Agricultural productivity and the sectoral reallocation of labor in rural India

Kyle Emerick[1]

August 24, 2016

Abstract

How do shocks to agricultural productivity affect the allocation of labor across sectors of the economy? To answer this, I use data from rural India to show that exogenous increases in agricultural productivity — caused by abnormally high levels of precipitation — lead to an increase in the labor share of the non-agricultural sector. I further show that the non-tradable sector expands significantly when agricultural output increases. This evidence is consistent with increasing agricultural output causing increased demand for local non-tradables, which in turn increases the non-agricultural labor share.

JEL codes: Q10, O13, J43
Keywords: Agricultural productivity, labor allocation

[1]Department of Economics Tufts University, 8 Upper Campus Rd, Medford, MA 02155. kyle.emerick@tufts.edu. I thank Raasika Gaugler and Xinxin Lyu for excellent research assistance.

1 Introduction

Many developing countries devote a substantial share of labor to agriculture. Yet, cross-country evidence indicates strongly that in most developing countries the share of agriculture in total value added falls far below the agricultural labor share (Restuccia, Yang, and Zhu, 2008; Duarte and Restuccia, 2010; Lagakos and Waugh, 2013; Gollin, Lagakos, and Waugh, 2014). This apparent "agricultural productivity gap" has implications for economy-wide total factor productivity and thus highlights the importance of labor reallocation in the development process (Vollrath, 2009; McMillan, Rodrik, and Verduzco-Gallo, 2014). Yet, there is little evidence on whether increases in agricultural productivity will increase the amount of labor allocated to the non-agricultural sector. On the one hand, productivity gains in agriculture could lead to an immediate increase in agricultural labor demand, thus crowding out non-agricultural employment. On the other hand, productivity gains could release labor that was otherwise necessary to produce enough food to meet subsistence requirements (Gollin, Parente, and Rogerson, 2002, 2007), or they could generate additional demand for locally produced non-tradables, which would additionally cause more labor to flow to the non-agricultural sector (Foster and Rosenzweig, 2004, 2007).

In this paper I show that exogenous increases in agricultural output cause modest increases in the non-agricultural labor share in rural India. Methodologically, I use variation in precipitation during the main growing season — an important driver of agricultural output — to isolate exogenous variation in agricultural productivity. In doing this I overcome the omitted variables bias issues that plague cross-sectional comparisons between high and low productivity areas.[1] However, an important caveat is that I use annual fluctuations in precipitation and the results therefore are not descriptive of long-term structural transformation of the economy.

I have four main results. First, using data from Indian districts, I show that rural households are significantly more likely to be employed in the non-agricultural sector during years when rainfall is abundant. In particular, a household is approximately 1.3 percentage points (2.8%) more likely to be a non-agricultural household during a year when rainfall during the main growing season is one standard deviation above the local average. Rescaling this estimate by the first stage effect of rainfall on agricultural output, the instrumental variables estimate indicate that a 10% increase in agricultural output causes the agricultural labor share to decline by 1.18 percentage points, or 2.2%. I further show that the declining

[1] The use of annual fluctuations in weather as an instrument for agricultural income is fairly common (Miguel, Satyanath, and Sergenti, 2004; Jayachandran, 2006). The approach relies on precipitation being unrelated to outcomes other than the effect that occurs via shocks to agricultural output. I consider this assumption extensively in Section 4.

agricultural labor share appears to result from an actual reallocation of workers and not growth in total employment. The data indicate that during good rainfall years households are likely to shift from agriculture being their primary to their secondary activity. Yet, the total amount of days worked in the non-agricultural sector increases, indicating that the effect on the primary sector of employment dominates.

Second, I analyze which sectors are most likely to grow when agricultural output increases. Most importantly, I show that local non-tradable sectors expand significantly during abnormally wet years. In particular, positive agricultural productivity shocks cause a significant increase in the labor share of the non tradable construction sector — mostly in residential construction. Further, I also show that the retail and education — mostly primary and secondary schooling — sectors grow when agricultural is more productive.[2] By showing that non-tradables account for a major share of the expansion after an increase in agricultural productivity, I am able to shed light on local demand effects as an important mechanism that explains the results. As is shown in Foster and Rosenzweig (2004), the local non-tradable sector in rural areas expands during agricultural booms because income gains from agricultural increase the demand for local goods and services. My results are consistent with this explanation. In addition, the result is similar to the recent experience in Africa where much of the labor that has left the agricultural sector has been absorbed by the local service sector (McMillan and Harttgen, 2014).

Third, while the overall effect of agricultural output on the non-agricultural labor share is seemingly modest, I use a simple decomposition to show that the gap in output per worker between the agricultural and non-agricultural sectors makes even a modest reallocation of workers economically significant. I use district-level GDP estimates from the Indian Planning Commission along with district-level labor shares from a nationally representative survey of workers to estimate output per worker by sector. With these estimates in hand, I am able to decompose the effects of an increase in agricultural output on GDP per capita into two parts: a direct effect due to higher output in the agricultural sector and an indirect effect that results from additional labor moving to the more productive non-agricultural sector. The results indicate that for an average Indian district, a 10 percent increase in agricultural output leads to an increase in GDP per capita of 477 rupees (2.5 percent) *while holding the sectoral allocation of labor constant*. However, the indirect effect that results from sectoral labor reallocation amounts to 211 rupees, or an additional 1.1 percent. Put differently, around 31 percent of the overall gains from the productivity shock are due to movement of labor to the more productive non-agricultural sector. The finding stresses the importance of

[2] One plausible explanation of the growth in the education sector is the positive relationship between demand for education and productivity in agriculture (Foster and Rosenzweig, 1996).

the indirect effects of agricultural productivity on welfare (de Janvry and Sadoulet, 2009).

Fourth, I show that the effect of agricultural productivity on sectoral labor reallocation is most prominent amongst households that belong to higher castes and individuals that have at least a primary education. This heterogeneity suggests that although labor reallocation may amplify the gains from positive agricultural productivity shocks, these additional gains are unlikely to reduce inequality. Instead, the better-off households are most likely to benefit from the growth in the rural non-farm sector.

Turning to robustness, the most obvious concern with my identification strategy is that year-to-year fluctuations in weather could potentially have an impact on the non-agricultural sector through channels other than agricultural productivity. As examples, labor demand in the non-agricultural sector could depend directly on rainfall if work takes place outdoors and is inhibited by rain. Or, If heavy rainfall affects the ability to transport complementary inputs, then the labor demand of the non-agricultural sector could respond directly to rainfall.[3]

In addition to only using precipitation during the wet season and conditioning on precipitation during the other months of the year, I show results from two falsification tests that are inconsistent with this potential violation of the exclusion restriction. First, I show that the sectors most likely to gain workers in rural areas during wet years are no more likely to grow during wet years in urban areas. We would expect to see effects for urban households if labor demand in the construction, retail, and education sectors depended directly on rainfall. Second, I exploit the spatial distribution of groundwater access — India's most popular irrigation source that reduces the susceptibility of agriculture to rainfall. I split my sample into districts that are above and below the sample median in terms of the share of cultivated area that is equipped with groundwater irrigation. The first-stage effect of rainfall on agricultural output exists only in districts with below-median access to groundwater. At the same time, the reduced-form effect of rainfall on the agricultural labor share exists only in these same districts. This lack of both first-stage and reduced-form effects in highly irrigated districts suggests that the effect of growing-season rainfall on the agricultural labor share operates through agricultural productivity.

These findings fill a gap left in the current literature on structural transformation and agricultural productivity, particularly in India. Kochar (1999) uses data from three Indian villages during the period from 1975 to 1984 to show that *both* positive and negative shocks

[3] It is noteworthy that both of these explanations would work against my finding that the non-agricultural labor share grows during wet years. However, there are likely other violations of the exclusion restriction that would induce the results that I observe. One example is that wet years reduce the marginal value of leisure, much of which takes place outside in rural India. This would induce a direct increase in the supply of labor, some of which would flow to the non-agricultural sector.

to agricultural income cause rural households to supply more off-farm labor. In a related paper, Adhvaryu, Chari, and Sharma (2013) find that industrial employment (for large manufacturing firms) grows during high-rainfall years, particularly in states that have laws that favor employers in terms of hiring and firing employees. Colmer (2016) shows that the manufacturing and services sectors absorb workers during hot years when agricultural productivity suffers.

In contrast to previous work, I combine data from the country's main agricultural area, employment data from all sectors, and several different measurements of local weather shocks to build a more comprehensive understanding of how the rural economy responds to local variation in agricultural productivity. Most importantly, I attempt to further quantity the economic significance of these labor reallocation effects by understanding how much of the local economic gains from positive agricultural shocks can be attributed to the reallocation of workers across sectors. Quantifying the magnitude of these local demand effects is important because a growing literature highlights the large gaps in labor productivity between the agricultural and non-agricultural sectors (Restuccia, Yang, and Zhu, 2008; Duarte and Restuccia, 2010; Gollin, Lagakos, and Waugh, 2014). These gaps in labor productivity across sectors when combined with my results suggest that labor reallocation constitutes an important component of the welfare gains from short-term agricultural productivity shocks in rural India. I quantify this importance more precisely by showing that gains from labor reallocation account for just less than one third of the overall economic gains from a short-term increase in agricultural productivity.

In focusing on the cross-sectoral effects of agricultural productivity shocks, the paper also relates to the literature on general equilibrium impacts of sector-specific productivity shocks. Focusing on the rural United States, Hornbeck and Keskin (2015) show that large gains in agricultural productivity due to groundwater irrigation resulted in only temporary gains to other sectors in the central United States. Bustos, Caprettini, and Ponticelli (2016) show that genetically engineered soybeans in Brazil save labor and thus their adoption causes the non-agricultural sector to grow. Finally, Allcott and Keniston (2014) show that positive shocks to the oil and gas sector in rural counties of the United States drive up wages but nonetheless lead to growth in the local manufacturing sector. I find that productivity of the agricultural sector in rural India has an economically meaningful impact on the size of the local nonagricultural sector.

The rest of this paper is organized as follows. In Section 2 I present a qualitative discussion of the major channels that could explain how shocks to agricultural productivity affect labor in other sectors. Section 3 discusses the data and identification strategy while Section 4 presents results. Section 5 presents various robustness checks and Section 6 discusses the implications of the findings and concludes.

2 Channels from agricultural productivity to labor allocation

In this section I provide a qualitative discussion of the economic mechanisms that could potentially explain how a shock in agricultural productivity affects the sectoral allocation of labor. I focus the discussion on general equilibrium effects in a two-sector economy. I start with the two competing mechanisms that guide the empirical analysis: the labor pull and local demand effects. I then discuss liquidity constraints and sectoral linkages — two alternate channels that I briefly consider in the empirical analysis.

Labor pull

Two-sector models such as the classic one in Harris and Todaro (1970) deliver the prediction that increases in agricultural productivity will pull labor away from the non-agricultural sector. To see this, note that a positive shock in agricultural productivity increases the marginal productivity of labor in agriculture, thus driving up wages and attracting labor away from the non-agricultural sector. If the labor endowment in the economy is fixed and perfectly mobile across sectors, then the effect of the gain in agricultural productivity is to reduce the amount of labor allocated to the non-agricultural sector. Taken together, higher wages and increased labor demand in agriculture combine to cause contraction of the non-agricultural labor force.

Local demand effects

Allowing for general equilibrium effects introduces a competing channel that works in the opposite direction as the labor pull effect. Rising incomes from higher agricultural productivity can lead to increased demand for local goods and services, which in turn draws labor away from agriculture and towards these sectors.[4]

The online appendix includes a simple general equilibrium model that clarifies two conditions that are most likely to make the local demand effect dominate. First, strong complementarity in consumption between agricultural and non-agricultural goods increases the demand effect of rising agricultural productivity (Foster and Rosenzweig, 2007). Second, an

[4] An important distinctive feature of this mechanism is the tradability of the good produced in the non-agricultural sector (Foster and Rosenzweig, 2004). Local price effects are absent for non-tradables.

income elasticity greater than one for the non-agricultural good — and less than one for the agricultural good — makes the local demand effect larger. The theoretical literature on structural transformation has commonly used Stone Geary preferences to achieve this latter condition (Matsuyama, 1992; Kongsamut, Rebelo, and Xie, 2001; Gollin, Parente, and Rogerson, 2002; Herrendorf, Rogerson, and Valentinyi, 2013; Gollin and Rogerson, 2014).

As I show in the online appendix, the relative price of the non-farm good needs to rise faster than agricultural TFP in order for the local demand effect to dominate. Alvarez-Cuadrado and Poschke (2011) show that across several developed countries the relative price in the manufacturing sector has increased noticeably since around 1960. At the same time, the employment share in agriculture has declined and agricultural TFP has risen relative to manufacturing TFP. These observations are broadly consistent with a model where rising agricultural productivity drives up prices and causes growth in the non-farm sector.

Liquidity constraints

Many general equilibrium models of sectoral labor allocation make the basic assumption that labor can move freely between sectors. However, in many cases moving sectors may require large upfront costs associated with migration. In the simplest case, non-farm activity is carried out in urban areas and thus changing sectors necessitates rural-to-urban migration. Recent empirical literature has indeed found evidence that the up-front costs of migration are an important barrier that prevents households from leaving rural areas (Bryan, Chowdhury, and Mobarak, 2014; Kleemans, 2014; Bryan and Morten, 2015; Angelucci, 2015; Bazzi, 2016).

If sectoral labor reallocation can be viewed synonymously with migration, then a positive shock in agricultural productivity could relieve liquidity constraints, lead to an increase in rural-urban migration, and thus generate an increase in the non-agricultural labor share. This line of reasoning suggests one reason why agricultural productivity growth would be "pro-poor". That is, the poorest households are most likely to face the most binding liquidity constraints. An exogenous increase in agricultural productivity that relieves these constraints would then be most likely to benefit the poorest households by increasing their propensity to work in the non-agricultural sector.

Sectoral linkages

Declining factor prices could explain sectoral labor reallocation if agricultural output is an input into non-farm production. In the simplest case, one can consider the food processing and manufacturing sector, such as a rice mill. If the price of agricultural output falls due to a positive shock in productivity, then complementarity between the input (agricultural output) and labor would lead to an increase in labor demand in the sector. Thus, linkages between sectors combined with falling agricultural prices could explain why labor flows to other sectors in response to an increase in agricultural productivity.

3 Data and Empirical Strategy

I combine data on agricultural output, household employment and wages, and annual weather to create a pooled cross-section of individual households over time. In this section I describe the most important details of the data used.

I measure agricultural output at the district-year level using the Indian Minsitry of Agriculture's annual crop production statistics. These data include the output and area of each crop for all districts from 1998-2013. The Indian agricultural year is split into two distinct seasons : a wet season (or "kharif" season) that runs from June to November and a shorter dry season (or "rabi" season) that runs from December to May. I focus on the wet season since dry-season crops are almost always irrigated and thus less susceptible to fluctuations in rainfall. Further, I measure agricultural productivity as the value of the output produced from the six main wet-season crops: rice, corn, soybeans, groundnuts, pearl millet, and sugarcane.[5] Annual prices for the entire country are used so that the value of output from the different crops can be added to create an overall measure of productivity.[6]

I use India's National Sample Survey (henceforth "NSS") to measure employment and wages. NSS is a widely-used nationally representative household survey that measures both employment outcomes and detailed household consumption. I use rounds 55, 61, 64, and 66 of "Schedule 10", or the employment/unemployment module of the survey. The surveys were administered beginning in July of 1999, 2004, 2007, and 2009, respectively.[7] The most relevant variables include wages and the industry of the principal occupation of all working household members and the household overall. Based on the time allocation of individual household members, NSS classifies each household in rural areas as self-employed in agriculture if more than 50% of its income comes from agriculture. Similarly, households are classified as agricultural laborer households if 50% of income comes from agricultural

[5]Taken together, these crops account for around 70% of the total area that is cultivated in the country during the wet season. Rice is by far the dominant crop, accounting for around 40% of annual wet season area on it's own.

[6]I use country-level prices from the FAOSTAT database so that the measure of productivity does not include local price effects.

[7]I use only the rounds from 1999 onwards to match the timing of the annual crop production statistics. Similarly to Deaton and Dreze (2002) and Topalova (2010), I also use only the "thick rounds" of the NSS survey due to their larger sample sizes. The thin rounds have smaller sample sizes and are not representative at the district level.

labor. I define households as working in the agricultural sector if they are either self employed in agriculture or agricultural laborer households.[8]

The surveys for each NSS round were carried out from July 1st to June 30th. The wet season harvest takes place in November or December of each year. Therefore, the season before the survey is the one where the harvest outcome had been realized at the time when surveying began. I define the precipitation shock variable accordingly by using the growing season immediately before the start of surveys. In other words, I use wet-season rainfall (June to November) from 1998, 2003, 2006, and 2008 for rounds 55, 61, 64, and 66, respectively.

The principal occupation of the household is determined based on the allocation of labor and income earned during the 365 days preceding the survey. This long reference window effectively means that the precipitation shock will be realized during or just before the reference window — making it reasonable to define the timing of the precipitation shock in such a way. A different approach would be to use the timing of the surveys to calculate an average of the two shocks that are relevant for each household. This approach to measuring the timing of shocks yields a similar main result to that reported in the paper.[9]

I primarily use the University of Delaware's terrestrial precipitation and temperature gridded monthly time series datasets to measure district-level weather. These gridded data consist of monthly observations on temperature and precipitation for a 0.5 degree grid for the period from 1950-2008 (Willmott and Matsuura, 2009).[10] I overlay the GIS boundaries of each district with the gridded climate data and calculate monthly precipitation and temperature as the weighted averages across all grids that intersect with the district, where the weights correspond to the share of district area that falls in each grid.

Finally, I limit the sample to the main agricultural states by excluding states in the far northwest and northeastern parts of the country.[11] In addition, I remove the less than 10%

[8]Unless otherwise noted, I show specifications using the household as the unit of observation throughout the paper. However, results are very similar when using individual-level observations and an indicator for working in the agricultural sector that is defined based on the 5 digit industry code of the individual's principal occupation.

[9]Households in the NSS are surveyed in four sub-rounds (July-September, October-December, January-March, and April-June). The previous wet-season harvest therefore occurred in the middle of the reference window for households in the first two sub-rounds. This alternative approach averages the rainfall shock measure from that previous season and the season before for these households. As an example, a household surveyed on August 15th 1999 will be asked about labor allocation from August 15th 1998 to August 15th 1999. Since the harvest occurs in November/December, I take the average of the 1998 and 1997 precipitation shock variables. Put simply, the 1997 wet season was the most recent wet season for the part of the reference window up until the 1998 harvest was realized. Another example demonstrates this modified measurement technique for households surveyed in sub-rounds 3 or 4. Consider a household surveyed on March 15th, 2000. The reference window for this household is March 15th 1999 to March 15th 2000. The 1998 harvest was the most recent wet-season harvest up until November/December 1999 when the 1999 harvest took place. I therefore take the average precipitation shocks from 1998 and 1999 for such a household.

[10]0.5 decimal degrees corresponds to approximately 50 kilometers when measured at the equator.

[11]The states I include are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil

of districts from the sample that had boundary changes during the period from 1998-2008. Finally, I show results primarily for the rural sample of NSS.

Table 1 displays summary statistics from rural households in the NSS and for weather variables from the UDEL dataset. Focusing first on the allocation of labor across sectors, the share of the labor force working in agriculture fell from 75% to 66% from 1998 to 2008. The construction sector is the one sector that absorbed much of the labor that left agriculture: while only 3% of the rural individuals worked in the construction sector in 1998, that number rose to 10% by 2008. Turning to the weather variables, the first and last years of the sample were above-average in terms of wet-season precipitation, while both 2003 and 2006 were relatively drier years.

I use district-level GDP estimates prepared by the Indian Planning Commission as a final source of data.[12] These data include sector-specific GDP estimates for each fiscal year from 1999-2007. Figure 1 shows the correlation between the share of GDP from agriculture and the share of the labor force that works primarily in agriculture, measured using the 61st NSS round in 2004. Across all districts 53% of the population works in agriculture, while only around 28% of GDP comes from agriculture. This is noticeable in the figure as most districts fall below the 45 degree line where the agricultural employment share is exactly equal to it's share of GDP. This evidence is suggestive that labor is misallocated across sectors even within districts.

Building on this, my empirical specification seeks to estimate the casual effect of agricultural productivity on the sectoral allocation of labor. Denoting a_{idt} as an indicator for whether household i in district d works primarily in the agricultural sector during year t, the equation of interest can be written as,

$$a_{idt} = \beta_1 product_{dt} + \alpha_d + \eta_t + \varepsilon_{idt}, \tag{1}$$

where $product_{dt}$ is the log of agricultural output in monetary terms in year t, α_d are district fixed effects, and η_t are time effects. Estimating β_1 by OLS is problematic because agricultural output and labor use are simultaneously determined. Also, unobserved local economic conditions are likely correlated with both agricultural output and labor use.

I use precipitation in the district, $precip_{dt}$, where $precip_{dt}$ is measured as the difference between total precipitation from June to November and the average precipitation in the district during the same period from 1950-2008, normalized by the district-specific standard

Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa. Together these states account for approximately 93% of agricultural area in the country.

[12] The data are available at http://planningcommission.nic.in/plans/stateplan/index.php?state=ssphdbody.htm (downloaded February, 2015).

deviation. Thus, the instrument is a measure of the distance in standard deviations between the current growing season's precipitation and the long-run average in the district. Importantly, I only use growing-season precipitation anomalies and also condition on precipitation during the previous dry season (December-May) in all specifications. I rely crucially on the assumption that conditional on district and time fixed effects, agricultural productivity is the only channel through which annual fluctuations *during the growing season* affect the share of the workforce in agriculture.

The corresponding first-stage relationship is therefore

$$product_{dt} = \delta_1 precip_{dt} + \alpha_d + \eta_t + \varepsilon_{dt} \quad (2)$$

My identification strategy obviously requires that precipitation during the growing season is a strong predictor of agricultural output. I first show that this is the case using visual evidence in the top panel of Figure 2. The figure shows the non-parametric relationship between the rainfall deviation and agricultural output, conditional on district and time fixed effects and dry-season precipitation. Reassuringly, there is a clear positive relationship between the rainfall shock measure and agricultural output. Going from the 5th to 95th percentiles of the rainfall shock — or from around 1 standard deviation below to 1.2 standard deviations above average rainfall — leads to an increase in the value of agricultural output of around 25 log points. This evidence is in line with other studies that have exploited rainfall during the monsoon season in India as an instrument for agricultural output (Jayachandran, 2006).

4 Results

I next show my main results on the effect of agricultural productivity on the equilibrium labor allocation to agriculture. I then show additional supplementary results on which sectors absorb the labor that leaves agriculture, effects on local wages, and simulation results to decompose the effect of agricultural productivity on local economic output into a direct production effect and a labor-reallocation effect.

4.1 Effect of agricultural productivity on the agricultural labor share

The regression results in Panel A of Table 2 confirm the visual evidence in Figure 2. Above-average rainfall during the wet season leads to large gains in agricultural productivity. More specifically, the total revenue from the six main wet-season crops increases by approximately 10.8% when total precipitation is one standard deviation above normal. The effect is highly significant (first-stage F=51.12).[13] As Column 2 shows, this first-stage relation is robust to controlling flexibly for region-specific time-trending unobservables by introducing region-by-year fixed effects.[14] Overall, these results are in no way surprising given the reliance of Indian agriculture on monsoon rainfall.

The reduced-form relation in Panel B shows a strong negative link between abnormally wet years and the amount of labor that works in agriculture. A one standard deviation increase in growing-season rainfall leads to a decrease in the agricultural labor share by 1.3 percentage points (the middle panel of Figure 2 shows the corresponding non-parametric relationship). This represents an approximate 2.4% decline in the share of labor allocated to agriculture.[15] If anything, the reduced-form effects become stronger when controlling for region-by-year fixed effects. In addition, the reduced-form effect remains significant when adjusting standard errors for spatial correlation across districts.[16]

Combining these reduced-form and first-stage results, Panel C shows the instrumental variables estimates. The negative and statistically significant effect of agricultural output on the agricultural labor share indicates that an increase in agricultural output by 10% leads to an approximate decrease in the agricultural labor share by 1.1 percentage points. A simple example is useful to help interpret the magnitude of the effect. One of the most severe droughts in recent history in India occurred in 2002: agricultural output fell by approximately 28 log points relative to the previous year. The coefficient estimate indicates that the productivity shock caused by the 2002 drought led to an increase in the agricultural labor share of 3.05 percentage points.

These effects of agricultural productivity on the reallocation of labor appear to be short term and not persistent over time. I show in Figure A2 results from a distributed lag model where the indicator for an agricultural household is regressed on the growing-season precipitation during the most recent season as well as lags for the six previous growing seasons. The

[13] I use the district as the unit of observation in Panel A since I only observe variation in the instrument at the district level. This explains the large difference in the number of observations across the rows of the table.

[14] The four regions are East (Bihar, Jharkhand, Orissa, West Bengal, Assam, and Chhattisgarh), West (Gujarat, Maharashtra, and Goa), North (Punjab, Haryana, Rajasthan, Uttar Pradesh, and Madhya Pradesh), and South (Tamil Nadu, Karnataka, Andhra Pradesh, and Kerala).

[15] The unweighted agricultural labor share of 55% from Table 2 is slightly lower that the weighted agricultural labor shares. While I do not apply the sampling weights in the regression analysis, I show in Table A8 that the main reduced-form results are unaffected when applying NSS sampling weights.

[16] Figure A1 shows 95 percent confidence bounds when standard errors allow for both spatial and temporal dependence (Hsiang, 2010). I vary the threshold for which two districts are assumed to be independent from 50 to 1,000 km, in increments of 10. Errors for a given district are assumed to be correlated across time for up to 10 years for each iteration. The reduced-form regression for each iteration is run on data that are collapsed to the district-year level. As the figure shows, the widest confidence bounds occur at a threshold of around 350 km. Nonetheless, the reduced-form effect remains statistically significant regardless of the spatial dependence threshold.

coefficients for the lagged rainfall measures are smaller and generally statistically insignificant. Further, the lagged coefficients sum to approximately zero, indicating that previous years' growing-season rainfall has no cumulative effect on the current labor allocation. This pattern of results suggests that any potential demand effects from positive shocks in agricultural productivity are temporary, i.e. they last no more than a year.

The overall negative effect of agricultural output on the agricultural labor share rules out the possibility that the labor pull effect dominates the local demand effect. Therefore, there must be meaningful local demand effects, binding liquidity constraints, or other linkages between sectors. The remainder of the analysis presents additional results to shed light on the most likely explanation.

4.2 Which sectors expand when agricultural productivity increases?

The NSS survey rounds 61, 64, and 66 included a more-detailed question on the primary sector of employment of the household. Each household was classified using a 5-digit industry classification code. The first 1-2 digits of the code allow households to be classified into broad categories such as agriculture, fishing, construction, and manufacturing. This more detailed breakdown of sectors allows me to establish which sectors employ more people when agricultural productivity increases.

While several sectors experience growing labor shares during wetter years, Table 3 shows that the manufacturing, construction and education sectors clearly gain workers when agricultural output increases. The reduced-form effect on the agricultural labor share during these three sample years was 2.1% — a slight variation from the overall effect during all four sample years in Table 2. In 2000 the average Indian district had a rural population of approximately 1.27 million individuals. A one standard deviation increase in growing-season precipitation would therefore be predicted to cause roughly 26,670 individuals to leave the agricultural sector. The remaining coefficients in Columns 2-10 indicate the sectors that grow during wet years: 5,080 individuals enter manufacturing, 10,160 enter the construction sector, 2,540 enter retail, 5,080 enter the education sector, and 3,810 enter other sectors.

Construction and education are the only sectors that are substantially over-represented in the inflow of workers. These two sectors account for 27.9% of the non-agricultural workforce. Yet, they account for 57.1% of the households that move sectors as a result of agricultural productivity gains. Notably, occupations in the construction sector in rural districts of India are almost entirely in residential construction.[17] In addition, occupations in the education

[17]Using the 5-digit industry codes to further classify households, 79.7% of households in the construction sector are listed as having a principal occupation in residential construction.

sector are concentrated in primary and secondary education.[18] The results are therefore consistent with general equilibrium models where rising agricultural income increases the demand for local non-tradables.

In contrast, the results do not appear to be driven by linkages in factor markets between the agricultural and non-agricultural sectors. Table A1 separates the retail and manufacturing sectors into food and non-food components. If agricultural markets are local and cheaper food prices cause the non-agricultural sector to expand, then the increases in the non-agricultural labor share should be concentrated in food retailing and manufacturing. The evidence in the table suggest this is not the case and suggests that linked factor markets are not responsible for the results.

4.3 Total employment, subsidiary occupations, and days of work

The household-level observations suggest strongly that the share of households working primarily in agriculture falls when agricultural productivity is high. This could result from either a direct reallocation of workers from the agricultural to the non-agricultural sectors, or a more rapid expansion of workers in the non-agricultural sector. In short, focusing on employment shares makes it impossible to distinguish between these two explanations. I next turn to the individual-level data from NSS to better understand this distinction.

Similarly to households, each individual household member in the survey lists their primary occupation and the corresponding industry of occupation. In addition, the survey identifies individuals that are not employed — either due to involuntarily unemployment or allocation of time to other tasks such as household duties. I can therefore estimate directly whether growing-season rainfall influences the overall employment probability. If so, then part of the effect on the agricultural labor share may be due to increased employment and not reallocation of workers between sectors.

I find no evidence that the probability of being in the labor force responds to rainfall during the growing season. Column 1 of Table 4 shows that the reduced-form effect on the probability of being employed — amongst all adults aged 18 to 70 — is close to zero and statistically insignificant. This zero effect is precisely estimated. The upper bound on the 95 percent confidence interval is 0.003, meaning that effects on the employment probability of over 0.5 percent can be ruled out. Focusing on employed individuals in column 2, the agricultural labor share decreases by 1.7 percentage points when rainfall in the growing season is one standard deviation above average — a result consistent with the household-level results above. In combination, the results in columns 1 and 2 suggest short-run gains in

[18] 89.4% of households in the education sector list their principal occupation as being in primary or secondary education.

agricultural productivity cause households to shift primary occupations from the agricultural to the non-agricultural sectors. The increased demand for labor due to the local economic effects of agricultural productivity appear to more than offset any direct increases in the demand for agricultural labor.

The individual-level data allow further investigation of how labor is allocated in response to shocks in agricultural productivity. In addition to primary occupations, each household member reports whether they have a subsidiary occupation, defined as an occupation that is pursued for at least month during the 365 days preceding the survey. Column 3 again shows small and statistically insignificant effects of rainfall shocks on the probability that an adult household member has a secondary occupation. However, column 4 shows that people with secondary occupations are *more likely* to hold these occupations in the agricultural sector during good rainfall years. More specifically, the probability of having a secondary occupation in the agricultural sector increases by 2.5 percentage points, or about 3.2 percent, when growing-season rainfall is one standard deviation above average.

Taken together, the results seem compatible with a somewhat simple explanation. Individuals allocate more of their time to non-agricultural work when productivity is high due to increased labor demand in the non-farm sector. However, individuals continue to work in agriculture, they just shift from being primarily engaged in agriculture to working in agriculture "on the side" while becoming primarily engaged in non-agricultural work. In short, agriculture shifts from a main activity to a side activity when productivity is high. This makes sense since agricultural labor in rainfed areas is seasonal. The growing season lasts around five months and thus individuals that obtain more non-agricultural work outside of the growing season are no less likely to be employed when productivity is high. Rather, they are more likely to earn a major share of their income from these additional non-farm employment opportunities. The result also highlights the importance of accounting for the multiple activities undertaken by rural households.

The results in Table 4 leave open the possibility that non-agricultural employment does not actually increase when rainfall is good for agriculture. I next show that I obtain similar results when using the actual number of days worked in different sectors during the 7 day window that precedes the date of the survey. I simply aggregate the total number of days worked in agriculture and non-agriculture and instead use these as dependent variables in an otherwise identical regression framework.

The number of days worked in agriculture decreases, and the number of days worked in non-agriculture increases by a similar amount when growing-season rainfall is above average. Column 1 of Table 5 shows that the average individual works 0.08 days fewer in agriculture when rainfall is one standard deviation above average. This is about a 2 percent effect since the average rural individual worked in agriculture for four of the seven days. The effect is therefore similar in percentage terms to the above effects on binary measures of principal occupations. Turning to column 2, the decreased employment in agriculture is entirely accounted for by non-agricultural work. A one standard deviation increase in growing-season rainfall causes the number of days worked in the non-farm sector to increase by 0.08 days. These results rule out the possibility that any of the above results are due to how households or individuals are classified into primary occupations. Instead, the estimated effects on sectoral labor shares reflect actual changes in work activity.

Finally, NSS also classifies individuals into sectors based on their primary sector during the 7 days preceding the survey. I show in the online appendix that the results using this measure are similar to the above results.[19]

4.4 Heterogeneous effects

I next show that the labor reallocation during good agricultural years tends to be concentrated amongst better-off households. Table 6 tests for heterogeneity along two dimensions: the caste and primary education of the household head. Focusing on column 1, much of the effect of rainfall on exit from agriculture is amongst the higher-caste households. A one standard deviation increase in wet-season precipitation causes the agricultural labor share to decrease by 0.5 percentage points amongst the 30% of households that belong the Scheduled Caste and Scheduled Tribe groups, while the effect amongst the higher caste groups is 1.4 percentage points. Column 2 shows a similar result that there is no effect of rainfall on labor reallocation amongst the individuals that have less than a primary education. Thus, all of the labor reallocation that occurs as a result of shocks to agricultural productivity is concentrated amongst the individuals that have at least a primary education. Columns 3 and 4 show that these heterogeneous reduced-form effects do not result from heterogeneous first-stage effects. Output in the agricultural sector is no more sensitive to growing-season rainfall in areas where higher-caste and more educated people reside.

Taken together, the results are all consistent with movement out of agriculture being restricted to the wealthiest and most educated households. One plausible explanation of the finding is that off-farm opportunities in rural areas tend to be greatest for better-off households (Lanjouw and Stern, 1998). If rural non-farm activities require semi-skilled labor, then only these better-off households will benefit when rising incomes cause increasing demand for labor in the non-farm sector. In addition, the findings on education are consistent with

[19]Table A2 shows that the agricultural labor share during the 7 days before the survey decreases by 1.3 percentage points with a one standard deviation increase in growing-season rainfall. This is the same size effect as the reduced-form effect in Table 2.

the observation that educational attainment restricts access to the off-farm labor market in rural areas of developing countries (de Janvry and Sadoulet, 2001).

4.5 Effect of agricultural productivity on the rural wage

If labor markets are competitive and labor is paid according to marginal productivity then the rural wage should increase when rainfall is favorable for agriculture. The empirical literature from India has indeed found that agricultural wages are generally higher during good rainfall years (Jayachandran, 2006; Kaur, 2015). Mobility of labor across sectors would additionally lead to rising non-agricultural wages.

Column 1 in Table 7 shows that there is a modest and weakly significant positive effect of growing-season rainfall on non-farm wages. Non-farm wages increase by about 1.3 percent when growing season rainfall is one standard deviation above its local average. The effect on agricultural wages is also positive, but is estimated imprecisely. However, the average rural wage increases by 2.6 percent with a one standard deviation increase in rainfall. These results are only compatible when high rainfall also causes labor to shift to the higher-wage non-agricultural sector, as the above results demonstrated. In other words, the increase in average wages is made up of two parts: increasing local demand for labor which drives up wages and movement of workers to the more productive non-farm sector. These results suggest that the latter effect accounts for at least half of the overall effect on the rural wage.

4.6 Decomposing the effect of gains in agricultural productivity

The results thus far indicate that improvements in agricultural productivity increase output per worker through two channels. The first is the direct effect of increased agricultural output, which is of course an increasing function of the agricultural labor share. The second is the indirect effect of re-allocating labor to the non-agricultural sector. This second effect depends on the difference in output per worker between the agricultural and non-agricultural sector. While some literature suggests that the rural non-farm sector is a low productivity sector, other results suggest that rural non-farm productivity in developing countries has grown significantly in recent decades (Lanjouw and Lanjouw, 2001).

The data show that agriculture is one of the lowest productivity sectors across Indian districts. To establish this, I use district-level measures of GDP by sector along with estimates of the number of workers that come from combining labor share estimates from NSS with total district population in 2000. Figure 3 shows that GDP per capita during 2003-2004 in the agricultural sector was around 11,000 Rupees, or approximately 240 USD. In contrast, estimates of GDP per capita in wholesale and retail trade, manufacturing, and construction are generally larger by a factor of 2-3 times.

Building on this productivity gap, I further decompose the economic effect of an agricultural productivity shock. I consider two sectors, agriculture (a) and non-agriculture (n). Output per worker in a district can then be written as $y = s_a y_a + s_n y_n$, where s_a and s_n denote labor shares and y denotes output per worker. I consider a hypothetical scenario where agricultural output goes up by 10%, i.e. Y_a goes up by 10%. The overall effect of the productivity shock on output per worker in the district is

$$dy = s_a dy_a + ds_n(y_n - y_a). \tag{3}$$

The first term, $s_a dy_a = \frac{dY_a}{L}$, is the direct effect of the productivity shock due to increased output in the agricultural sector — when holding the allocation of labor constant. The second term $ds_n(y_n - y_a)$ is the labor reallocation effect, or the increase in GDP per capita that results from more workers moving to the more productive non-agricultural sector. Empirically, I observe Y_a and Y_n in the data, and am able to estimate L_a and L_n by multiplying labor shares from NSS with district-level population estimates. In addition, I use the IV results in Table 2 to estimate $ds_n = 0.1 * 0.109 = 0.0109$.

The results of the decomposition exercise indicate that a significant share of the gains from the productivity shock in agriculture are due to labor reallocation. Figure 4 shows two distributions of predicted changes in GDP per capita across Indian districts. The first is the counterfactual distribution of changes across districts when assuming that the productivity shock does not cause labor reallocation ($s_a dy_a$). Returning to the terminology above, this is the direct effect of the productivity shock. The second distribution includes both this direct effect and the additional effect due to movement of labor to the non-agricultural sector ($ds_n(y_n - y_a)$). The clear rightwards shift in the distribution demonstrates the economic importance of labor reallocation. More specifically, the average direct effect due to increasing agricultural productivity is an increase in GDP per capita of 477 rupees, or around 2.5%. The average labor reallocation effect is around 212 rupees, or around 1.1%. Combining these two figures, around 31% of the gains from a 10% increase in agricultural output are expected to arise due to reallocation of labor to the non-agricultural sector. Therefore, the reallocation of labor across sectors is an important channel through which improvements in agricultural productivity affect welfare.

While this calculation is based on the assumption that labor mobility is the only spillover effect of agricultural productivity on the non-agricultural sector, regression evidence in the online appendix paints a similar picture. Table A3 shows that a one standard deviation increase in growing-season rainfall causes agricultural GDP to increase by approximately 5.2%. However, the same rainfall shock leads to a 1.2% increase in GDP of the non-agricultural sector. The corresponding effect on overall GDP is approximately 2.4%. This pattern of results is qualitatively consistent with the results on labor reallocation.

5 Robustness

I next show that the main results are robust to several alternate assumptions and modeling choices. I start with the most pressing concern which is the exclusion restriction assumption that agriculture productivity is the only channel through which growing-season precipitation shocks influence labor allocation. I then turn to several other robustness tests which include using different measures of weather shocks, choice of climate data, and estimation of local price effects.

Exclusion restriction

The main identification assumption of the instrumental variables specification is that agricultural productivity is the only channel through which short-term shocks in growing-season precipitation affect how rural households allocate labor across sectors. This assumption is obviously violated if the non-agricultural sector benefits directly from wet weather during the growing season. As one example, my strategy would over-state the labor-reallocation benefits of agricultural productivity if the rural non-farm sector benefits from an abnormally wet growing season because labor productivity is higher when weather is more favorable.[20]

I show two falsification exercises that are inconsistent with the results being driven by violations of the exclusion restriction. First, I use the individual-level NSS data to look at effects on labor force participation rates and sectoral labor shares in urban areas. An increase in growing-season rainfall in urban areas is less likely to affect outcomes via agriculture because the local agricultural sector is much smaller in these areas. More concretely, 9 percent of adults in urban households work in agriculture while this number is 64 percent amongst rural adults.

Column 1 in Table 8 shows that participation in the labor force is unaffected by growing-season rainfall in urban areas. Therefore, if the results are driven by direct effects of growing-season rainfall, then the sectors that expand in rural areas — and presumably benefit from wetter years — should also expand in urban areas by drawing in labor from sectors other than agriculture. Column 2 shows no reduced-form effect on the agricultural labor share in urban areas. More importantly, manufacturing, construction, retail, and education —

[20]Recent work has suggested that labor productivity is one of the channels through which weather (largely temperature) influences productivity. See Dell, Jones, and Olken (2012) and Zivin and Neidell (2014) as examples.

the sectors that expand during wet years in rural areas — are unaffected by growing-season rainfall in urban areas. There is no reason to expect these strong differences between rural and urban households if my estimates were being driven by direct linkages between growing-season rainfall and the non-agricultural sector. Therefore, the result provides suggestive evidence that the results are not explained by a direct effect of rainfall during the growing season on labor demand in the non-agricultural sector.

As a second piece of evidence, I exploit the fact that there is spatial variation across districts in access to groundwater irrigation. Access to groundwater makes agriculture less susceptible to weather variation because farmers can extract groundwater during dry spells and unlike surface water, short-term availability of groundwater is not governed by recent rainfall. I divide the sample into districts that are above and below the median in terms of the share of crop area that is irrigated with groundwater. There is no reason to expect the reduced-form estimates to vary across these two subsamples if violations of the exclusion restriction are responsible for the findings.

The results in Table 9 confirm that all of the labor reallocation effects of rainfall shocks occur in districts where groundwater is less abundant and therefore agriculture is most susceptible to rainfall. Starting with Panel A, the first-stage effect of growing-season rainfall on agricultural output is larger by a factor of nearly five in districts with less access to groundwater. At the same time, Panel B shows that the reduced-form effect of growing-season rainfall on the agricultural labor share is mostly present in districts with less groundwater. In addition, the effect of precipitation in more-irrigated districts is smaller by a factor of three and is statistically insignificant. This pattern of results is consistent with agriculture being the key channel through which precipitation during the growing season affects the allocation of labor across sectors.

Symmetry of effects

The main empirical specification treats positive and negative rainfall shocks symmetrically. That is, I don't measure whether positive and negative shocks have the opposite effects on labor reallocation. Recent work has shown that agricultural labor markets do not respond symmetrically to weather shocks as standard theory would predict. Specifically, Kaur (2015) shows that agricultural wages in rural India increase during good-rainfall years, but fail to decrease during bad-rainfall years.

I test for asymmetric effects by measuring growing-season rainfall shocks with a set of four indicator variables for the quintiles of each district's 1950-2008 rainfall distribution, where years with rainfall from the 40th to the 60th percentile represent the omitted base group. Measuring rainfall in this way — as opposed to with the deviation from the district's long-term mean, allows me to estimate whether the main results are due to positive shocks, negative shocks, or both.

Figure A3 shows evidence suggesting that the labor reallocation effects are primarily due to positive shocks. An extremely dry year in the bottom quintile of the district's rainfall distribution results in a decrease in agricultural output of around 20 percent. Yet, the agricultural labor share remains roughly the same as in an average year. In contrast, the agricultural labor share decreases — and agricultural output increases — when rainfall during the growing season falls in the top two quintiles of the distribution.

One possible explanation for this pattern is that government relief during bad rainfall years mitigates the negative income effects of droughts and effectively shuts down the local income channel. Another possibility is that agricultural wages only adjust upwards during good years — as in Kaur (2015) — and thus the local income channel is more operative during these years. These explanations are not mutually exclusive and I do not distinguish between them in the analysis.

Different measures of weather shocks

The main analysis uses the normalized difference between a district's wet-season rainfall and its' long-term average as a precipitation-shock measure.[21] The literature relating agricultural outcomes to weather has not arrived at a consistent way of measuring shocks to precipitation — or temperature. I next show that my results remain similar when alternative plausible measures are used.

The timely arrival of monsoon rainfall is important for Indian agriculture. Based on this, Kaur (2015) uses rainfall shocks during the first month of the monsoon to test the sensitivity of agricultural wages to productivity shocks. I show in Figure A4 that June is the month with the sharpest increase in monthly rainfall, corresponding to the most frequent month of monsoon arrival. I show results in the first row of Table A4 where I instead use rainfall during June, conditioning on monthly measures of rainfall during the other months. This alternative measure of the precipitation shock produces results that are roughly similar to the main results.

Jayachandran (2006) and Adhvaryu, Chari, and Sharma (2013) instead define rainfall shocks with a numeric variable taking on 3 values depending on whether a district's rainfall is above the 80th percentile, in between the 20th and 80th percentiles, or below the 20th percentile. Results in the second row of Table A4 show that I also obtain similar results with this more discrete measure.

[21] Sekhri and Storeygard (2014) use the same measurement approach when relating rainfall shocks to violence against women in Indian districts.

As a final measure, I consider temperature. To my knowledge, none of the studies on the economic impacts of weather shocks in India consider how most crops respond differently to high temperatures during different phases of the growing season. I show this in Figure A5 which shows that temperatures in September-October have the largest impact on wet-season crops. This is consistent with basic agronomics because cereal crops (rice and maize) are in the flowering stage at this time, which is the stage when output is most sensitive to temperature fluctuations (Yoshida, 1981).[22] In addition to being the most important for agricultural productivity, temperature variations during the months of September and October are the most important for labor reallocation (bottom panel of Figure A5). Combining these results, the last row of Table A4 shows that the IV estimate is almost the same when using temperature during September-October — the most sensitive time for cereal crops — as the instrument.

Combining these additional tests, it does not appear that the main relationship I estimate is an artifact of how the precipitation shock was measured. This is an important verification because the literature has taken several different approaches to measuring weather shocks when relating such shocks to agricultural productivity.

Local price effects

I generated the productivity measure based on national-level agricultural prices. The effect of rainfall shocks on revenue productivity at the district level may be smaller if local prices are responsive to variation in rainfall. I test this by estimating the relationship between local prices for rice and corn — two important wet-season crops — and rainfall shocks. The price data come from the Indian Ministry of Agriculture which makes available monthly wholesale prices based on auctions by the Agricultural Produce Market Committee.[23]

Using these price data, I show that local price responses are likely to offset only a small portion of the physical productivity gains that arise during good rainfall years. Table A5 shows that a one standard deviation increase in rainfall during the most recent wet season decreases the prices of rice and corn by 1.8 and 0.8 percent respectively. Recall that the first-stage effect in Table 2 is 0.108. This is an effect on physical productivity since the prices used to construct the measure do not vary across districts. Combining these two estimates, the modest effects on local prices appear to offset less than 20 percent of the gains in physical productivity when wet-season rainfall is one standard deviation above average.

[22] Feng, Oppenheimer, and Schlenker (2015) take advantage of this to measure the effect maize productivity in the U.S. corn belt on migration. Their first-stage effects show that maize yields in the U.S. are most responsive to extreme heat during the flowering stage.

[23] The data are available at http://agmarknet.dac.gov.in (accessed August 2016).

Migration and composition of the sample

One possibility is that positive rainfall shocks increase wealth and these wealth effects cause agricultural households to migrate. This could reduce the agricultural labor share by changing the composition of the sample — and have little to do with labor demand in the non-agricultural sector.

While the data are somewhat limited, I test for effects on outmigration using rounds 55 (1999 to 2000) and 64 (2007 to 2008) of the NSS survey. These two rounds include a module on temporary labor outmigration of household members.[24] Table A6 shows small and insignificant effects of growing-season rainfall on outmigration. These results are inconsistent with migration, or at least temporary labor migration, being the driving force behind the effect of growing-season rainfall on labor reallocation.

These non-results on migration also suggest that liquidity constraints are not the important mechanism for the relationship between agricultural productivity and labor reallocation. As the literature on migration has pointed out, labor reallocation could follow positive income shocks if there are large upfront costs to moving sectors or migrating (Munshi, 2003; Bryan, Chowdhury, and Mobarak, 2014; Kleemans, 2014; Angelucci, 2015; Bazzi, 2016). Large upfront costs of labor reallocation are likely to be important when such reallocation requires actual labor migration. The absence of an effect on temporary labor migration therefore suggests that upfront costs of labor reallocation do not explain the results.

District-specific linear time trends

The main specifications control only for time-varying unobservables that do not vary over space, i.e. across districts. The additional specifications in Table 2 control for time-varying unobservables that vary across — but not within — regions. Another approach is to include district-specific linear time trends in order to eliminate time-varying unobservables that are specific to districts, but vary approximately linearly over time.

Table A7 shows that the results are robust to these additional controls. While the first-stage coefficient decreases slightly and becomes less precisely estimated (F=6.38), the reduced-form effect becomes larger and the corresponding IV estimate is approximately the same as the one obtained in Table 2 when controlling for region-specific time effects.

[24] One limitation is that the question is asked differently between the two rounds of the survey. The 55th round identifies migrants as those that left the house for at least 60 days, but no more than 6 months during the last year while the 64th round uses a lower threshold of 30 days.

Sampling weights

While I simplify the analysis by not applying NSS sampling weights to the IV estimator, this has no meaningful impact on the results. Table A8 shows the reduced-form estimates where the regression is weighted using the NSS sampling weights. The reduced-form estimates — both with and without region-by-year fixed effects — are similar to those that were estimated without applying sampling weights. Therefore, the results are unaffected by the choice of whether or not to apply sampling weights.

Climate data

The main source of climate data in the analysis is the UDEL terrestrial precipitation and temperature gridded monthly time series. These data are derived from individual station-level observations and thus one minor disadvantage is the lower spatial resolution (approximately 50 km). An additional source of precipitation data is NASA's Tropical Rainfall Measuring Mission (TRMM). These satellite-based observations are less commonly used, but have the advantage of a higher spatial resolution of approximately 25 km. Also, the TRMM climate data consist of daily — rather than monthly — observations.

The main instrumental variables results are if anything larger when using this alternative climate data source. Table A9 shows the first-stage, reduced-form and instrumental variables results with the TRMM data.[25] While the first-stage and reduced-form effects are smaller than those observed with the UDEL data, the instrumental variables estimates are larger by a factor of around 1.5. The choice of climate datasets therefore has no meaningful impact on the results.

6 Discussion and Conclusion

This paper has shown that short-term increases in agricultural productivity cause a decrease in the agricultural labor share and a corresponding increase in the amount of labor allocated to the non-agricultural sector. Most importantly, short-term gains in agricultural productivity in India cause expansion in the manufacturing, construction, retail, and education sectors. The strong re-allocation effects in the latter three sectors are consistent with agricultural productivity causing incomes to rise and leading to increased demand for local goods and services. Additional results suggest that the findings are less consistent with alternative explanations such as liquidity constraints or linkages in production between sectors.

[25] The available years are 1998-2013. Thus, the district-level long-term average and standard deviations of precipitation are based on a shorter time period.

I have also shown that even small gains in agricultural productivity can have meaningful impacts on total economic output. This occurs through two channels: a direct effect and the above-mentioned labor-reallocation effect. The importance of this latter effect is driven by the gaps in output per worker across sectors. In the average Indian district output per worker in agriculture is around one third of that in the other sectors of the economy. As a consequence, the increase in the non-agricultural labor share caused by gains in agricultural productivity is economically significant. My best estimate is that a 10% increase in agricultural productivity would lead to a 3.6% increase in GDP per capita. I estimate that around 31% of this increase would be due to re-allocation of labor to the non-agricultural sector.

One argument against strong investment in agricultural development is that such investments will keep people engaged in a relatively low productivity sector. On the contrary, my results demonstrate that labor reallocation is an important channel through which pro-agricultural policies will enhance growth. In other words, the data from rural India do not support the idea that growth in agricultural productivity will prevent or even slow the movement of labor out of agriculture. Rather, the results indicate that the movement of labor out of agriculture should be considered as an important consequence of policies that enhance agricultural productivity.

It is important to stress that I use short-term variation in annual weather. As a result, my results are not descriptive of slow-moving gains in productivity over time. One example is the well-studied Green Revolution, which led to large increases in the productivity of staple crops over the period from 1960-2000 (Evenson and Gollin, 2003). This long and drawn-out period of agricultural growth differs fundamentally from year-to-year shocks in agricultural output. Thus, my results do not offer insights into how sustained periods of growth such as the Green Revolution affect the allocation of labor across sectors of the economy.

Nonetheless, the results highlight the importance of the agricultural sector in rural development. The reliance of such a large portion of the labor force on agriculture in developing countries suggests that cross-sectoral spillovers may be important. My results are consistent with this argument that the non-agricultural sector can grow by gaining workers during years when agriculture is particularly productive.

References

Adhvaryu, Achyuta, AV Chari, and Siddharth Sharma. 2013. "Firing Costs and Flexibility: Evidence from Firms' Employment Responses to Shocks in India." *Review of Economics and Statistics* 95 (3):725–740.

Allcott, Hunt and Daniel Keniston. 2014. "Dutch Disease or Agglomeration? The local Economic effects of natural resource booms in modern America." Tech. rep., National Bureau of Economic Research.

Alvarez-Cuadrado, Francisco and Markus Poschke. 2011. "Structural change out of agriculture: Labor push versus labor pull." *American Economic Journal: Macroeconomics* :127–158.

Angelucci, Manuela. 2015. "Migration and Financial Constraints: Evidence from Mexico." *Review of Economics and Statistics* 97 (1):224–228.

Bazzi, Samuel. 2016. "Wealth heterogeneity and the Income Elasticity of Migration." *forthcoming, American Economic Journal: Applied Economics* .

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica* 82 (5):1671–1748.

Bryan, Gharad and Melanie Morten. 2015. "Economic Development and the Spatial Allocation of Labor: Evidence From Indonesia." Unpublished.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli. 2016. "Agricultural Productivity and Structural Transformation: Evidence from Brazil." *The American Economic Review* 106 (6):1320–1365.

Colmer, Jonathan. 2016. "Weather, labour reallocation, and industrial production: Evidence from india." Unpublished.

de Janvry, Alain and Elisabeth Sadoulet. 2001. "Income strategies among rural households in Mexico: The role of off-farm activities." *World development* 29 (3):467–480.

———. 2009. "Agricultural growth and poverty reduction: Additional evidence." *The World Bank Research Observer* :lkp015.

Deaton, Angus and Jean Dreze. 2002. "Poverty and inequality in India: a re-examination." *Economic and Political Weekly* :3729–3748.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2012. "Temperature shocks and economic growth: Evidence from the last half century." *American Economic Journal: Macroeconomics* 4 (3):66–95.

Duarte, Margarida and Diego Restuccia. 2010. "The role of the structural transformation in aggregate productivity." *The Quarterly Journal of Economics* 125 (1):129–173.

Evenson, R.E. and D. Gollin. 2003. "Assessing the impact of the Green Revolution, 1960 to 2000." *Science* 300 (5620):758–762.

Feng, Shuaizhang, Michael Oppenheimer, and Wolfram Schlenker. 2015. "Weather Anomalies, Crop Yields, and Migration in the US Corn Belt." Unpublished.

Foster, Andrew D and Mark R Rosenzweig. 1996. "Technical change and human-capital returns and investments: evidence from the green revolution." *The American Economic Review* :931–953.

———. 2004. "Agricultural productivity growth, rural economic diversity, and economic reforms: India, 1970–2000*." *Economic Development and Cultural Change* 52 (3):509–542.

———. 2007. "Economic development and the decline of agricultural employment." *Handbook of Development Economics* 4:3051–3083.

Gollin, Douglas, David Lagakos, and Michael E Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2):939–993.

Gollin, Douglas, Stephen Parente, and Richard Rogerson. 2002. "The role of agriculture in development." *American Economic Review* :160–164.

Gollin, Douglas, Stephen L Parente, and Richard Rogerson. 2007. "The food problem and the evolution of international income levels." *Journal of Monetary Economics* 54 (4):1230–1255.

Gollin, Douglas and Richard Rogerson. 2014. "Productivity, transport costs and subsistence agriculture." *Journal of Development Economics* 107:38–48.

Harris, John R and Michael P Todaro. 1970. "Migration, unemployment and development: a two-sector analysis." *The American Economic Review* :126–142.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. 2013. "Two perspectives on preferences and structural transformation." *The American Economic Review* 103 (7):2752–2789.

Hornbeck, Richard and Pinar Keskin. 2015. "Does agriculture generate local economic spillovers? Short-run and long-run evidence from the ogallala aquifer." *American Economic Journal: Economic Policy* 7 (2):192–213.

Hsiang, Solomon M. 2010. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *Proceedings of the National Academy of Sciences* 107 (35):15367–15372.

Jayachandran, Seema. 2006. "Selling labor low: Wage responses to productivity shocks in developing countries." *Journal of Political Economy* 114 (3):538–575.

Kaur, Supreet. 2015. "Nominal wage rigidity in village labor markets." Unpublished.

Kleemans, Marieke. 2014. "Migration Choice under Risk and Liquidity Constraints." Unpublished.

Kochar, Anjini. 1999. "Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural India." *Review of Economics and Statistics* 81 (1):50–61.

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie. 2001. "Beyond balanced growth." *The Review of Economic Studies* 68 (4):869–882.

Lagakos, David and Michael E Waugh. 2013. "Selection, Agriculture and Cross-Country Productivity Differences." *American Economic Review* 103 (2):948–980.

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

Lanjouw, Peter and Nicholas Stern. 1998. *Economic development in Palanpur over five decades*. Oxford University Press.

Matsuyama, Kiminori. 1992. "Agricultural productivity, comparative advantage, and economic growth." *Journal of Economic Theory* 58 (2):317–334.

McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo. 2014. "Globalization, structural change, and productivity growth, with an update on Africa." *World Development* 63:11–32.

McMillan, Margaret S and Kenneth Harttgen. 2014. "What is driving the'African Growth Miracle'?" Tech. rep., National Bureau of Economic Research.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. "Economic shocks and civil conflict: An instrumental variables approach." *Journal of Political Economy* 112 (4):725–753.

Munshi, Kaivan. 2003. "Networks in the modern economy: Mexican migrants in the US labor market." *The Quarterly Journal of Economics* :549–599.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. 2008. "Agriculture and aggregate productivity: A quantitative cross-country analysis." *Journal of Monetary Economics* 55 (2):234–250.

Sekhri, Sheetal and Adam Storeygard. 2014. "Dowry deaths: Response to weather variability in India." *Journal of Development Economics* 111:212–223.

Topalova, Petia. 2010. "Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India." *American Economic Journal: Applied Economics* 2 (4):1–41.

Vollrath, Dietrich. 2009. "How important are dual economy effects for aggregate productivity?" *Journal of Development Economics* 88 (2):325–334.

Willmott, CJ and Kenji Matsuura. 2009. "Terrestrial air temperature and precipitation: 1990-2008 Gridded Monthly Time Series, Version 2.01." *University of Delaware* .

Yoshida, Shouichi. 1981. *Fundamentals of rice crop science.* International Rice Research Institute.

Zivin, Joshua Graff and Matthew Neidell. 2014. "Temperature and the allocation of time: Implications for climate change." *Journal of Labor Economics* 32 (1):1–26.

Tables

Table 1: Summary statistics for rural individuals, by year

	Year				
	1998	2003	2006	2008	Total
Panel A: Primary occupation in:					
Agriculture	0.75	0.69	0.70	0.66	0.70
	(0.43)	(0.46)	(0.46)	(0.47)	(0.46)
Manufacturing	0.07	0.09	0.08	0.07	0.08
	(0.26)	(0.28)	(0.26)	(0.26)	(0.27)
Construction	0.03	0.05	0.06	0.10	0.06
	(0.18)	(0.22)	(0.24)	(0.29)	(0.24)
Retail	0.04	0.05	0.05	0.05	0.05
	(0.20)	(0.23)	(0.21)	(0.23)	(0.22)
Education	0.01	0.02	0.02	0.02	0.02
	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)
Panel B: Weather variables					
Wet-season rainfall, mm	1,049.51	982.81	1,007.47	1,139.94	1,042.79
	(529.50)	(527.59)	(592.49)	(532.33)	(550.05)
Dry-season rainfall, mm	186.99	114.39	178.89	171.20	160.85
	(149.52)	(116.62)	(164.41)	(165.12)	(152.40)
Wet-season deviation in SD's	0.30	-0.17	-0.12	0.52	0.12
	(0.60)	(0.77)	(0.90)	(0.82)	(0.84)
Dry-season deviation in SD's	0.89	-0.39	0.55	0.44	0.33
	(1.56)	(0.55)	(0.71)	(0.94)	(1.09)
Total observations	328,525				

The data in Panel A are for the NSS rural sample of households. The table shows means and standard deviations across the four sample years that are included in the analysis. Each observation is an individual that was active in the labor market during the year. NSS sampling weights were applied. The unit of observation in Panel B is the district. The states included are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa

Table 2: Effects of agricultural productivity on the agricultural labor share in rural areas

	Panel A: First-Stage Estimates	
	(1)	(2)
Rainfall deviation	0.108***	0.099***
	(0.015)	(0.016)
Region by year FE	No	Yes
Mean of Dep Variable	21.11	21.11
Number of districts	440	440
Number of Observations	1427	1427
R squared	0.588	0.657
	Panel B: Reduced-Form Estimates	
	(1)	(2)
Rainfall deviation	-0.013***	-0.020***
	(0.003)	(0.004)
Region by year FE	No	Yes
Mean of Dep Variable	0.55	0.55
Number of districts	443	443
Number of Observations	214519	214519
R squared	0.070	0.071
	Panel C: IV Estimates	
	(1)	(2)
Log value	-0.109***	-0.195***
agriculutral output	(0.031)	(0.043)
Region by year FE	No	Yes
Mean of Dep Variable	0.53	0.53
Number of districts	438	438
Number of Observations	177553	177553
R squared	0.056	0.048

Data are for rural households only. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable in Panel A is the log of the value of output from 6 important wet-season crops (rice, soybeans, millet, maize, groundnut and sugarcane). The dependent variable in Panel B is an indicator for households with a principle occupation in the agricultural sector, either as laborers or farmers. Panel C contains instrumental variable estimates where the dependent variable is the indicator for households in the agricultural sector and the log value of agricultural output is instrumented with the rainfall deviation. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 3: Effects of rainfall deviations on sectoral labor shares in rural areas

	Household principal occupation in:									
	(1) Ag	(2) Manufact.	(3) Const.	(4) Retail	(5) Hospitality	(6) Trans.	(7) Public	(8) Educ.	(9) Service	(10) Other
Rainfall deviation	-0.021***	0.004**	0.008***	0.002*	0.000	-0.001	0.000	0.004***	0.001*	0.002
	(0.004)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.56	0.09	0.08	0.08	0.01	0.04	0.02	0.03	0.02	0.06
Number of districts	443	443	443	443	443	443	443	443	443	443
Number of Observations	147974	147974	147974	147974	147974	147974	147974	147974	147974	147974
R squared	0.060	0.025	0.039	0.017	0.008	0.014	0.011	0.012	0.009	0.024

Data are for rural households only. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable is an indicator variable for the principal industry of the household being the sector corresponding to the column title. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 4: Reduced-form effects on employment probability and primary / secondary occupations

	Primary Occupation:		Subsidiary Occupation:	
	(1) Employed	(2) Ag. Sector	(3) Employed	(4) Ag. Sector
Rainfall deviation	-0.001	-0.017***	-0.002	0.025***
	(0.002)	(0.003)	(0.003)	(0.004)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.59	0.64	0.22	0.79
Number of districts	410	410	410	410
Number of Observations	545577	319681	545577	119712
R squared	0.045	0.087	0.062	0.073

The data are for individuals from 18-70 years old living in rural households. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. Columns 1 and 3 include all individuals. Column 2 includes only individuals that have a primary occupation, while column 4 only includes individuals that have a secondary occupation. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 5: Reduced-form effects on number of days worked in 7 days preceding survey

	Number of days worked in last 7 days in:	
	(1)	(2)
	Ag	Non-Ag
Rainfall deviation	-0.080***	0.084***
	(0.022)	(0.022)
District FE	Yes	Yes
Year FE	Yes	Yes
Mean of Dep Variable	3.98	2.32
Number of districts	410	410
Number of Observations	350199	350199
R squared	0.087	0.074

The data are for individuals from rural households that worked for at least 0.5 days during the week before the survey. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable in column 1 is the total number of days worked in agriculture and the dependent variable in column 2 is the total number or days worked in the non-agricultural sector. Both regressions are estimated by OLS and include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 6: Heterogeneous effects of precipitation shocks on the agricultural labor share

	Reduced Form		First Stage	
	(1)	(2)	(3)	(4)
Rainfall deviation	-0.005	-0.003	0.110***	0.139***
	(0.005)	(0.003)	(0.016)	(0.019)
Rainfall deviation * High Caste	-0.009*		0.010	
	(0.005)		(0.009)	
Rainfall deviation * Primary Education		-0.020***		-0.012
		(0.004)		(0.011)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year FE X High Caste	Yes	No	Yes	No
Year FE X Primary Education	No	Yes	No	Yes
Mean of Dep Variable	0.55	0.64	21.22	21.26
Number of districts	443	410	438	394
Number of Observations	214519	339334	177553	277224
R squared	0.072	0.111	0.916	0.927

The dependent variable in columns 1 and 2 is an indicator for households or individuals with a principle occupation in the agricultural sector, either as laborers or farmers. The dependent variable in columns 3 and 4 is the log of the value of output from rice, soybeans, millet, maize, groundnut, and sugarcane in the wet season. The rainfall deviation is the difference between rainfall during the main growing season (wet season) and the long-term average in the district, normalized by the standard deviation. Higher caste is an indicator for non-Scheduled Caste / Scheduled Tribe households. Primary education is an indicator for individuals with at least a primary school education. Columns 1 and 3 include the dry-season precipitation deviation, the high caste indicator, and the interaction between the two. Columns 2 and 4 include the dry-season precipitation deviation, indicator for primary education, and the interaction between the two. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 7: Reduced-form effects of precipitation shocks on wages

	(1) Non-Agricultural	(2) Agricultural	(3) All
Rainfall deviation	0.013*	0.008	0.026***
	(0.008)	(0.007)	(0.006)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mean of Dep Variable	4.53	3.77	4.10
Number of districts	410	409	410
Number of Observations	62564	72565	133587
R squared	0.190	0.496	0.337

The dependent variable in all regressions is the log of earnings per day worked. Column 1 is for non-agricultural activities, column 2 is for agricultural activities, and column 3 is average daily wages across all activities. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average (1950-2008) in the district, normalized by the standard deviation. Both regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 8: Effects of rainfall deviations on sectoral labor shares in urban areas

		Conditional On Employment, Primary Occupation In:									
	(1) Employed	(2) Ag.	(3) Manufact.	(4) Const.	(5) Retail	(6) Hospitality	(7) Trans.	(8) Public	(9) Educ.	(10) Service	(11) Other
Rainfall deviation	-0.001	0.001	0.002	0.001	-0.000	0.000	-0.002*	0.002	0.000	-0.002*	-0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.50	0.09	0.20	0.09	0.20	0.03	0.07	0.06	0.06	0.04	0.17
Number of districts	412	412	412	412	412	412	412	412	412	412	412
Number of Observations	306843	151892	151892	151892	151892	151892	151892	151892	151892	151892	151892
R squared	0.008	0.068	0.053	0.022	0.021	0.009	0.012	0.019	0.010	0.010	0.028

Data are for adults from 18 to 70 living in urban households. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable is an indicator variable for the principal industry of the household being the sector corresponding to the column title. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table 9: Main estimates separated by access to groundwater irrigation

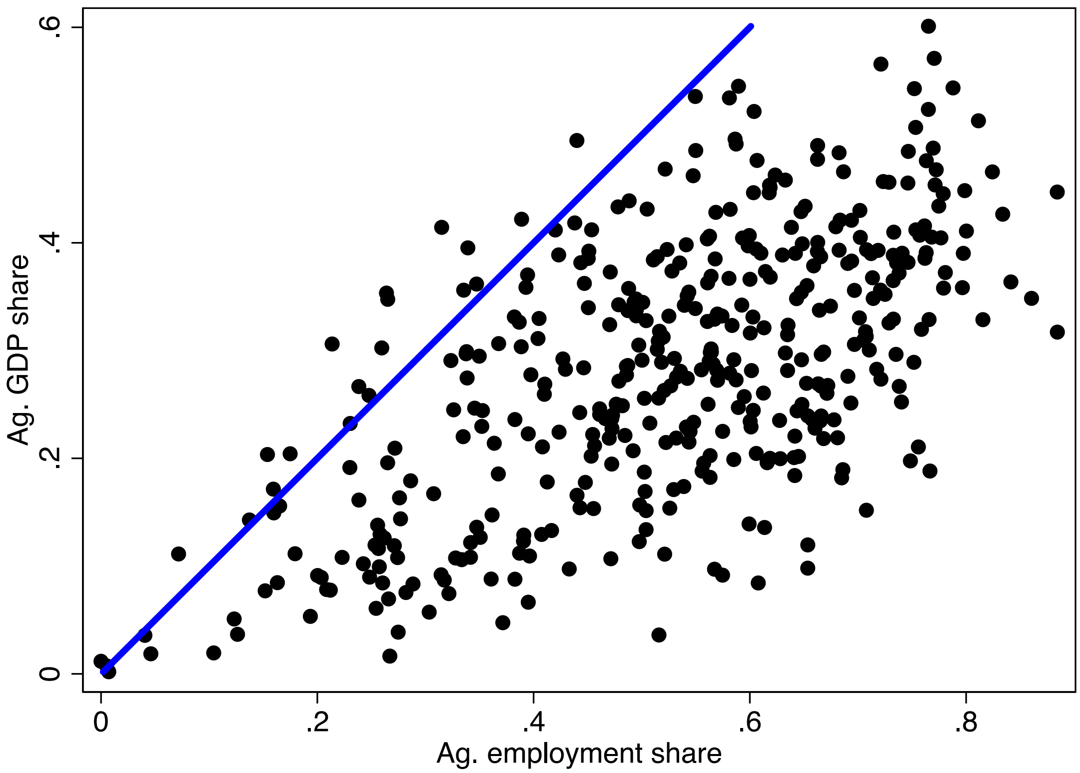
	Panel A: First-stage for districts with groundwater access:	
	(1)	(2)
	Below median	Above median
Rainfall deviation	0.172***	0.036
	(0.021)	(0.024)
District FE	Yes	Yes
Year FE	Yes	Yes
Mean of Dep Variable	20.97	21.26
Number of districts	219	213
Number of Observations	716	689
R squared	0.609	0.605

	Panel B: Reduced-form for districts with groundwater access:	
	(1)	(2)
	Below median	Above median
Rainfall deviation	-0.020***	-0.006
	(0.005)	(0.006)
District FE	Yes	Yes
Year FE	Yes	Yes
Mean of Dep Variable	0.54	0.56
Number of districts	218	213
Number of Observations	97837	99509
R squared	0.085	0.053

The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable in Panel A is the log of the value of output from 6 important Kharif crops (rice, soybeans, millet, maize, groundnut and sugarcane). The dependent variable in Panel B is an indicator for households with a principle occupation in the agricultural sector, either as laborers or farmers. Groundwater access is defined as the share of average annual cultivated area from 2000-2002 in the district that is equipped with groundwater irrigation. The median value of groundwater access is 20.4% of cultivated area. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

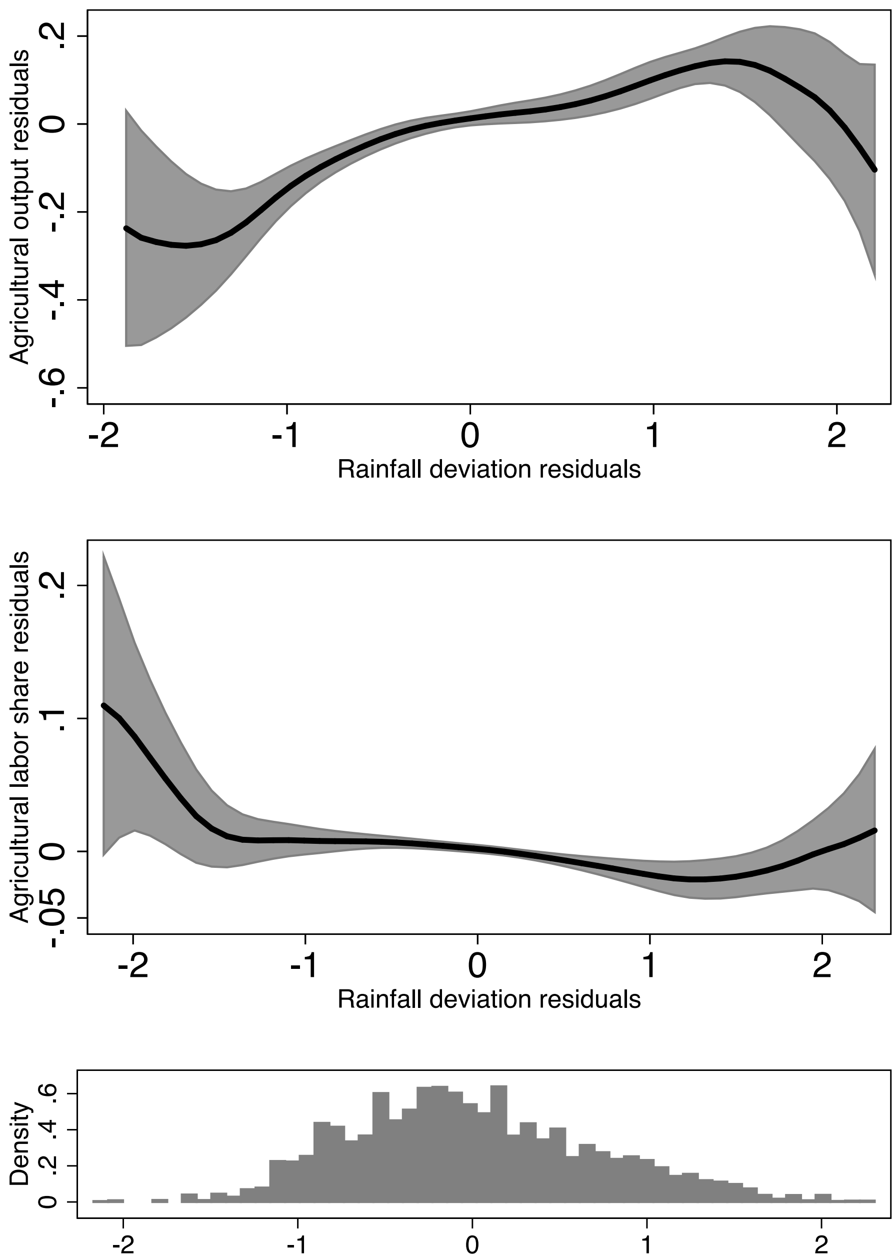
Figures

Figure 1: Agriculture's share of labor and share of output across Indian districts in 2004-2005



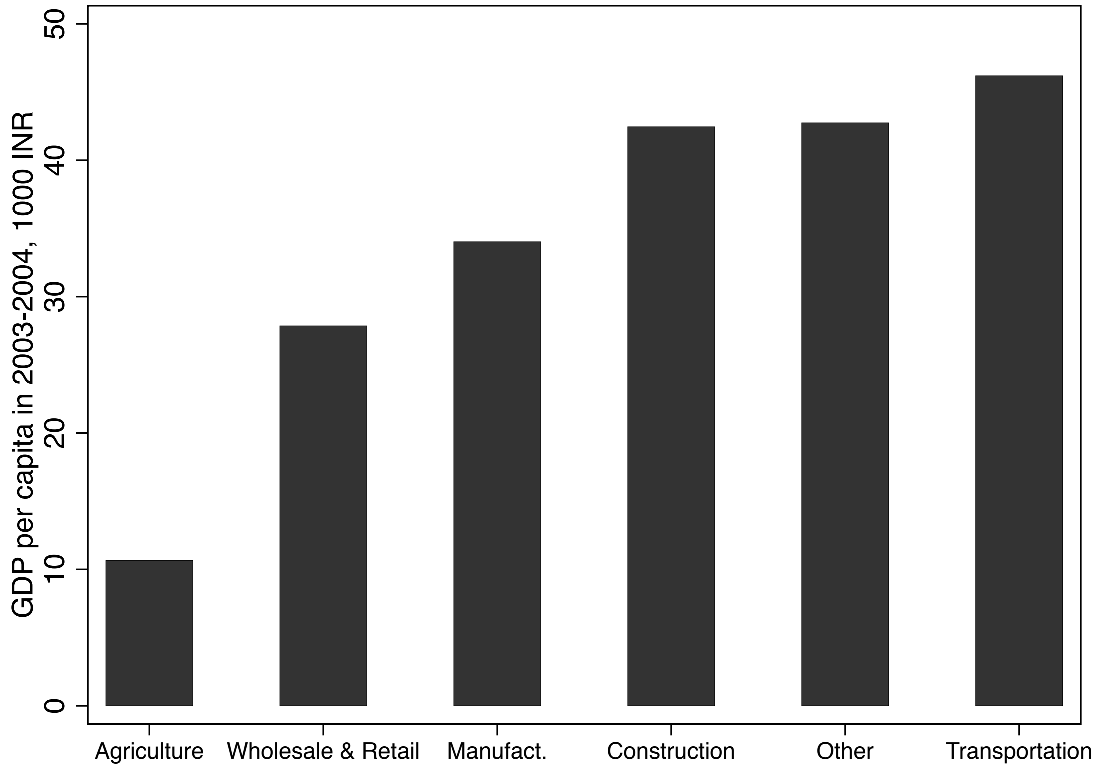
Notes: District-level GDP per capita data are from the Indian planning commission. Labor shares are calculated using NSS Round 61, with the application of sampling weights. The data include all districts in the main agricultural states of Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa. The blue line is the 45° line

Figure 2: Partial relationships between growing-season precipitation, the agricultural labor share, and agricultural output



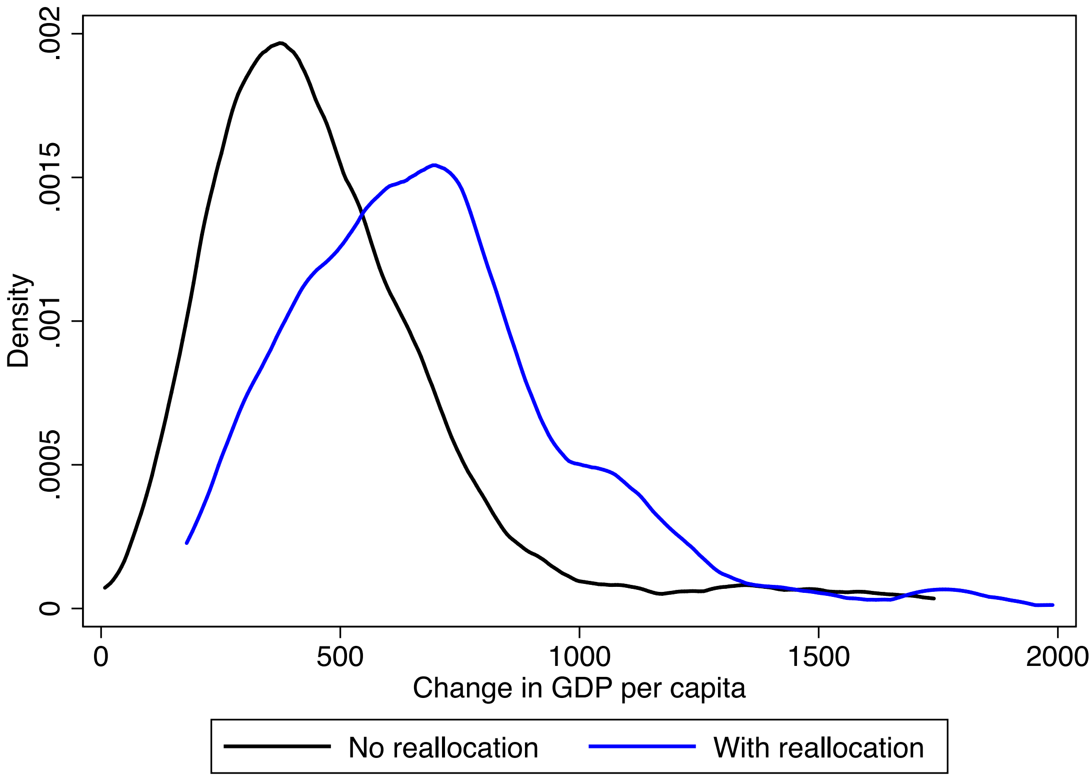
Notes: The data are at the district-level for years 1998, 2003, 2006, and 2008. The figure shows the partial relationship between growing-season rainfall and agricultural output (top panel), the partial relationship between growing-season rainfall and the agricultural labor share (middle panel), and the distribution of growing-season rainfall residualized z-scores (bottom panel). Growing season rainfall, agricultural output, and agricultural employment were first residualized to net out the effects of dry-season rainfall, district, and year fixed effects. The black lines in the top two panels are estimated using non-parametric fan regressions and the shaded areas represents the 95% confidence intervals. Standard errors were estimated using bootstrapping to correct for clustering at the district level.

Figure 3: GDP per capita across Indian districts in 2003-2004, by sector



Notes: Figure shows average GDP per capita estimates across 416 districts from 2003-2004. The data on total GDP by sector is combined with estimates of the total population dependent on each sector to calculate sector-wise GDP per capita. The estimates of the total population dependent on each sector are generated by multiplying sectoral labor shares from NSS with district-level population.

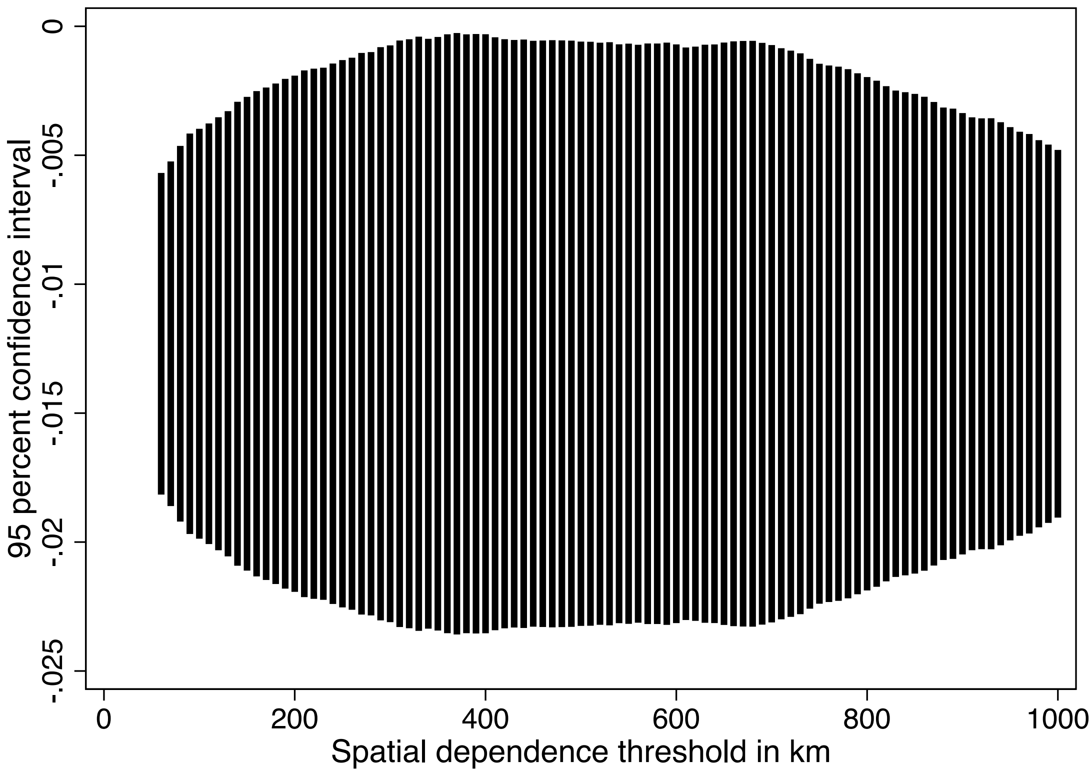
Figure 4: Simulated effects of a 10% increase in agricultural productivity, with and without sectoral reallocation of labor



Notes: Black density is the predicted impact of a 10% increase in agricultural productivity on GDP per capita, while holding constant the sectoral allocation of labor. The density is drawn across Indian districts. The blue line is the density when the productivity shock also causes more labor to move to the more productive non-agricultural sector.

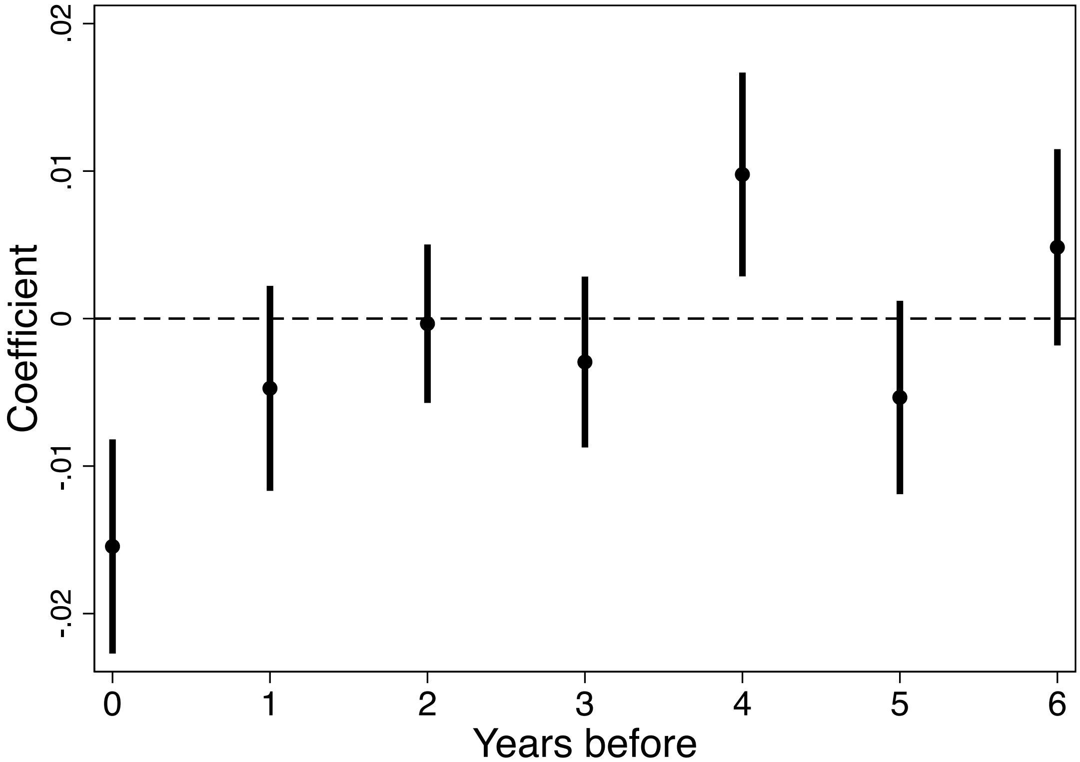
Appendix A: Additional Tables and Figures

Figure A1: Sensitivity of reduced-form standard errors to spatial dependence



Notes: The figure shows 95 percent confidence intervals when standard errors are corrected for both spatial and temporal dependence, as in Hsiang (2010). Errors for nearby districts are allowed to be correlated up to the threshold corresponding to the value along the horizontal axis. Serial correlation is assumed to decay linearly up to a threshold of 10 years, after which errors are assumed uncorrelated.

Figure A2: Effects of current and lagged rainfall shocks on the agricultural labor share



Notes: The figure shows the coefficients (dots) and 95 percent confidence intervals (bars) from a regression of an indicator for working in agriculture on the current growing-season precipitation deviation and 6 lags. The regression also includes the current and lagged deviations of rainfall during the months outside the growing season. The sum of all seven coefficients for growing-season rainfall is -.014 and its p-value is 0.08.

Table A1: Reduced-form effect of precipitation shocks on probability of working in the food sector

	Retail occupation in:		Manufacturing occupation in:	
	(1)	(2)	(3)	(4)
	Food	Non-Food	Food	Non-Food
Rainfall deviation	-0.0002	0.0025***	0.0002	0.0040***
	(0.0010)	(0.0008)	(0.0006)	(0.0016)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.05	0.03	0.01	0.07
Number of districts	443	443	443	443
Number of Observations	147974	147974	147974	147974
R squared	0.018	0.008	0.008	0.025

The rainfall deviation is the difference between rainfall during the main growing season and the long-term average (1950-2008) in the district, normalized by the standard deviation. The dependent variable in all columns is an indicator for the household having a principal occupation in the sector corresponding to the column name. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table A2: Reduced form effects on sector of occupation during 7 days preceding the survey

	Individual sector of employment:									
	(1) Ag	(2) Manufact.	(3) Const.	(4) Retail	(5) Hospitality	(6) Trans.	(7) Public	(8) Educ.	(9) Service	(10) Other
Rainfall deviation	-0.013***	0.004***	0.002	0.003***	0.000	-0.000	-0.000	0.002***	-0.000	0.001
	(0.003)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.65	0.09	0.06	0.06	0.01	0.03	0.01	0.03	0.02	0.05
Number of districts	410	410	410	410	410	410	410	410	410	410
Number of Observations	350199	350199	350199	350199	350199	350199	350199	350199	350199	350199
R squared	0.083	0.042	0.033	0.014	0.008	0.011	0.006	0.008	0.009	0.025

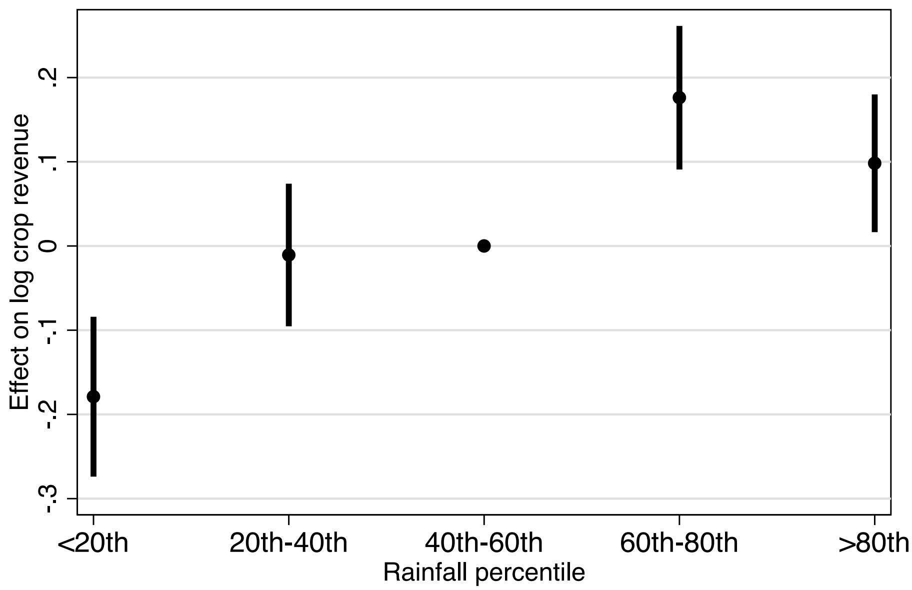
Data are for individuals living in rural households. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable is an indicator variable for the principal industry of employment of the individual during the 7 days preceding the survey. Dependent variables correspond to column titles. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

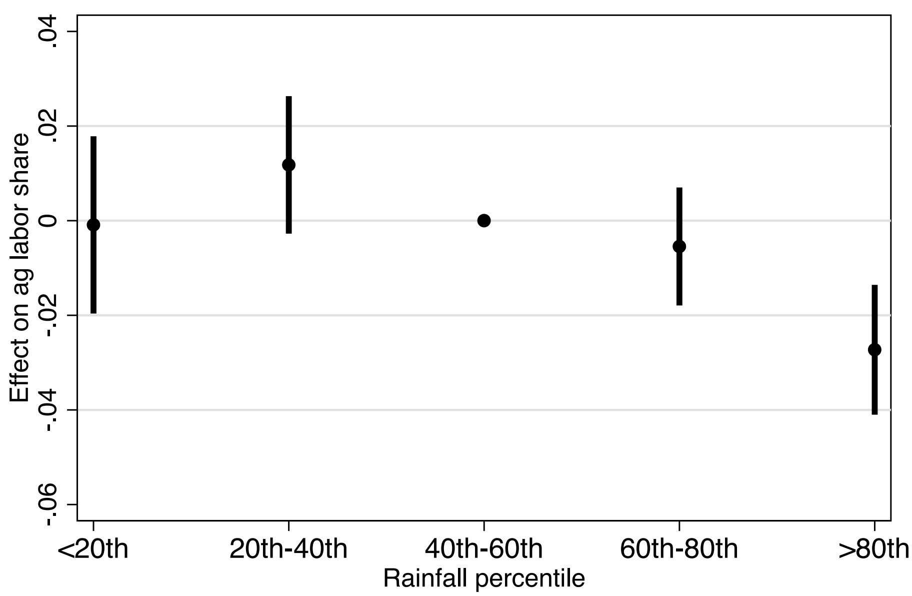
Table A3: Effects of precipitation shocks on district-level GDP

	Log of total GDP in:		
	(1) Agriculture	(2) Non-agriculture	(3) Total
Rainfall deviation	0.052***	0.012***	0.024***
	(0.006)	(0.002)	(0.002)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mean of Dep Variable	11.10	12.16	12.51
Number of districts	402	402	402
Number of Observations	2914	2914	2914
R squared	0.957	0.995	0.993

The rainfall deviation is the difference between rainfall during the main growing season and the long-term average (1950-2008) in the district, normalized by the standard deviation. The dependent variable in all columns is the log of total GDP, where the sector is given by the column label. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, and Kerala. The data do not include West Bengal and Goa.

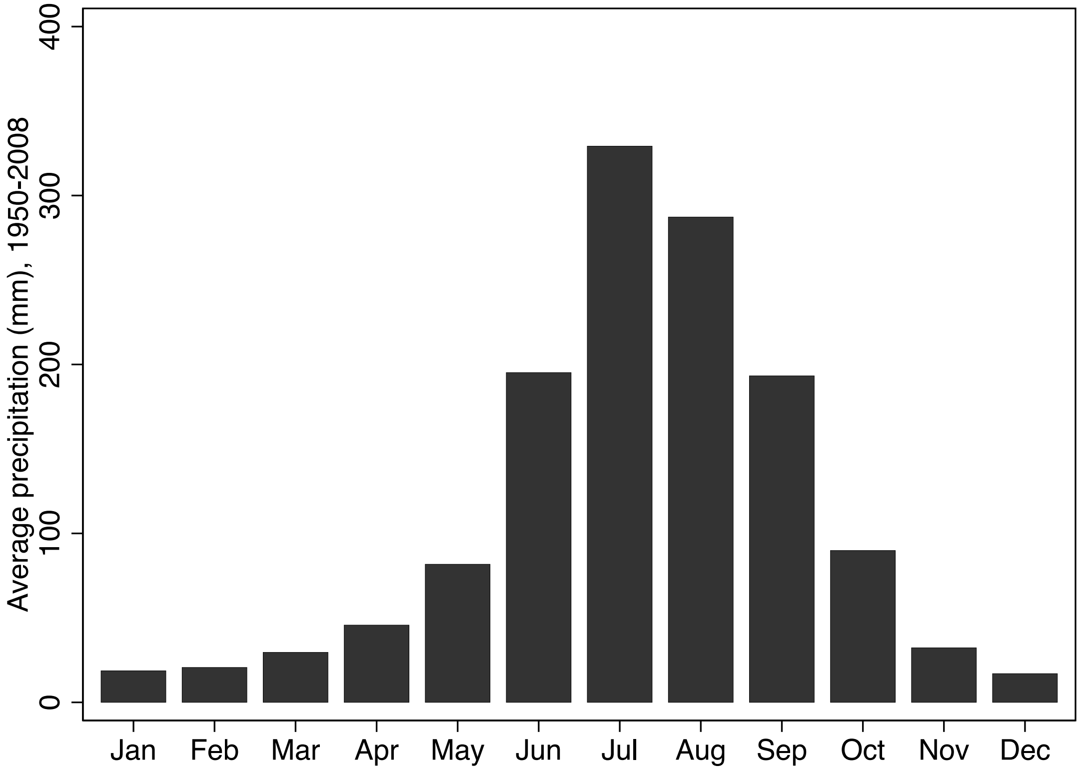
Figure A3: Reduced-form and first-stage effects of positive and negative precipitation shocks





Notes: The top panel shows the coefficients (dots) and 95 percent confidence intervals (bars) for the effects of growing season precipitation on log of crop revenue. Coefficients are effects of precipitation being in that quintile relative to the base group of the 40th-60th percentile. The estimates are based on the same data used to estimate the first-stage regressions in the main text. The bottom panel shows the corresponding reduced-form estimates from NSS. Similarly to the main text, both regressions include district and year fixed effects and standard errors are clustered at the district level.

Figure A4: Monthly average rainfall, 1950-2008



Notes: Graph shows monthly average precipitation in 1950-2008. The data are limited to the sample of districts for which the regressions in the main text are estimated.

Table A4: Robustness to alternative ways of measuring climate shocks

	First stage	Reduced form	IV
A: June Rainfall	0.051***	-0.010**	-0.171*
	(0.015)	(0.004)	(0.095)
B: Discrete Shock (-1, 0, or 1)	0.114***	-0.017***	-0.100**
	(0.023)	(0.005)	(0.039)
C: September-October Temperature	-0.300***	0.033***	-0.109***
	(0.039)	(0.006)	(0.027)

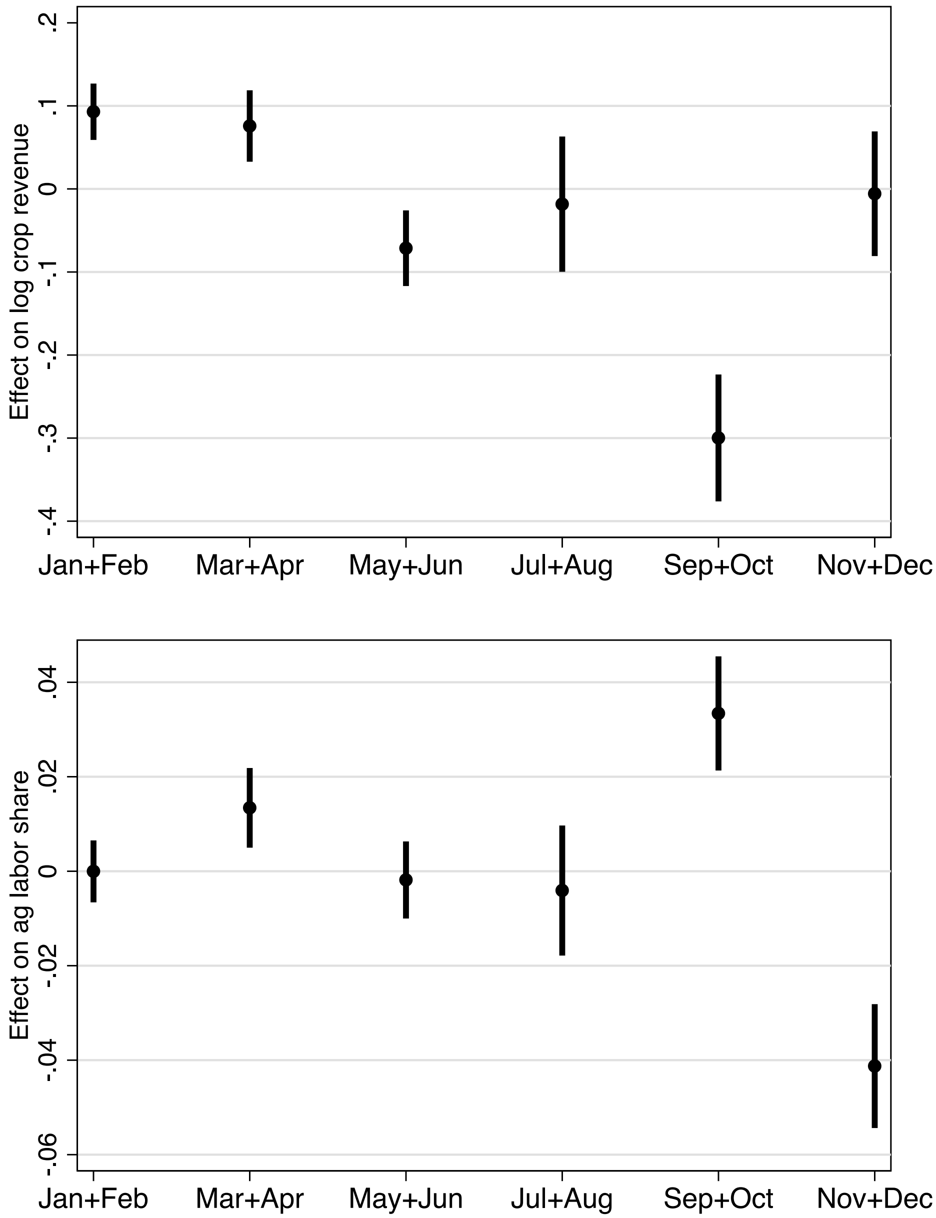
Each row shows the first-stage, reduced-form, and IV (2SLS) estimates when weather shocks are measured as labeled in the first column. More specifically, row A instruments for agricultural productivity using June rainfall — the most frequent month of the onset of the monsoon. The specification controls for monthly rainfall measures in the remaining 11 months of the year. Row B uses a discrete measure that takes on a value of -1 if rainfall during the growing season (June-November) is less than the 20th percentile in the district's distribution from 1950-2008, 0 if it is in between the 20th and 80th percentiles, and 1 if it is greater than the 80th percentile. The regressions in row B control for the z-score of precipitation in the remaining months of the year. Row C instruments for agricultural productivity using average temperature (° C) in September and October, controlling for average temperatures in January/February, March/April, May/June, July/August, and November/December.

Table A5: Effects of rainfall shocks on monthly agricultural prices

	(1) Rice (paddy)	(2) Corn
Rainfall deviation	-0.0179***	-0.0076**
during most recent wet season	(0.0038)	(0.0032)
District FE	Yes	Yes
Year FE	Yes	Yes
Mean of Dep Variable	6.51	6.43
Number of districts	341	272
Number of Observations	10627	8256
R squared	0.531	0.553

The data consist of monthly observations on wholesale agricultural prices at the district level. The dependent variables in column 1 and 2 are the logs of wholesale prices of rice and corn, respectively. The rainfall deviation is the difference between rainfall during the most recent wet season and the long-term average (1950-2008) in the district, normalized by the standard deviation. The most recent wet season is defined as the most recent (complete) period from June to November. Thus, this is June to November of the previous year if the price observation is in the months from January to November and June to November of the current year for December. All standard errors are clustered at the district level. The sample includes all states that are included in the main analysis.

Figure A5: Reduced-form and first-stage effects of intra-annual variation in temperature



Notes: The top panel shows the coefficients (dots) and 95 percent confidence intervals (bars) for the effects of average temperature (degrees Celsius) on log of crop revenue. The estimates are based on the same data used to estimate the first-stage regressions in the main text. The bottom panel shows the corresponding reduced-form estimates from NSS. Similarly to the main text, both regressions include district and year fixed effects and standard errors are clustered at the district level.

Table A6: Reduced-form effects on temporary labor migration

		By age group:	
	(1)	(2)	(3)
	All	18-40	41-70
Rainfall deviation	0.0002	-0.0003	0.0018
	(0.0030)	(0.0040)	(0.0017)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mean of Dep Variable	0.045	0.058	0.022
Number of districts	408	408	408
Number of Observations	299642	188625	111017
R squared	0.048	0.068	0.027

The data are for individuals from 18-70 years old living in rural households in NSS rounds 55 (1999-2000) and 64 (2007-2008). The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable is an indicator if the individual left the household for work for at least 60 days but no more than 6 months during the last year (55th round) or 30 days but no more than 6 months during the last year (64th round). All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table A7: Robustness of main results to district-specific linear time trends

	(1) First Stage	(2) Reduced Form	(3) IV
Rainfall deviation	0.0790**	-0.0186***	
	(0.0313)	(0.0041)	
Log value agricultural output			-0.2061***
			(0.0582)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
District-specific linear trend	Yes	Yes	Yes
Mean of Dep Variable	21.105	0.546	0.531
Number of districts	440	443	438
Number of Observations	1427	214519	177553
R squared	0.962	0.077	0.061

The rainfall deviation is the difference between rainfall during the main growing season and the long-term average in the district, normalized by the standard deviation. The dependent variable in column 1 is the log of the value of output from 6 important wet-season crops (rice, soybeans, millet, maize, groundnut and sugarcane). The dependent variable in columns 2 and 3 is an indicator for households with a principle occupation in the agricultural sector, either as laborers or farmers. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa

Table A8: Robustness of reduced-form estimates to using NSS sampling weights

	Reduced-Form Estimates	
	(1)	(2)
Rainfall deviation	-0.016***	-0.018***
	(0.003)	(0.004)
Region by year FE	No	Yes
Mean of Dep Variable	0.62	0.62
Number of districts	443	443
Number of Observations	214519	214519
R squared	0.073	0.074

Both regressions are weighted by NSS sampling weights. The rainfall deviation is the difference between rainfall during the main growing season and the long-term average (1950-2008) in the district, normalized by the standard deviation. The dependent variable in Panel B is an indicator for households with a principle occupation in the agricultural sector, either as laborers or farmers. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Table A9: Robustness of main effects to using TRMM satellite rainfall data

	Panel A: First-Stage Estimates	
	(1)	(2)
Rainfall deviation	0.043***	0.032**
	(0.013)	(0.014)
Region by year FE	No	Yes
Mean of Dep Variable	21.11	21.11
Number of districts	440	440
Number of Observations	1427	1427
R squared	0.573	0.645
	Panel B: Reduced-Form Estimates	
	(1)	(2)
Rainfall deviation	-0.008***	-0.011***
	(0.003)	(0.003)
Region by year FE	No	Yes
Mean of Dep Variable	0.55	0.55
Number of districts	443	443
Number of Observations	214519	214519
R squared	0.070	0.071
	Panel C: IV Estimates	
	(1)	(2)
Log value	-0.166**	-0.323***
agriculutral output	(0.067)	(0.125)
Region by year FE	No	Yes
Mean of Dep Variable	0.53	0.53
Number of districts	438	438
Number of Observations	177553	177553
R squared	0.049	0.021

The rainfall deviation is the difference between rainfall during the main growing season and the long-term average (1998-2013) in the district, normalized by the standard deviation. Climate data are TRMM satellite data from the National Oceanic and Atmospheric Administration. The dependent variable in Panel A is the log of the value of output from 6 important wet-season crops (rice, soybeans, millet, maize, groundnut and sugarcane). The dependent variable in Panel B is an indicator for households with a principle occupation in the agricultural sector, either as laborers or farmers. Panel C contains instrumental variable estimates where the dependent variable is the indicator for households in the agricultural sector and the value of agricultural output is instrumented with the rainfall deviation. All regressions include the rainfall deviation during the dry season (December-May) as a control. All standard errors are clustered at the district level. The states included in the analysis are Bihar, Jharkhand, Orissa, West Bengal, Chhattisgarh, Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Assam, Punjab, Haryana, Rajasthan, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Kerala, and Goa.

Appendix B: Theoretical Model

In this appendix I outline a simple general equilibrium model that shows how exogenous increases in agricultural productivity can affect prices and the allocation of labor across sectors. I set up the most simple model that can still highlight the conditions that are necessary for the agricultural labor share to be decreasing in agricultural productivity.

There are two sectors: agriculture (denoted as A) and non-agriculture (denoted as N). Labor is the only factor in the economy and output of the agricultural sector is given by $Y_A = \theta L_A^\beta$, where L_A is labor in agriculture, $\beta < 1$ is the output elasticity of labor, and θ is the agricultural TFP parameter. Output of the non-agricultural sector is simply $Y_N = L_N$, where L_N is the amount of labor allocated to the non-agricultural sector. The total labor endowment of the economy is $L = L_A + L_N$. The (endogenous) price of the non-agricultural good is p_N and agriculture is assumed to be the numeraire so that $p_A = 1$.

There is a representative consumer that derives utility from consuming the two goods according to the utility function $u(c_A, c_N)$. The consumer maximizes this utility subject to the budget constraint that the total value of consumption be equal to total income in the economy which is $p_N L_N + \theta L_A^\beta$.

The competitive equilibrium is defined as the price and labor allocation such that the consumer maximizes utility and at the same time labor is allocated efficiently across sectors. Since labor is mobile across sectors, its' allocation is pinned down by the straightforward condition that values of marginal products are equalized across sectors. This condition collapses to

$$L_A = \left(\frac{p_N(\theta)}{\beta\theta} \right)^{\frac{1}{\beta-1}}. \tag{B1}$$

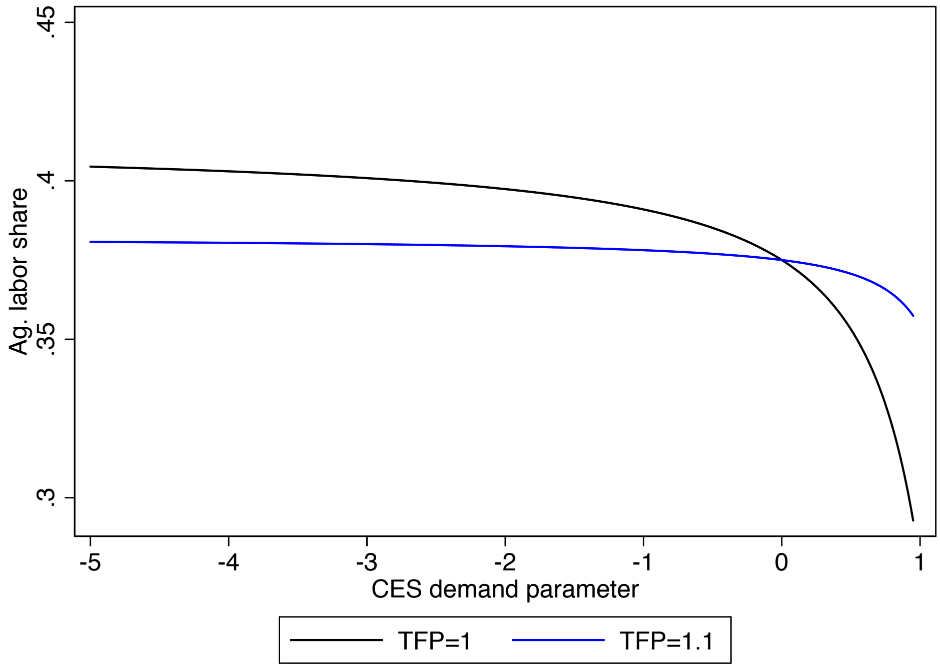
Equation B1 highlights the demand channel discussed in the main text. In order for agricultural labor to be decreasing in agricultural TFP it is necessary that the price of the non-agricultural good rises more than proportionately with θ. Put differently, the sign of $\frac{\partial L_N}{\partial \theta}$ can be positive due to an income effect where an increase in agricultural TFP increases income and this in turn increases the demand for the non-agricultural good and thus the non-agricultural labor share. This income effect is increasing in the income elasticity of the non-agricultural good and the share of income that comes from agriculture.

The price of the non-agricultural good is such that supply equals demand from the representative consumer. The sign of $\frac{\partial L_N}{\partial \theta}$ will therefore depend on consumer preferences over the two goods. As one example, if preferences are CES, i.e. $u(c_A, c_N) = (c_A^\rho + c_N^\rho)^{\frac{1}{\rho}}$, then the agricultural labor share will be decreasing in θ as long as the two goods are complementary (see Figure B1 for a numerical example). As ρ becomes more negative — and thus the goods become more complementary — the increase in the non-agricultural labor share is larger for a given increase in agricultural TFP. If preferences are instead Cobb Douglas, then the non-agricultural price rises at the same rate as θ leading to independence between agricultural productivity and the non-agricultural labor share.[26] As a final example, non-homothetic preferences such as a Stone-Geary utility function with $u(c_A, c_N) = ln(c_A - \bar{a}) + ln(c_N - \bar{n})$ will deliver the prediction that $\frac{\partial L_N}{\partial \theta} > 0$ if $\bar{a} > 0$ and $\bar{n} < 0$. In this case the income elasticity of the non-agricultural good will be greater than one, leading to more resources shifting to the non-agricultural sector when agricultural productivity increases and thus income grows. In this case the effect of agricultural productivity on the non-agricultural labor share increases as $\bar{n}$ becomes more negative, i.e. as the income elasticity of the non-agricultural good increases. This particular form of preferences is commonly used in the theoretical literature on structural transformation (Kongsamut, Rebelo, and Xie, 2001; Gollin, Parente, and Rogerson, 2002) and is approximately consistent with structural transformation in the U.S (Herrendorf, Rogerson, and Valentinyi, 2013).

In sum, the simple framework points to two features that can generate an increasing relationship between agricultural productivity and the labor share in the non-agricultural sector: strong complementarity in consumption between the agricultural and non-agricultural good and preference structures with relatively steep Engel curves for non-agricultural goods.

[26]Foster and Rosenzweig (2007) show this for a similar model.

Figure B1: Effect of a 10 percent increase in agricultural TFP with CES preferences



Notes: Figure shows the effect of an exogenous 10% increase in agricultural TFP on the equilibrium agricultural labor share. The 10% increase in TFP is the movement from the black curve to the blue curve. The utility function of the representative consumer is $u(c_A, c_N) = (c_A^\rho + c_N^\rho)^{\frac{1}{\rho}}$. The value of ρ varies along the horizontal axis of the graph. The agricultural production function is $Y_A = \theta L_A^{0.6}$ and the non-agricultural production function is $Y_N = L_N$.