures and delicensing reforms compiled by Aghion, Burgess, Redding and Zillibotti (2008). A more complete description of the Indian policy environment can be found in appendix A.2.6.

B.1. **Definitions of Raw Material Inputs.**

- Raw Material Groups using definitions from Mineral Atlas of India/NCAER Economic Plans (1960s-1990s)
  - *Agriculture:* Proportion of land classified as barren; Mean Kharif Rainfall; Groundwater Recharge Rate.
  - *Forestry* : Proportion of District Covered in Forests
  - *Metal*: Alumnium, Chromium, Copper, Iron Ore, Lead, Manganese, Zinc.
  - *Ceramics:* Kaolin, Feldspar, Glass and Foundry Sand.
  - *Construction Sector:* Calcite, China Clay, Limestone, Sandstone.
  - *"Strategic"* Chemicals - Asbestos, Baryte, Dolomite, Fluorite and Limonite.
  - *Energy:* Coking and Non-Coking Coal and Electricity Prices

B.2. **Mineral Atlas of India and Density Measures.**

B.3. **Industries by Raw Materials.**

Exerpt from the National Mineral Atlas – Map showing Limestone, Barytes, and Asbestos Deposits in Cuddapah, Andhra Pradesh

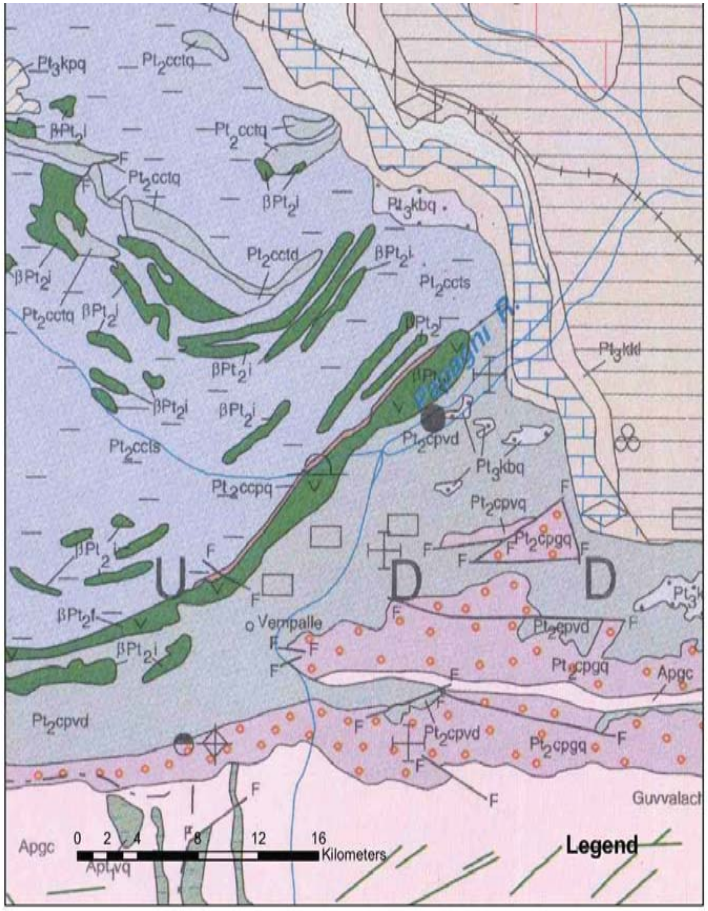

Table A.1: Physical Costs of Transportation of Raw Materials

Mineral Group (Compiled by end use)	Density		Domestic Price/Ton		Density /Price
	Mean	SD	Mean	SD	
Tobacco	320	-	36000	-	0.01
Timber	400	-	-	-	-
Cotton (w Lint)	560	-	12900	-	0.04
Rice	580	-	18400	-	0.03
Sugar (Raw)	960	-	32000	-	0.03
Water	1000	-	-	-	-
Coal (Bitumious)	1200	-	390	-	3.08
Clay (Kaolin)	2160	-	180	-	12.00
Ceramics	2620	300.00	187	160.10	14.04
Construction	2660	120.00	203	40.41	13.08
Chemicals	3684	1100.00	509	297.59	7.24
Steel	4610	2410.00	1050	1257.98	4.39
Metals	8850	6400.00	45135	61042.80	0.20
Gold	19300	-	-	4.40E+09	4.39E-06

$$Density = \frac{Mass}{Volume}$$

Specific Gravity, Source: Compiled from Various Engineering Texts
Prices: a) Minerals - Domestic Prices for Ores in 1990 Rs at Mine Head. b) Crops - Wholesale price of Crops, from Agricultural Situation in India. Source Mineral Data: Mineral Yearbook, Indian Bureau of Mines.

Industries in top quintiles for different raw materials

	Agriculture	Wood	Ceramics	Chemicals
1	Slaughtering of Animals	Bidi Manufacture	Refractory and Structural Clay Products	Tea Processing
2	Dairy Industry	Sawing of Wood	Glass and Glass Products	Organic and Inorganic Chemicals
3	Fruit and Vegetable Preparation	Veneer Manufacture	Earthen and Plaster Products	Fertilizers and Pesticides
4	Grain Milling	Structural Wooden Products	Non-Structural Ceramics	Paints and Varnishes
5	Sugar Refinement	Wooden and Cane Boxes	Cement and Plaster	Drugs and Medecines
6	Gur Making	Wood Industrial Products	Mica Products	Perfumes and Cosmetics
7	Salt Production	Cork Products	Structural Stone Goods	Refining of Sugar
8	Ghee Production	Wooden Furniture	Asbestos Cement	Tyre and Tubes
9	Vegetable Oil Production	Bamboo Furniture	Misc Non-Metallic Mineral Products	Footwear
10	Animal Oil Production	Wooden Products nec	Radiographic Aparatus	Rubber Products

	Metal	Energy	Proportion of Literate Labor
1	Fertilizers and Pesticides	Pulp, Paper and Paper Board	Plastics
2	Refractory and Structural Clay	Containers and Boxes	Drugs and Medecines
3	Semi Finished Iron	Paper n.e.c	Coke Oven Products
4	Ferro-Alloys	Organic and Inorganic Chemicals	Batteries
5	Copper Manufacturing	Coke Oven Products	Printing and Publishing of Books
6	Brass Manufacturing	Coal and Coal Tar Products n.e.c	Tyre and Tubes
7	Aluminium Manufacturing	Cement, Lime and Plaster	Refined Petroleum Products
8	Zine Manufacturing	Ferro-Alloys	Metal Furniture and Fixtures
9	Processing of Metal Scraps	Brass Manufacturing	Insulated Wires and Cables
10	Other Non-Ferous Metal Industries	Aluminium Manufacturing	Electrical Equipment n.e.c.

B.4. **Description of Indian Industrial Environment.** *Post Independence Indian Industrial Policy* The Nehruvian view of economic development, dominant in the period ensuing independence, endorsed the need for economic development driven by state directed economic activity and central planning. India's economic strategy during this period is articulated in a series of Five Year Plans which laid down the philosophy and strategies for development in the subsequent years. The Second Five Year plan (1957-1962) laid down the foundations of India industrial policy until the 1990s (Krueger and Chiony, 2002). The "Licence Raj" entailed a series of rigid controls over the establishment, capacity, investment and production of the industrial sector (Srinivasan, 2000). Certain industries were reserved for public ownership, whilst others were circumscribed to public-private partnerships. Targets were established for permissable production levels and licences were required to expand capacity beyond that permitted. Import-substitution policies, aimed at stimulating growth, gave rise to a restrictive trade regime with high average and peak nominal tariffs and non-tariff barriers.

Reforms initiated during the 1980s and 1990s represented a major break from the inward oriented, state directed and public sector driven' approach pursued since independence (Srinivasan, 2000). The first phase of reforms aimed at unraveling the regulatory framework occurred in 1985, after a political crisis sparked by the assassination of the incumbent prime minister, Indira Gandhi, in 1984. During this period, a subset of manufacturing industries was removed from the jurisdiction of the license regime whilst others were allowed more flexibility in their functioning. The government adopted an expansionary fiscal stance leading to large fiscal deficits, high levels of external debt and falling foreign reserves by 1991. A serious macroeconomic and balance of payments crisis ensued; in the wake of this crisis, systemic structural reforms were introduced alongside more "traditional" stabilization measures (Krueger and Chiony, 2002). Reforms to industrial policy, trade and payments regime, excise tariffs and financial sector substantially decreased the role of the public sector in economic activity and altered the institutional framework and profitability of private sector firms. Reforms affected all firms in that industry or mining sector irrespective of location.

Fiscal reforms initiated by the central government resulted in substantial reductions in customs duties and some reduction and rationalization of union excise duties (excise duties levied by the central government). Both average and peak tariffs have been drastically reduced since 1990-91, from an average nominal tariff of 125% to 25% in 1997-98. Excise duties were simplified and harmonized across products.

Indian Industrial Location and Expansion - Industrial Policies and Raw Materials:

In the model presented in section 3 I assume that firms take into account natural resource prices and labor costs when making their location decisions. I argue that this assumption is appropriate prior in the period prior to 1985 during which Indian industrial policy restricted the ability of larger regulated industrial units to choose their locations. Location decisions appear to have been guided by political concerns (such as providing employment to "backward" areas or targeting balanced regional development (Mookherjee, 1995)) as well as by consideration for costs. However it does appear that access to inputs were an important condition precondition to the establishment of an industry.

The Industrial Programs component of the Fourth Five Year Plans suggests that access to inputs was one of the primary motivations given for the dispersal of industries across geographic areas. “The present pattern of regional distribution of industries has been formed by the accessibility of base raw materials. Thus agriculture based industries are mostly concentrated in North and South Bihar, while industries based on minerals are to be found chiefly in Chotanagpur, with a dispersal of some small-scale industries in South Bihar.” (NCAER Bihar, 1969). The inter-linkages between industries appears to also have been taken into consideration, as this excerpt on the production of sulphuric acid indicates: “The location of these suggested capacities will be linked with the consuming industries to avoid transportation” (NCAER Andhra Pradesh, 1969). Conditional on accessibility of resources, the location of final demand was also a component to be considered, as demonstrated by this excerpt on the establishment of fertilizer factories: “The demand for fertilizers is likely to be large in each State and it is considered justifiable to contemplate setting up of suitable capacities in each State to meet the demand arising within the State, provided of course suitable raw materials and other facilities are there for the production of fertilizers.” (NCAER Andhra Pradesh, 1969).

Appendix C. Derivations

C.1. **Obtaining the Wage Equations.** Linearizing labor demand and supply for manual and skilled workers, we get:

$$\begin{aligned}
\text{Labor Demand}_{manual}^{Manuacturing} &= a_0 + a_1 w_{mdt} + a_2 w_{sdt} + a_3 X_{dt} + a_4 Z_{dt}^{M} + u_{dt}^{a} \\
\text{Labor Demand}_{skilled}^{Manuacturing} &= b_0 + b_1 w_{mdt} + b_2 w_{sd,t} + b_3 X_{dt} + b_4 Z_{dt}^{M} + u_{dt}^{b} \\
\text{Labor Demand}_{manual}^{Agriculture} &= c_0 + c_1 w_{mdt} + c_2 w_{sdt} + c_3 X_{dt} + c_4 Z_{dt}^{A} + u_{dt}^{c} \\
\text{Labor Demand}_{skilled}^{Agriculture} &= d_0 + d_1 w_{mdt} + d_2 w_{sdt} + d_3 X_{dt} + d_4 Z_{dt}^{A} + u_{dt}^{d} \\
\text{Labor Supply}_{manual} &= f_0 + f_1 w_{mdt} + f_2 w_{sdt} + f_3 X_{dt} + f_4 Z_{dt}^{S} + u_{dt}^{d} \\
\text{Labor Supply}_{skilled} &= g_0 + g_1 w_{mdt} + g_2 w_{sdt} + g_3 X_{dt} + g_4 Z_{dt}^{S} + u_{dt}^{d}
\end{aligned}$$

Setting labor demand equal to labor supply for both manual and skilled labor, but keeping manufacturing labor demand in it's raw form, we get wages for skilled and manual labor:

$$\begin{aligned}
w_{mdt} &= \left(\frac{1}{f_1 - c_1}\right)[(c_0 - f_0) + (c_2 - f_2) w_{sd,t} + (c_3 - f_3) X_{dt} + c_4 Z_{dt}^{A} - f_4 Z_{dt}^{S} + \mathrm{L}_{manual,dt}^{Manufacturing} + u_{dt}^{c} - u_{dt}^{f}] \\
w_{sdt} &= \left(\frac{1}{g_2 - d_2}\right)[(d_0 - g_0) + (d_1 - g_1) w_{md,t} + (d_3 - g_3) X_{dt} + d_4 Z_{dt}^{A} - g_4 Z_{dt}^{S} + \mathrm{L}_{skilled,dt}^{Manufacturing} + u_{dt}^{f} - u_{dt}^{c}]
\end{aligned}$$

Solving the system of simultaneous equations, we get:

$$\begin{aligned}
(26)\ w_{mdt} &= \alpha_0 + \alpha_1 \mathrm{L}_{manual,dt}^{M} + \alpha_2 \mathrm{L}_{skilled}^{M} + \alpha_3 \text{Ag Productivity}_{dt} + \alpha_4 Z_{dt}^{A} + \alpha_5 Z_{dt}^{S} + \alpha_6 X_{dt} + \epsilon_{dt} \\
(27)\ w_{sdt} &= \beta_0 + \beta_1 \mathrm{L}_{manual,dt}^{M} + \beta_2 \mathrm{L}_{skilled}^{M} + \beta_3 \text{Ag Productivity}_{dt} + \beta_4 Z_{dt}^{A} + \beta_5 Z_{dt}^{S} + \beta_6 X_{dt} + \varepsilon_{dt}
\end{aligned}$$

where $\xi = \left(\frac{(g_2-d_2)(f_1-c_1)}{(g_2-d_2)(f_1-c_1)-(c_2-f_2)(d_1-g_1)}\right)$, $\alpha_1 = \xi * \frac{1}{f_1-c_1}$, $\alpha_2 = \alpha_1 * \left(\frac{c_2-f_2}{g_2-d_2}\right)$. $\beta_2 = \xi * \frac{1}{g_2-d_2}$ and $\beta_1 = \beta_1 * \left(\frac{d_1-g_1}{f_1-c_1}\right)$.

Under the assumption that manual and skilled labor demand are complements in agricultural and manufacturing, $c_1 < 0$, $c_2 < 0$, $d_1 < 0$ and $d_2 < 0$. Under the assumption that total labor supply is perfectly inelastic, it follows that $f_1 + g_1 + f_2 + g_2 = 0$ and that $f_1 > 0$, $g_1 < 0$, $f_2 < 0$ and $g_2 > 0$. $\xi > 0$ if the own-price labor supply and demand response is greater than the cross-price labor demand and supply response for both manual and skilled labor. Therefore $\alpha_1 > 0$ and $\beta_2 > 0$. Under the same conditions, $\alpha_2 < \alpha_1$ since $f_1 + f_2 > 0$ and $c_1 - c_2 < 0$ therefore $f_1 - c_1 > -c_2 - f_2 > c_2 - f_2$. Similarly $\beta_1 < \beta_2$.

α_1 increases as agricultural manual labor demand becomes more wage elastic - the intuition here is that as agricultural labor demand becomes less responsive to changes in wages, a large increase in agrarian wages is required to "release" a given quantity of labor from agriculture. α_1 increases as the manual wage elasticity of manual labor supply decreases - the intuition is that if manual labor supply is unresponsive to manual labor wages, an increase in manual labor demand in the manufacturing sector will have a greater impact on wages than if the labor supply of manual laborers is responsive to changes in the wage. Similarly, α_2 increases as skilled labor supply becomes more wage inelastic and increases as the skilled agricultural labor demand becomes more inelastic.

C.2. **Obtaining the Income equation.** Household income is given by the following equation:

$$y_h = w_u\lambda_h + w_s * \xi_h + \Pi^{Ag}(w_u, w_s, Land_h, \theta, s_h^a, X_h, p, Rain) + M_h \tag{28}$$

where y_h denotes household income, λ denotes the proportion of household time spent working in the manual labor market, ξ denotes the proportion of household time spent in the skilled labor market. Π^{Ag} is the household cultivation profit function, which is a function of household land endowments, $Land_h$, the agricultural technology θ, the education of the household head, farm assets (X_h) and p denotes local prices of inputs (for example, the price of agricultural seeds). Rain denotes district level weather. M_h denotes all other sources of household income, for example income earned through the bullock rental market or through local non-agricultural enterprizes.

The specifications for wages were derived in section (**??**) and empirically estimated in section (**??**).

Under the assumption that agricultural productivity alters the return to land and education assets, the relationship between land and education will vary across districts according to district level agricultural productivity (Foster and Rosenzweig, 1996). In addition, if large landowners hire more labor on the labor market, the relationship between wages and agricultural profits will vary according to land holdings. A linear approximation to the cultivation profit function is given by:

$$\begin{aligned}\Pi^{Ag}_{hdt} \quad = \quad & \zeta_0 + \zeta_1 w_{mdt} + \zeta_2 w_{sdt} + \zeta_3\theta_{dt} + \zeta_4 Land_{hdt} + \zeta_5 s_{hdt} + \zeta_6\theta_{dt} * A_{hdt} \\ & +\zeta_7\theta_{dt} * s_{hdt} + \zeta_8 Land_{hdt} * w_{mdt} + \zeta_9 Land_{hdt} * w_{sdt} + \zeta_{10} Rain_{dt} \\ & +\zeta_{11}\theta_{dt} * w_{mdt} + \zeta_{12}\theta_{dt} * w_{sdt} + \mu_h + \mu_d + \nu_{hdt}\end{aligned} \tag{29}$$

where the error term contains household level farm assets, local prices and potentially the interaction of these term with our measure of agricultural productivity.

Substituting the empirical specifications for the profit function (29) and wages (**??** and 16) into incomes and assuming (at first) that there is no variation across households in the time devoted to the unskilled and skilled labor market:

$$
\begin{aligned}
y_{mdt} \quad = \quad & g_0 + g_1 \mathrm{E}^M_{total,dt} + g_2 \mathrm{E}^M_{sdt} + g_3\theta_{dt} + g_4 A_{dt} + g_5 Land_{hdt} + g_6 s^a_{hdt} + g_7 \mathrm{E}^M_{total,dt} * Land_{hdt} \\
& + g_8 \mathrm{E}^M_{sdt} * Land_{hdt} + g_9\theta_{dt} * Land_{hdt} + g_{10} s_{hdt} * Land_{hdt} + g_{11}\mathrm{E}^M_{mdt} * \theta_{dt} + g_{12}\mathrm{E}^M_{sdt} * \theta_{dt} \\
& + g_{13} s * \theta + g_{14} A_{dt} * Land_{hdt} + g_{15} A_{dt} * \theta + g_{16} Rain + g_{17}\theta^2 + \mu_h + \mu_d + \nu_{hdt}
\end{aligned} \tag{30}
$$

where the g terms are linear functions of the amount of time devoted to the skilled and unskilled labor market, where: $g_0 = \zeta_0 + \zeta_1\alpha_0 + \zeta_2\beta_0 + \lambda\alpha_0 + \xi\beta_0, g_1 = \alpha_1(\zeta_1+\lambda) + \beta_1(\zeta_2+\xi), g_2 = \alpha_2(\zeta_1+\lambda) + \beta_2(\zeta_2+\xi), g_3 = \alpha_3(\zeta_1+\lambda) + \beta_3(\zeta_2+\xi) + \zeta_{11}\alpha_0 + \zeta_{12}\beta_0, g_4 = \alpha_4(\zeta_1+\lambda) + \beta_4(\zeta_2+\xi), g_5 = \zeta_4, g_6 = \zeta_5, g_7 = \zeta_8\alpha_1 + \zeta_9\beta_1, g_8 = \zeta_8\alpha_2 + \zeta_9\beta_2, g_9 = \zeta_6, g_{10} = \zeta_6, g_{11} = \zeta_{11}\alpha_1 + \zeta_{12}\beta_1, g_{12} = \zeta_{11}\alpha_2 + \zeta_{12}\beta_2, g_{13} = \zeta_7, g_{14} = \zeta_8\alpha_4 + \zeta_9\beta_4, g_{15} = \zeta_{11}\alpha_4 + \zeta_{12}\beta_4, g_{16} = \zeta_{10}, g_{17} = \zeta_{11}\alpha_3 + \zeta_{12}\beta_3$. The error term includes other sources of household income, household level farm assets, local prices and potentially the interaction of these term with agricultural productivity.

I estimate a compressed version of this regression for landed and landless households, where I separate households into groups according to their landholding status and the education of the household head. As such, I separate households into groups according to the likelihood that they are working in the skilled and unskilled labor market, as well as whether they are earning profits from agricultural activities. The coefficients on agricultural productivity and predicted manufacturing employment capture the average effect of a change in these terms on households within that endowment category:

$$y_{mdt} = \gamma_0 + \gamma_1 \mathrm{E}^M_{total,dt} + \gamma_2 \mathrm{E}^M_{sdt} + \gamma_3\theta_{dt} + \gamma_4 A_{dt} + \gamma_5 Land_{hdt} + \gamma_6 s^a_{hdt} + \mu_h + \mu_d + \nu_{hdt}$$

where $\gamma_1 = \alpha_1(\zeta_1+\lambda) + \beta_1(\zeta_2+\xi) + Land_h * (\zeta_8\alpha_1 + \zeta_9\beta_1)$ $\gamma_2 = \alpha_2(\zeta_1+\lambda) + \beta_2(\zeta_2+\xi) + Land_h * (\zeta_8\alpha_2 + \zeta_9\beta_2)$ $\gamma_3 = \alpha_3(\zeta_1+\lambda) + \beta_3(\zeta_2+\xi) + \zeta_{11}\alpha_0 + \zeta_{12}\beta_0 + Land_h\zeta_6 + s_h\zeta_7$

C.3. **Alternative specification for the wage equation.** In the first stage specification, employment in industry i is written as a function of industry i's policy changes, industry i's resource usage and regional endowments. The error term in the first stage specification includes the local wage, all other industries' policy changes and potentially the interaction of industry i's policies with industry j's policies.

For these omissions to violate my identification strategy, the omitted variables which are found in the error term of the first stage specification need to be correlated with the triple interaction terms *as well as* with unobserved determinants of the wage regression. If the instrumental variables are valid then the *partial* correlation between the agrarian wage and each instrumental variable should be zero. If policy changes are correlated across industries over time, the covariance between the interaction between own-industry tariffs, own-industry resource use and regional raw material endowments and any interaction involving industry j's tariffs may not be zero.

If the only overlap between manufacturing and agriculture is in the labor market, it is highly unlikely that these additional terms are partially correlated with the wage error term. In addition, if there is another plausible channel through which the additional interactions directly enter into the second stage regressions, these effects should in any case be largely absorbed by the interaction of average industry tariffs and local natural resource characteristics included in the second stage. Finally the nature of the policy reforms conducted greatly reduces the possibility of concern about restrictions on the first stage specification. Topalova (2004) has shown that the differential changes in import tariffs across industries during the trade policy reforms conducted between 1991 and 1997 were unrelated to the state of the industries at the beginning of the reform.

However, in order to ensure that my results are robust to this alternative specifications, I estimate two alternative specification. In the first, I allow industry i to be affected by industry j's tariff change through the labor market. The specifications is derived below. In the second, I take a more general approach and allow industry i's employment to vary with the policy changes across all industries. With a long panel of industries, this alternative specification could in principal be conducted. With only four data points for each industry across time, for reasons of parsimony I do not attempt this specification. Therefore, for reasons of parsimony, I use 2-digit industry categories in this specification.

In the following, I allow shocks in industry j to enter into industry i's employment through the labor market, notably through the wage. Write manufacturing employment in industry i, district d, time t as:

$$(31)\; l_{i,d,t} = \beta_0 + \beta_1\tau_{i,t} + \beta_2\tau_{i,t} * r_i + \beta_3\tau_{i,t} * r_i * n_d + \beta_4\tau_{i,t} * n_d + \beta_5 r_i * n_d$$
$$+\beta_6 w_{d,t} + \beta_7 w_{d,t} * s_i + \beta_8\tau_{i,t} * w_{d,t} + \beta_9\tau_{i,t} * w_{d,t} * s_i + \beta_{10}\tau_{i,t}s_i + \beta_{11}X_{d,t} + \delta_d + \delta_t + u_{i,d,t}$$

Where $\tau_{i,t}$ is the import tariff covering products in an industry at a given moment in time, n_d is the natural resource stock in district d, r_i is a measure of natural resource use, $w_{d,t}$ are equilibrium district wages at time t, s_i is a measure of labor usage, $X_{d,t}$ are other explanatory variables. δ_d are district level fixed effects and δ_t are time dummies. The only difference between equation xx and the first stage equation introduced in the main body of the text are terms $w_{d,t}$ through $\tau_{i,t} * s_i$. The presence of $w_{d,t}$ in equation (5) highlights the simultaneity problem since in a structural equation equilibrium employment is clearly a function of equilibrium wages.

Aggregating equation this equation over J industries to obtain employment in the manufacturing sector at a district level:

$$(32)\; l_{d,t} = J\beta_0 + J\beta_1\overline{\tau}_t + J\beta_2\overline{\tau r}_t + J\beta_3\overline{\tau r}_t * n_d + J\beta_4\overline{\tau}_t * n_d + J\beta_5\overline{r} * n_d$$
$$+J\beta_6 w_{d,t} + J\beta_7\overline{s}w_{d,t} + J\beta_8\overline{\tau}_t * w_{d,t} + J\beta_9\overline{\tau s}_t * w_{d,t} + J\beta_{10}\overline{\tau s}_t + J\beta_{11}X_{d,t} + J\delta_d + J\overline{u}_{d,t}$$

where $\overline{x} = \frac{1}{J}\sum_i^J x_i$. Since district and time dummies are included in the specification, only variables that vary at a district time level are identified leaving:

$$(33)\quad l_{d,t} = J\beta_0 + J\beta_3\overline{\tau r}_t * n_d + J\beta_4\overline{\tau}_t * n_d + J\beta_6 w_{d,t} + J\beta_7 w_{d,t} * \overline{s} + J\beta_8 w_{d,t} * \overline{\tau}_t$$
$$+J\beta_9 w_{d,t} * \overline{\tau s}_t + J\beta_{11}X_{d,t} + J\delta_d + J\overline{u}_{d,t}$$

Inserting (33) into (31) we get:[67]

$$
\begin{array}{lrcl}
(34) & l_{d,t} & = & b_0 + b_1\overline{\tau r}_t * n_d + b_2\overline{\tau}_t * n_d + b_{3,t}l_{d,t} + b_{4,t}X_{d,t} + D_{d,t} + \xi_{d,t} \\
(35) & b_3 & = & \alpha_2(\beta_6 - \beta_7\overline{s} - \beta_8\overline{\tau}_t - \beta_9\overline{\tau s}_t) \\
(36) & l_{d,t} & = & \frac{1}{(b_{3,t})}[b_0 + b_1\overline{\tau r}_t * n_d + b_2\overline{\tau}_t * n_d + b_{4,t}X_{d,t} + D_{d,t} + \xi_{d,t}]
\end{array}
$$

Inserting district level wages into the industry-district level wage regression, and ignoring interactions of industry level variables with district and time dummies:[68]

$$
\begin{array}{lrcl}
(37) & l_{i,d,t} & = & a_0 + a_1\tau_{i,t} + a_2\tau_{i,t} * r_i + a_3\tau_{i,t} * r_i * n_d + a_4\tau_{i,t} * n_d + a_5 r_i * n_d + a_6X_{d,t} \\
& & & +a_7X_{d,t} * s_i + a_8X_{d,t} * \tau_{i,t} + a_9X_{d,t} * \tau_{i,t} * s_i + a_{10}l_{d,t} + a_{11}l_{d,t} * s_i + a_{12}l_{d,t} * \tau_{i,t} \\
& & & +a_{13}l_{d,t} * \tau_{i,t} * s_i + a_{14}\tau_{i,t} * s_i + D_d + D_t + \epsilon_{i,d,t}
\end{array}
$$

Substituting equation (34) into industry-level employment above:[69]

$$
\begin{array}{lrcl}
(38) & l_{i,d,t} & = & c_0 + c_1\tau_{i,t} + c_2\tau_{i,t} * r_i + c_3\tau_{i,t} * r_i * n_d + c_4\tau_{i,t} * n_d + c_5 r_i * n_d + c_6X_{d,t} \\
& & + & c_7X_{d,t} * s_i + c_8X_{d,t} * \tau_{i,t} + c_9X_{d,t} * \tau_{i,t} * s_i + c_{10,t}X_{d,t} + c_{11,t}X_{d,t} * s_i + c_{12,t}X_{d,t} * \tau_{i,t} \\
& & + & c_{13,t}X_{d,t} * s_i * \tau_{i,t} + c_{14,t}\overline{\tau r}_t * n_d + c_{15,t}\overline{\tau r}_t * n_d * s_i + c_{16,t}\overline{\tau r}_t * n_d * \tau_{i,t} \\
& & + & c_{17,t}\overline{\tau r}_t * n_d * s_i * \tau_{i,t} + c_{18,t}\overline{\tau}_t * n_d + c_{19,t}\overline{\tau}_t * n_d * s_i \\
& & + & c_{20,t}\overline{\tau}_t * n_d * \tau_{i,t} + c_{21,t}\overline{\tau}_t * n_d * s_i * \tau_{i,t} + D_d + D_t + \varepsilon_{i,d,t}
\end{array}
$$

$c_{10,t}X_{d,t}$,$c_{11,t}X_{d,t}*s_i$,$c_{12,t}X_{d,t}*\tau_{i,t}$ and $c_{13,t}X_{d,t}*s_i*\tau_{i,t}$ capture third-order effects of employment responses to changes in aggregate employment driven by changes in average tariffs vary with $X_{d,t}$ and interactions of $X_{d,t}$ with industry level labor usage and tariffs. Since these third-order effects are likely to be small, I omit these variables from the analysis. A similar case can be made for excluding $c_{16,t}\overline{\tau r}_t * n_d * \tau_{i,t}$, $c_{17,t}\overline{\tau r}_t * n_d * s_i * \tau_{i,t}$, $c_{20,t}\overline{\tau}_t * n_d * \tau_{i,t}$, $c_{21,t}\overline{\tau}_t * n_d * s_i * \tau_{i,t}$. In addition, under the assumption that $\beta_8 - \beta_{10} = 0$, $c_{13,t}$ through $c_{19,t}$ don't vary over time.[70] This leaves:

$$
\begin{array}{lrcl}
(39) & l_{i,d,t} & = & c_0 + c_1\tau_{i,t} + c_2\tau_{i,t} * r_i + c_3\tau_{i,t} * r_i * n_d + c_4\tau_{i,t} * n_d + c_5 r_i * n_d + c_6X_{d,t} \\
& & + & c_7X_{d,t} * s_i + c_8X_{d,t} * \tau_{i,t} + c_9X_{d,t} * \tau_{i,t} * s_i + c_{14}\overline{\tau r}_t * n_d + c_{15}\overline{\tau r}_t * n_d * s_i + \\
& & + & c_{18}\overline{\tau}_t * n_d + c_{19}\overline{\tau}_t * n_d * s_i + D_d + D_t + \varepsilon_{i,d,t}
\end{array}
$$

---

[67]where $b_0 = (\tilde{\beta}_0+\alpha_0(\beta_6+\beta_7\overline{s}))$, $b_1 = \beta_3$, $b_2 = \beta_4$, $b_{3,t} = \alpha_2(\beta_6-\beta_7\overline{s}-\beta_8\overline{\tau}_t-\beta_9\overline{\tau s}_t)$, $b_{4,t} = (\beta_{11}+\alpha_1(\beta_6+\beta_7\overline{s}+\beta_8\overline{\tau}_t+\beta_9\overline{\tau s}_t))$, $D_{d,t} = (\delta_d + \alpha_0(\beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t) + d_d(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t)) + (\delta_t + d_t(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t))$ $\xi_{d,t} = (u_{i,d,t} + v_{d,t}(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t))$

[68]$a_0 = (\tilde{\beta}_0 + \alpha_0(\beta_6 + \beta_7 s_i))$, $a_1 = \beta_1 + \beta_8\alpha_0$, $a_2 = \beta_2$ $a_3 = \beta_3$, $a_4 = \beta_4$ $a_5 = \beta_5$ $a_6 = \beta_{11} + \alpha_1\beta_6$, $a_7 = \alpha_1\beta_7$, $a_8 = \alpha_1\beta_8$, $a_9 = \alpha_1\beta_9$, $a_{10} = \alpha_2\beta_6$, $a_{11} = \alpha_2\beta_7$, $a_{12} = \alpha_2\beta_8$, $a_{13} = \alpha_2\beta_9$, $a_{14} = \beta_{10}$ $\epsilon_{i,d,t} = u_{i,d,t} + v_{d,t}(\beta_6 + \beta_7 * s_i + \beta_8\tau_{i,t} + \beta_9\tau_{i,t} * s_i)$

[69]$c_0 = (\tilde{\beta}_0 + \alpha_0(\beta_6 + \beta_7 s_i)) + b_0 a_{10}$, $c_1 = \beta_1 + \beta_8\alpha_0$, $c_2 = \beta_2 + b_0 a_{12}$ $c_3 = \beta_3$, $c_4 = \beta_4$ $c_5 = \beta_5$, $c_6 = \beta_{11} + \alpha_1\beta_6$, $a_7 = \alpha_1\beta_7$, $a_8 = \alpha_1\beta_8$, $a_9 = \alpha_1\beta_9$, $a_{10} = \alpha_2\beta_6$, $c_{10,t} = a_{10}\frac{1}{b_{3,t}}(\beta_{11} + \alpha_1(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t))$, $c_{11,t} = a_{11}\frac{1}{b_{3,t}}(\beta_{11} + \alpha_1(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t))$, $c_{12,t} = a_{12}\frac{1}{b_{3,t}}(\beta_{11} + \alpha_1(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t))$, $c_{13,t} = a_{13}\frac{1}{b_{3,t}}(\beta_{11} + \alpha_1(\beta_6 + \beta_7\overline{s} + \beta_8\overline{\tau}_t + \beta_9\overline{\tau s}_t))$ $c_{14,t} = a_{10} * \frac{1}{b_{3,t}} * \beta_3$, $c_{15,t} = a_{11} = a_{11} * \frac{1}{b_{3,t}} * \beta_3$ $c_{16,t} = a_{12} * \frac{1}{b_{3,t}} * \beta_3$ $c_{17,t} = a_{13} * \frac{1}{b_{3,t}} * \beta_3$ $c_{18,t} = a_{10} * \frac{1}{b_{3,t}} * \beta_3$, $c_{19,t} = a_{11} = a_{11} * \frac{1}{b_{3,t}} * \beta_3$ $c_{20,t} = a_{12} * \frac{1}{b_{3,t}} * \beta_3$ $c_{21,t} = a_{13} * \frac{1}{b_{3,t}} * \beta_3$

[70]Including the interactions of these variables with time trends has little impact on the coefficient estimates in the second stage. These results are therefore not included in this robustness checks but are available upon request.

C.4. **Levels of Aggregation.** Let the outcome regression be given by:

$$y_d = \beta_0 + \beta_1 x_{1,d} + ... + \beta_{k-1} x_{k-1,d} + \beta_k x_{k,d} + u_d \tag{40}$$

$$E[u_d] = 0 \tag{41}$$

where $x_{1,d}$ through $x_{k-1,d}$ are exogenous variables and $x_{k,d}$ denotes the endogenous variable. The outcome variable of interest, y_d, is measured at a district level and varies over time. Let each district d belong to a region r, where there are a total of D districts and R districts, where each region has an identical number of districts, R_d. Write the regression as:

$$y = \mathbf{x}\beta + u \tag{42}$$

where $\mathbf{x^d} = (1, x_{1,d}, ..., x_{k-1,d}, x_{k,d})$ is a $1 \times (K+1)$ vector of variables at a district level.

Assume that we have more than one instrumental variable for x_K. Let $z_{1,d}, z_{2,d}, ...z_{M,d}$ denote the variables such that:

$$Cov(z_{h,d}, u_d) = 0 \quad h = 1, 2, ..., M \tag{43}$$

so that each $z_{h,d}$ is exogenous. Let $z_{1,r} = \sum_{d \in r}^{R_d} z_{1,d}$ denote

Each z_h is exogenous in equation under the assumption that $Cov(z_{d',h}, u_d) = 0$, $\forall d' \neq d \in R_d$

$$Cov(z_{h,r}, u) = Cov(\sum_{d}^{R_d} z_{h,d}, u) = \sum_{d}^{R_d} Cov(z_{h,d}, u) = 0 \tag{44}$$

Define the vector of exogenous variables at a district level as: $\mathbf{z^d} = (1, x_{1,d}, ..., x_{k-1,d}, z_{1,d}, z_{2,d}, ...z_{M,d})$, a $1 \times L$ vector $(L = K + M)$. Similarly, define the vector of exogenous variables at a region level as: $\mathbf{z^r} = (1, x_{1,r}, ..., x_{k-1,r}, z_{1,r}, z_{2,r}, ...z_{M,r})$, a $1 \times L$ vector $(L = K + M)$.

The population moment conditions of $E[u_d] = 0$ and $Cov(z_{h,d}, u_d)$ imply the L population orthogonality conditions:

$$\text{Assumption 2SLS.1 district} \quad E[\mathbf{z^d}' u] = \mathbf{0} \tag{45}$$

$$\text{Assumption 2SLS.1 region} \quad E[\mathbf{z^r}' u] = \mathbf{0} \tag{46}$$

The rank conditions at a district level are given by:

$$\text{Assumption 2SLS.2a district} \quad rank E(\mathbf{z^d}' \mathbf{z^d}) = L \tag{47}$$

$$\text{Assumption 2SLS.2b district} \quad rank E(\mathbf{z^d}' \mathbf{x}) = K \tag{48}$$

While the analogue conditions at a region level are given by:

$$\text{Assumption 2SLS.2a region} \quad rank E(\mathbf{z^r}' \mathbf{z^r}) = L \tag{49}$$

$$\text{Assumption 2SLS.2b region} \quad rank E(\mathbf{z^r}' \mathbf{x}) = K \tag{50}$$

Under $2SLS.1$, $2SLS.2a$ and $2SLS.2b$, β is identified and the 2SLS estimator obtained from a random sample is consistent for β.

References

[1] Aghion, Philippe; Burgess, Robin; Redding, Stephen, and Zilibotti, Fabrizio, (2009) "The Unequal of Liberalization: Evidence from Dismantling the License Raj in India", *The American Economic Review*

[2] Ahluwalia, M.S. (2002) "State-Level Performance under Economic Reforms in India", in Anne O. Krueger (ED.) Economic Policy Reforms and the Indian Economy, New Delhi: Oxford University Press.

[3] Ahluwalia, M.S. (1975) "Income Distribution and Development: Some Stylized Facts", *The American Economic Review*

[4] Atkin, David (2009)

[5] Badiani, Reena (2009) "Location and Growth of Employment in the Manufacturing Sector in India, 1983-2001" *Work in Progress*

[6] Besley, Timothy and Robin Burgess, (2004) "Can Labour Regulation Hinder Economic Performance? Evidence from India," *The Quarterly Journal of Economics*, MIT Press, vol. 119(1), pages 91-134, February.

[7] Datt, Gaurav and Martin Ravallion (2002) "Growth Poverty in Rural India" World Bank Policy Research Working Paper No. 2846.

[8] Edmonds, Eric, Nina Pavnick and Petia Topalova

[9] Ellison, Glenn and Edward L. Glaeser (1999) "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?," *American Economic Review*, American Economic Association, vol. 89(2), pages 311-316, May.

[10] Foster, Andreas and Mark Rosenzweig (2003) "Information, Learning and Wage Rates in Low-Income Rural Areas" *Journal of Human Resources*, Fall 1993, 28(4) pp759-790

[11] Foster, Andrew D and Rosenzweig, Mark R (1996) "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review*, American Economic Association, vol. 86(4), pages 931-53, September.

[12] Goldberg, Penny, Amit Khandewal, Nina Pavnick and Petia Topalova

[13] Government of India, 2001, Concepts and Definitions Used in NSS: Golden Jubilee Publication. Government of India, Ministry of Statistics and Program Implementation, National Sample Survey Organisation.

[14] Griliches, Zvi, *Technology, Education and Productivity: Essays in Applied Econometrics*, New York: Basil Blackwell

[15] Kim, Sukkoo (1995), "Expansion of Markets and the Geographic Distribution of Economic Activity: The trends in U.S. Regional Manufacturing Structure, 1860-1987" *Quarterly Journal of Economics* 110, 881-908

[16] Kochhar, Kalpana, Utsav Kumar, Raghuram Rajan, Arvind Subramanian and Ioannis Tokatlidis (2006), "Indias Pattern of Development: What Happened, What Follows?" IMF Working Paper, WP/06/22

[17] Krueger, Anne, and Sajjid Chiony (2002), "The Indian Economy in Global Context" in Anne O. Krueger (ED.) Economic Policy Reforms and the Indian Economy, New Delhi: Oxford University Press.

[18] Krugman, Paul, (1991), *Geography and Trade*, Cambridge, Mass: MIT Press.

[19] Lanjouw, Peter and Rinku Murgai (2008), "Poverty Decline, Agricultural Wages and Non-Farm employment" Working Paper, World Bank.

[20] Mehta, R. (1999): Tariff and Non-tariff Barriers of Indian Economy: A Profile, Research and Information System for the non-aligned and Other Developing Countries (RIS), New Delhi.

[21] Munshi, Kaivan and Mark Rosenzweig (2005), "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth," BREAD Working Paper No. 092.

[22] NCAER Publication, (1969) "Techno-Economic Survey of Andhra Pradesh" and other states; National Council of Applied Economic Research: New Delhi

[23] Srinivasan, T.N. (2000), "Eight Lectures of India's Economic Reforms", Oxford India Paperbacks, New Delhi: Oxford University Press.

[24] Srinivasan, T.N. (2002), "The Indian Economy in Global Context" in Anne O. Krueger (ED.) Economic Policy Reforms and the Indian Economy, New Delhi: Oxford University Press.

[25] Topalova, Petia (2005) "Trade Liberalization, Poverty, and Inequality: Evidence from Indian Districts," NBER Working Papers 11614, National Bureau of Economic Research, Inc.

[26] Wooldridge, Jeffrey (2002) *Econometric Analysis of Cross Section and Panel Data* MIT Economic Press.