Complementarity between non-agricultural and agricultural shocks in rural industrialization: Evidence from China*

Jessica Leight[1]

Williams College

October 19, 2016

Abstract

This paper analyzes patterns of structural transformation in China between 2000 and 2010, seeking to estimate the impact of shocks to labor demand in the secondary (industrial and mining) sector on local economic outcomes, and analyze whether there is any evidence of complementarity between these shocks and agricultural productivity shocks, proxied by county-level rainfall. I employ a newly assembled panel including a nationwide sample of 2000 counties, and construct secondary labor demand shocks following Bartik (1991), using the baseline composition of county employment and national employment fluctuations by subsector. The empirical results indicate that first, there is a robust response to secondary labor demand shocks in terms of increased employment, GDP and value added in the secondary sector, and increases in total GDP. Second, there is evidence of significant complementarity between these shocks and agricultural shocks, but this pattern is restricted to counties that are less industrialized in baseline. In these counties, non-agricultural growth is observed only following positive shocks to both the non-agricultural and agricultural sectors. Further exploration suggests this pattern may be driven by capital constraints in heavily agricultural regions.

1 Introduction

The Chinese economy has in recent years been characterized by extremely rapid growth of non-agricultural production even in rural areas. Micro data shows that the share of rural income constituted by agricultural income fell by half between 1987 and 2001, from 40 percent to 20 percent (Benjamin et al., 2008). At the macro level, agriculture's

*Preliminary and not for citation. Please do not circulate. Michael Neubauer, Jillian Stallman, and Jane Zhao provided excellent research assistance. Thanks also to Loren Brandt, Valerie Mueller, Katrina Kosec and seminar audiences at Williams, University of Massachusetts - Amherst, Northeastern, the World Bank Research Group, and Colgate for detailed comments.

share of GDP declined from 28 to 16 percent between 1985 and 2000, while its share in employment declined from 62 to 50 percent (Huang et al., 2008). Needless to say, this transformation has coincided with a period of extremely rapid overall growth, and evidence suggests that increasing non-agricultural productivity has been a key driver of this process (Brandt and Zhu, 2010).

What is the role of agriculture in the transformation of the Chinese economy? The question of the relationship between agricultural and non-agricultural growth in developing countries given the presence of large and unproductive agricultural sectors has long been a focus of academic and policy debate. In the early post-war period, it was widely argued that increases in agricultural productivity could stimulate overall growth (Johnston and Mellor, 1961), and recent work has presented evidence that agricultural expansion has significant ancillary effects in boosting aggregate growth and reducing poverty.[1]

In the context of China, structural estimation using macro data suggests that at least in the earlier reform period, growth in agricultural productivity has been rapid and has contributed substantially to overall productivity growth, both directly and by stimulating substitution of factors out of agriculture (Brandt and Zhu, 2010; Zhu, 2012). However, relatively little is known about within-country, local-level heterogeneity in the relationship between agricultural and non-agricultural growth. More generally, the literature on cross-sector spillovers in China as well as other developing economies continues to be constrained by the absence of plausibly exogenous sectoral shocks.

This paper analyzes patterns of structural transformation in China using a newly assembled panel of county-level data from approximately 2000 counties observed between 2000 and 2010. More specifically, I seek to answer two questions. First, I estimate the impact of shocks to labor demand in the secondary (industrial and mining) sector on local economic outcomes, including employment, investment, and output in agricultural and non-agricultural production. Second, I analyze whether there is any evidence of complementarity between shocks to the secondary sector and agricultural productivity shocks, where the latter are proxied by rainfall at the county level.

Following a large literature based on Bartik (1991), I construct a variable capturing secondary labor demand shocks by employing the baseline composition of the county secondary sector and data on national fluctuations in subsector employment over time. I will demonstrate that the use of these shocks is robust to various endogeneity concerns; there is little evidence that national-level employment fluctuations are driven by local trends, and also little evidence that counties experiencing more positive shocks in this period show evidence of differential trends prior to widespread rural industrialization.

I then analyze the response of local economic outcomes to secondary labor demand shocks, as well as the interaction of these shocks and rainfall shocks. In addition, I test whether there is evidence of heterogeneity comparing across counties that are initially characterized by limited industrialization and a primary dependence on agriculture, and counties that already have sizeable non-agricultural sectors. China provides a uniquely favorable setting for this analysis given the wide within-country diversity in patterns of industrialization.

Agricultural productivity shocks could affect the local response to secondary labor demand shocks via three main channels. First, a positive shock to agricultural produc-

[1] Gollin et al. (2014) provide a useful overview of this debate, though focusing primarily on sub-Saharan Africa.

tivity will increase the returns to labor and capital in agriculture, and this will render counties less responsive to subsequent positive shocks to the returns to factors in the non-agricultural sector; I denote this the factor demand channel. Second, increased rainfall will increase local income and may increase demand for secondary output; I denote this the local demand channel. Third, if counties have limited access to external capital, they may be more responsive to positive shocks to the non-agricultural sector only after a boom in agriculture that increases local income and local savings. I denote this third channel the capital channel. All three channels have been identified in other recent papers analyzing structural transformation in developing countries, particularly Marden (2015), Santangelo (2016), and Colmer (2016).

The empirical results indicate that first, there is a robust response to secondary labor demand shocks on average. A positive shock generates an increase in secondary employment, as well as gross domestic product and value added in the secondary sector. There is also evidence of decreased investment and a decline in output in agriculture. Perhaps most important, a one standard deviation increase in predicted secondary employment leads to a 43% percent increase in GDP per capita.[2]

In addition, I find that there is significant complementarity between positive shocks in the agricultural and non-agricultural sectors, but only in counties that are initially less industrialized; less industrialized counties are identified as those in the lowest quantile of non-primary employment as a proportion of total employment at baseline. Consider two counties characterized by initially low levels of industrialization that are subject to shocks to secondary labor demand of similar magnitude; one county also experiences a one standard deviation increase in rainfall prior to this positive shock. The results suggest that increases in secondary and total employment, and secondary and total GDP, are observed only in the county that recently experienced a boom in agriculture. In the less industrialized county in which there was no recent positive shock in agriculture, the effects of the secondary labor demand shock are minimal.

In counties that are initially more industrialized, the evidence is very different. The response of these counties to a secondary labor demand shock at the mean level of rainfall is generally larger in magnitude, compared to the response of counties that are not highly industrialized. However, there is no evidence that this response varies systematically with respect to agricultural productivity shocks, perhaps unsurprising given that the direct effect of rainfall shocks in more industrialized counties is also small in magnitude.

Further exploration of the potential channels for these results suggests that the complementarity of agricultural and non-agricultural shocks in less industrialized counties is not driven by local demand effects; rather, it may reflect capital constraints. Given that rural counties have limited access to external credit to fund investment in new productive sectors, the primary source of capital is re-investment from agriculture. Accordingly, there are potentially significant complementarities between non-agricultural and agricultural shocks. For counties that are already industrialized, however, there is no significant interaction between the two sets of shocks.

This paper fits into several related literatures. First, there is a substantial and growing literature analyzing the relationship between shocks to agricultural production, whether

[2]As described in more detail in the section presenting the empirical results, the effect magnitudes are benchmarked using the within-prefecture standard deviation. This is employed as a plausible proxy for local shocks.

technological or climatic, and transitions from agriculture to industry. Foster and Rosenzweig (2004) estimate the impact of shocks to the returns to agriculture in India induced by the adoption of Green Revolution technology, and find that industrial growth is fastest in areas where agricultural growth is lagging. Jedwab (2011) finds that positive price shocks to cocoa crops during cocoa booms in both Ghana and the Ivory Coast lead to an increase in urbanization. Hornbeck and Keskin (2015) argue there is no evidence that positive agricultural growth generated by the construction of an aquifer in the U.S. stimulated growth in non-agricultural sectors, while Bustos et al. (2015) present evidence that the introduction of genetically engineered soybean seed in Brazil led to industrial growth only when the technological change was labor-saving.

Also relevant is recent work employing both micro- and macro-level data that has documented that productivity is much lower in the agricultural sectors of developing economies compared to the non-agricultural sectors in the same economies – and that by extension, the gap in productivity comparing agricultural sectors across less developed and developed countries is much larger than the corresponding gap when comparing non-agricultural sectors (Gollin et al., 2014; Lagakos and Waugh, 2013). This evidence has raised the question of whether market frictions of various forms may prevent substitution of productive factors out of agriculture into non-agricultural production, generating the large observed productivity gaps. While the China literature suggests that these productivity gaps may not be as large there as in other developing countries, the relationship between agricultural shocks and structural transformation is nonetheless informative given the broader debate about constraints limiting exit from agriculture.

Second, a number of papers have explored whether there are spillover effects of agricultural growth in other sectors, or substantial effects of agricultural growth in poverty reduction. Gemmell et al. (2000) and Tiffin and Irz (2006) examine this question using time series methods. Ravallion and Chen (2007) present evidence that in China, growth in agriculture has a much more significant impact on poverty reduction, compared to growth in other sectors. Ligon and Sadoulet (2008) and Christiansen et al. (2011) provide similar evidence for broader samples, using cross-country data. This literature has generally focused on analyzing cross-country data, and has not exploited local shocks to agricultural or non-agricultural production.

Third, there is also a substantial literature analyzing the relationship between climatic shocks to temperature and rainfall and economic growth, including growth in the industrial sector, using cross-country or macro-level data. Dell et al. (2014) provides a useful overview of the literature on climatic shocks. Dell et al. (2012) find a negative relationship between temperature and industrial value added in a global country-level sample, though only in poor countries, and Hsiang (2010) finds a similar effect analyzing only data from Central and South America. Jones and Olken (2010) find evidence of a decline in industrial imports from poor countries following an increase in temperature. Most recently, Santangelo (2016) and Colmer (2016) both analyze the relationship between climatic shocks and industrialization at the local level in India. Santangelo (2016) suggests a significant negative impact of local agricultural shocks on local industry in India, consistent with a negative local demand effect, while Colmer (2016) concludes that negative weather shocks result in enhanced growth in local manufacturing productivity.

Fourth, this paper connects to a smaller literature analyzing rural industrialization in China, a set of papers that is largely descriptive. Mukherjee and Zhang (2007) argue that rapid rural industrial growth in China primarily reflects a favorable institutional structure. de Janvry et al. (2005) argue that non-farm employment has generated significant increases in rural income, but also significant increases in rural inequality; Benjamin et al. (2005) also present evidence that rising non-farm incomes drive increasing inequality. Marden (2015) provides evidence that agricultural growth in the early reform period is associated with significantly faster growth in non-agricultural output.

Relative to the existing literature, this paper makes several contributions. It is the first paper to construct county-level Bartik shocks in China. It is also the first paper to my knowledge to examine not only the direct response of non-agricultural economic outcomes to productivity shocks in agriculture in a developing country context, but also interactions between local non-agricultural and agricultural shocks.

The remainder of the paper proceeds as follows. Section 2 outlines a simple conceptual framework and provides an overview of the data. Section 3 describes the empirical strategy. Sections 4 and 5 present the primary empirical results and robustness checks, and Section 6 explores potential channels for the observed effect. Section 7 concludes.

2 Background

2.1 Conceptual framework

In this section, I will briefly describe a simple conceptual framework in order to generate predictions to guide the empirical analysis. The local economy – in this case, a county – is assumed to encompass two productive sectors, agricultural and non-agricultural production. There are two factors of production, capital and labor, and output in each sector is described by the function $Y_j = p_j \theta_j L_j^{\alpha} K_j^{\beta}$, where θ_j denotes a sector-specific productivity parameter.

The empirical analysis will explore the relationship between positive, temporary shocks to agriculture in the form of rainfall shocks — conceptualized as increases in θ_A — and more permanent positive shocks to the secondary sector, captured by Bartik shocks. The latter could correspond to positive shocks to productivity θ_I or price p_I; for concision, I will assume these shocks are positive shocks to θ_I, though the subsequent analysis would also apply to price shocks. The primary focus in the empirical analysis will be the sign of the cross-partial derivative $\frac{\partial Y_I}{\partial \theta_A \partial \theta_I}$: i.e., I wish to identify whether non-agricultural and agricultural shocks are complements or substitutes in generating expansion of the non-agricultural sector.

I hypothesize that there are three channels through which rainfall shocks may affect the subsequent response of local economic outcomes to positive shocks to the secondary sector. First, an increase in rainfall will increase the local return to factors in the agricultural sector, at least in the short term, and thus slow the substitution of productive factors into the secondary sector. I denote this channel the factor reallocation channel.

Second, an increase in rainfall will increase local income. If markets are not fully integrated and local demand is important, then a positive income shock will result in an increase in prices p_I, and this would be manifest in a direct positive effect on the secondary and tertiary sectors. This channel may not lead to any change in the slope of the response function $\frac{\partial Y_I}{\partial \theta_I}$. However, firms are only responsive to positive shocks when both prices and productivity are increasing, it may lead to a more robust response to secondary labor demand shocks. I denote this channel the local demand channel.

Third, an increase in rainfall may also increase local capital available for investment, via either public or private channels. If counties are constrained in making some minimum indivisible investment required to substantially expand non-agricultural production, then positive rainfall may relax this constraint and enable them to enter non-agricultural production for the first time. I will denote this channel the capital channel. The objective of the empirical analysis is to identify the channel that dominates.

2.2 Data

This paper employs data from four primary sources: the 2000 county-level census in China, annual data on secondary employment compiled by the national statistical bureau, annual provincial statistical yearbooks that report economic outcomes at a county level, and monthly climatic data from a network of climate stations. (In the robustness checks, I will also employ separate data on exports by disaggregated secondary subsector; this data is described in Section 5.) I will discuss each data source in turn.

First, the 2000 county-level census provides an overview of demographic and employment information in each county. The decennial censuses in China collect basic information from a full sample of Chinese households, and more detailed information (known as the long-form survey) from a 9.5% sample. Households are surveyed in their place of residence if they have their permanent registration at that location, or if their permanent registration is elsewhere but they have resided in that location for more than six months. Among other indicators, the census provides information about total employment (and employment by gender), as well as employment in a number of disaggregated subsectors.

I employ the county census to generate estimates of total employment by sector at the county level at the beginning of the period of interest. Primary employment is defined as the sum of employment in farming, forestry, animal husbandry, fishery and agricultural services. Secondary employment is the sum of employment in the 34 subsectors specified in industrial production and mining; the share of initial employment in each subsector will be employed to construct the labor demand shocks, as described in more detail in Section 3.1.[3] Tertiary employment is the sum of employment reported in a number of disaggregated service sectors.[4]

[3]The secondary sectors are coal mining and dressing, extraction of petroleum and national gas, mining and dressing of ferrous metals, mining and dressing of nonferrous metals, mining and dressing of nonmetal minerals, mining and dressing of other minerals, logging and transport of wood and bamboo, food processing, food production, beverages, tobacco, textiles, garments and other fiber products, leather and related products, timber processing, furniture manufacturing, paper-making, printing, cultural and sports goods, petroleum processing and coking, raw chemical materials and chemical products, medical and pharmaceutical products, chemical fiber, rubber products, plastic products, nonmetal mineral products, smelting and pressing of ferrous metals, metal products, ordinary machinery, specialized equipment, transport equipment, weapons and ammunitions, electric equipment and machinery, electronic and telecommunications equipment, instruments and office machinery, other manufacturing, production of electric power and hot water, production of gas, and production of tap water.

[4]This includes employment in construction, geological prospecting, transportation and telecommunication services, wholesale and retail trade and catering services, finance and insurance, real estate, social services, public services, hotels, tourism and recreational services, information services, health care, sports and social welfare, education, culture and the arts, scientific research and polytechnic services,

Additional information reported in the census includes detailed county demographics, including population with different residency status; population by gender, age, ethnicity, and educational categories; demographic information about births and deaths; and some information about population mobility and the housing stock.

The second source of data is data on secondary employment provided by the All-China Data Center. This data includes annual data on disaggregated employment between 1999 and 2011 in 43 specific subsectors.[5] These categories are matched to the country categories in order to generate measures of annual growth in each secondary subsector. It is important to note that annual growth is not reported at the county level; the national statistical yearbooks report disaggregated employment only nationwide and by province.

The third source of data employed is annual data on county economic indicators that is reported in provincial economic yearbooks. Each year, every province in China publishes a statistical yearbook. In general, the statistical yearbook primarily reports economic indicators for the full province or for larger aggregate units such as prefectures. However, the majority of provincial yearbooks also include some economic indicators reported at the county level. This data was compiled and digitized for each county available in all Chinese provinces (excluding the autonomous regions and provincial-level cities) for every year available between approximately 1996 and 2011. (Each yearbook reports data from the prior year; thus the final year observed in the data is 2010.) This is the first time, to my knowledge, a comprehensive panel of this size has been assembled. More detail about the variables of interest observed in the county yearbooks is provided in Section 3.4.

The fourth source of data is climatic data on rainfall from the Asia-Pacific Data-Research Center. The APDRC provides monthly rainfall data measured in millimeters for a 2.5 degree grid across Asia between 1990 and 2011.[6]

3 Empirical strategy

3.1 Constructing shocks to secondary labor demand

The primary analysis will construct variables proxying for shocks to secondary labor demand following a methodology that has been widely used in the labor literature since Bartik (1991). The intuition is to generate a predicted shock to labor demand in the secondary sector using the composition of employment in a baseline year (here 2000) and growth per subsector as reported at the national level. The choice of the base year is driven by the fact that disaggregated secondary employment composition at the county level was reported in the census only in 1990 and 2000, and annual data on national employment at the subsector level is reported only post-1998.

The objective of using this instrument is to construct a variable capturing shocks to the secondary sector that is uncorrelated with shocks to the local, county-level economy. Given that there are more than 2000 counties in China, it is implausible that even shocks

and government and party organizations.

[5] In some cases, an industry that was reported as one category in the county data reports employment in two separate categories in the national data. In addition, employment is reported in four categories in the national data that is not reported in the county data: waste processing and three residual types of mining employment.

[6] More specifically, I am employing the GPCC data, Version 6.

to a relatively larger county would affect the national labor market; this assumption will be further evaluated in the robustness checks reported in Section 5.

More specifically, I construct the predicted growth rate of secondary employment G_{ipt} for county i in province p in year t as follows.

$$G_{ipt} = \sum_j EMP_{ijb} \times \frac{N_{j,t} - N_{j,t-1}}{N_{j,t-1}} \tag{1}$$

EMP_{ijb} is the employment share in county i of industry subsector j in the base year (2000), and $N_{j,t}$ is the national employment share of industry j in year t. For each year, this is a predicted growth rate for employment in county i from period t-1 to t, holding the industrial composition fixed at baseline and assuming each county sector grows at the same rate observed for nationwide employment in that sector.

I then employ these predicted growth rates to construct predicted secondary employment for each county in each year, employing reported total secondary employment in the county census in 2000 and the predicted growth rates.[7] More specifically, I estimate predicted employment for county i in year t as follows; $Base_{ip}$ denotes total secondary employment for the county of interest as reported in the 2000 census.

$$Pred_{ipt} = (\prod_{j=2001}^{t} G_{ipt})\ Base_{ip} \tag{2}$$

Similar empirical strategies have been employed by a number of other authors using U.S. data, including Autor and Duggan (2002), Blanchard et al. (1992), Bound and Holzer (2000), and Luttmer (2005). In the China literature, Bartik-style instruments have only been employed for analyses of large urban areas.[8] This is the first paper to employ a Bartik instrument for secondary labor demand using a nationwide panel of Chinese county-level data.

While the choice of the base year is driven primarily by data availability as previously noted, it is also helpful to verify that the observations for employment and GDP in the year 2000 are not dramatically different relative to the trends generally observed over this period. Figure 1 in the Appendix shows the trends from 1990 to 2010 for secondary employment, total employment, secondary GDP, and total GDP for all of China; in all cases, rapid growth is observed over this period. For secondary employment, there is some evidence of a slowdown between 1997 and 2002, followed again by very rapid growth. However, there is no evidence that the year 2000 in particular corresponds to an outlier or a trend break for any of these variables, suggesting that no systematic bias arises from its use as a base year.

3.2 Constructing agricultural shocks

Using the APDRC data for the 2.5 degree grid, I interpolate between the grid points at which rainfall is reported using the inverse distance weighting method and the coordinates

[7] While some provinces and years also report secondary employment in the provincial yearbooks, not all counties report secondary employment in 2000. Accordingly, I utilize the reported employment from the county census in order to ensure consistency and maximize the size of the sample.

[8] Relevant papers include Deng et al. (2001), Jenq and Yi (2013), Li et al. (2011), and Liu and Shang (2012).

of the country centroid. Only grid points within 200 kilometers of the county centroid are employed to construct rainfall; the average county measure employs data from three grid points.

I then construct a measure of mean rainfall in the cultivation period for the primary grain crop, rice or wheat; the cultivation seasons are defined and the primary crops are identified following Tao et al. (2008).[9] I will demonstrate that rainfall in this period is correlated with increased agricultural output. Altering this measure to include rainfall from all twelve months does not significantly increase the predictive power.

3.3 Sample and outcome variables

The primary question of interest in this analysis is the relationship between shocks to the non-agricultural and agricultural sectors, and thus I wish to restrict the sample to exclude county-level units that are already fully urbanized. Accordingly, I exclude the four large provincial-level cities: Beijing, Tianjin, Shanghai, and Chongqing. I also exclude provincial-level autonomous regions: Tibet, Xinjiang, Ningxia, Inner Mongolia, and Guangxi, as well as the island of Hainan. Full data is normally not available for these regions, and they are characterized by very different patterns of economic development. Otherwise, all counties where the name and code of the county can be matched between the county census and the provincial yearbooks are included.

More details about the process of matching counties between the census and the provincial yearbooks is provided in Appendix B. Overall, 87% of counties reported in the 2000 census are matched to provincial yearbook data. Counties that are missing in the provincial yearbooks are disproportionately counties that are part of the urbanized areas of larger, prefecture-level cities, as some provinces do not report county-level data for these units in the yearbook. Accordingly, any bias due to missing counties will orient the sample toward rural areas that are not already fully industrialized.

I will also summarize the variables of interest here. It is important to note that the sample size available varies for different outcomes, as many are not reported in some provinces or years. Tables A12 through A14 in the Appendix summarize the provinces and years in which each variable is observed.

The first set of outcome variables includes employment as reported in the primary, secondary, and tertiary sectors, as well as total employment. Primary employment includes employment in agriculture and related economic activities (husbandry and fishing), secondary employment is employment in industry and mining, and tertiary employment is employment in services. Employment is reported in thousands of persons. The sample size is far smaller for employment than for other indicators of interest.

The second set of outcomes is gross domestic product at the county level, as well as GDP in the primary, secondary, and tertiary sectors, and per capita GDP. All measures are reported in millions of yuan with the exception of per capita GDP, reported in yuan. Provincial yearbooks generally report GDP in nominal terms, and accordingly, these variables are deflated using the World Bank deflator.

[9] Provinces in northern and eastern China are designated as primarily wheat-producing; provinces in southern China are designated as primarily rice-producing. The cultivation season for wheat is designated as October to April; the key cultivation season for rice is designated as April to October.

The third set of outcomes is investment in agriculture. Here, variables reported include cultivated area, reported in thousands of hectares; and output of grain and cash crops, reported in thousands of tons. Production of cash crops is the sum of production of meat and edible oils, the most commonly reported cash crops.

Finally, the fourth set of outcome variables includes value added as reported in the primary and secondary sectors; value added in the tertiary sector is reported only for a very small sample. Value added is again reported in millions of yuan and deflated to generate real values.

3.4 Summary statistics

The core dataset includes data on 2062 counties from 21 provinces. The median county is observed in the panel for 12 years, including 2000; Table 1 reports summary statistics. The sampled counties have a mean population of around 480,000. Average employment is around 248,000 people, of which 54% is constituted by employment in the primary sector, 21% by the secondary sector, and 28% by the tertiary sector.

GDP per capita is over 14,000 yuan, or around $1700. The highest share of GDP is constituted by the secondary sector (41%), followed by the tertiary and primary sectors. The average county reports around 58,000 hectares of cultivated area, and produces 220,000 tons of grain and around 50,000 tons of cash crops.

Figure 1 presents some graphical evidence on trends in GDP and employment, including GDP and employment by sector, over the period of interest. Each graph shows the median of the indicator of interest in the specified year. To avoid bias due to entry and exit of counties, the sample included in the figures is restricted to counties that report the indicator of interest in at least eight of the ten years.

We can observe in Figure 1a that GDP nearly triples over this period, and similarly for GDP per capita. Figure 1b reports GDP by sector; growth is clearly most rapid in the secondary sector, where GDP increases by a factor of five, followed by tertiary and primary growth. I also report a simple measure of labor productivity, value added per worker by sector, in Figure 1c. Productivity is higher and increasing more rapidly in the secondary sector, but also increasing gradually in the primary sector; value added in the tertiary sector is reported only for a very small sample.

Figure 1d reports the evolution of employment. A gradual decline over time in employment in the primary sector is observed. By contrast, both secondary and tertiary employment double. Figure 1e illustrates that sown area is roughly constant in absolute terms, and approximately two thirds of sown area is devoted to grain cultivation, and the remainder to cash crops. Finally, Figure 1f shows growth over time in crop production, which does not exhibit dramatic fluctuations. Production of grain is increasing, and still dwarves the production of cash crops.

Overall, the evidence suggests that the counties observed in this sample are experiencing dramatic growth, primarily driven by growth in non-agricultural sectors. While there is no evidence that agriculture is shrinking – primary GDP is growing and investment in agriculture is roughly constant – labor engaged in the primary sector is declining.

Figure 2 shows a map of the sampled counties; the lightest color denotes regions that are outside the sample. Counties included in the sample are shaded according to the quantile of initial non-primary employment, as reported in the 2000 census. This provides an overview of the salience of non-agricultural production by province at baseline. Unsurprisingly, the counties with the lowest levels of non-primary employment are concentrated in central and western China, while the highest levels are observed in eastern and southeastern China.[10] This variation in the initial level of industrialization will also be important in the main analysis.

4 Empirical results

4.1 Non-agricultural shocks

First, I analyze the response of county-level economic outcomes to secondary labor demand shocks, proxied by predicted employment as constructed above, and denoted $Pred_{ifpt}$ for county in prefecture f in province p in year t. In addition to province-year fixed effects and county fixed effects, I include differential trends for counties in different quantiles of the initial absolute level of secondary employment, λ_{sec}, and interact the province-year fixed effects and the secondary quantile-trends with a dummy variable for counties defined to be industrialized at baseline. This variable I_{ifp} is defined using information about the fraction of the population engaged in non-primary employment as reported in the 2000 census; more details are provided in subsection 4.2.[11]

This yields the following specification.

$$Y_{ifpt} = \beta_1 Pred_{ifpt} + \sum_{i=1}^{4} \lambda_{sec}^{i} \times t + \gamma_{pt} + I_{ifp} \times (\sum_{i=1}^{4} \lambda_{sec}^{i} \times t + \gamma_{pt}) + \nu_i + \epsilon_{ifpt} \quad (3)$$

Standard errors are estimated employing two-way clustering at the county and year level. All the dependent variables, as well as the predicted employment shocks, have the top and bottom 3% of observations trimmed to eliminate outliers.

The results of estimating equation (3) are reported in Table 2. Again, throughout this discussion, the primary sector refers to agriculture and related activities (husbandry and fishing); the secondary sector refers to mining and industrial production; and the tertiary sector refers to services.

In Panel A, we first observe a negative response of primary employment to predicted secondary labor demand shocks in Column (1), and, unsurprisingly, a significant and positive response of secondary employment to the same shocks in Column (2). The hypothesis that the coefficient β_1 is equal to one in the specification employing secondary employment cannot be rejected: i.e., employment appears to increase one for one with predicted employment based on observed national growth in the secondary subsectors represented in the county at the start of the period. In addition, the hypothesis that the coefficient on primary employment is equal in magnitude to the coefficient on secondary employment in absolute value cannot be rejected, suggesting a symmetric shift from the agricultural to the secondary sector.

[10]The only exception is some counties in western provinces, including Gansu and Qinghai, that report high levels of employment in services; this is primarily government employment. These counties are geographically large, but small in terms of population and economic activity.

[11]Any county for which this fraction is above the 25th percentile comparing across all counties at baseline (21%) is designated an industrialized county, $I_{ifp} = 1$; the remainder are designated less industrialized counties, or counties that are primarily agricultural.

No significant impact is observed for tertiary employment. Given the evidence of symmetric but opposite shifts in the primary and secondary sectors, unsurprisingly, no significant change in total employment is observed.

To benchmark the magnitude of the effects, I will employ the within-province standard deviation in the independent variable; this provides a measure of relatively local variation in secondary labor demand shocks. A one standard deviation increase in predicted secondary employment leads to a 22% increase in secondary employment, and a 15% decline in primary employment. (Note that despite the fact that the shifts are comparable in absolute magnitude, the proportional effects are different given that the initial level of employment is higher in the primary sector.)

Panel B reports the results for GDP. An increase in secondary labor demand yields a 15% decline in primary GDP, and large and significant increases in secondary, tertiary and total GDP and GDP per capita.[12] The estimated coefficients suggest a one standard deviation increase in predicted employment results in increases in secondary, tertiary and total GDP of at least 35%. GDP per capita increases by 43%.

Panel C reports measures of agricultural investment and value added. Here, we observe again some evidence of substitution out of agriculture. Counties experiencing positive shocks to the secondary sector show evidence of lower sown area and reduced production of grain; a one standard deviation increase in predicted employment leads to declines in sown area and grain production of around 7%. There is no evidence of a decline in cash crop production, or in primary value added; however, Column (5) indicates there is a large increase in secondary value added.

Heterogeneous effects Additional results examining potential heterogeneity in the observed effects are reported in the appendix. For concision, I focus only on GDP. Table A1 reports the coefficients on the secondary labor demand shock variable estimated separately for provinces of China in three broadly defined regions: north and northeast, south and east, and southwest and northwest.[13] There is no evidence that the GDP response to shocks varies consistently across regions, and the differences between the estimated coefficients are generally not statistically significant.[14]

I also analyze whether there is any evidence of heterogeneity respect to the initial level of industrial concentration in the counties: i.e., is a larger response to the secondary labor demand shock observed in a county in which a single subsector constitutes a higher proportion of total employment at baseline? In order to test this hypothesis, I generate a dummy equal to one if a single subsector within the secondary sector constitutes more than 20% of baseline secondary sector employment, and interact this dummy with the primary shock of interest in equation (3).[15] The results are reported in Panel B of Table A1, and show interaction effects that are negative but generally insignificant.

[12]Again, GDP is reported in millions of yuan, while GDP per capita is reported in yuan.

[13]The north and northeast region includes Hebei, Shanxi, Liaoning, Jilin and Heilongjiang; the south and east region includes Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henen, Hubei, Hunan, and Guangdong; and the southwest and northwest region includes Sichuan, Guizhou, Yunnan, Shaanxi, Gansu and Qinghai.

[14]The same pattern is evident if the coefficients are estimated separately for all six regions.

[15]In a small number of counties, more than one subsector constitutes 20% or more of baseline secondary employment. These counties are also coded with the dummy variable equal to one.

This suggests that the observed pattern is not restricted to counties in which secondary employment is initially highly concentrated.

In addition, given that there is considerable within-region variation in the level of initial economic development, it is useful to examine heterogeneity with respect to the initial size of the local economy. I employ total employment as reported in the baseline census in 2000 to capture the baseline size of the county economy, define a dummy variable equal to one if a county is above the median level of employment, and interact this dummy with the primary shock of interest in equation (3). The results reported in Panel C of Table A1 suggest that the response to secondary labor demand shocks may be somewhat smaller in initially larger counties. However, the same pattern of substitution away from agriculture and into non-agricultural production, as well as increases in overall GDP and GDP per capita, is observed for all counties.

Finally, the substitution of factors into non-agricultural production may be slower in counties where initial levels of human capital are lower, though this phenomenon may be of somewhat limited relevance given that manufacturing in China is predominantly low-skilled. The census reports the number of individuals in the county in different categories of educational attainment, and thus I define a dummy variable for highly educated counties equal to one for those counties where the proportion of the adult population with more than primary education is above the median (49%). I then estimate heterogeneous effects for these counties. The results, reported in Panel D of Table A1, suggest that counties characterized by a higher level of initial human capital do show evidence of relatively larger increases in secondary, tertiary and total GDP following positive shocks to the secondary sector. However, the same basic pattern of increased secondary and total GDP is evident even in counties with lower levels of initial human capital.

Migration destination shocks Given the pattern of widespread outmigration from rural China to large cities, particularly cities in the heavily industrialized coastal region, these results raise the question of whether it is secondary labor demand shocks in rural counties themselves that are more important in stimulating structural transformation, or shocks at the migration destination. The 2000 census reports at the county level the number of outmigrants from that county resident in each province in China. The average county reports 20% of the registered population living outside the county, and 13% outside the province, but there is considerable heterogeneity: the proportion of residents living outside of the county ranges from virtually zero to nearly 70%.

Unfortunately, data on migration is not reported in the provincial statistical yearbooks. However, I can use the data reported in the 2000 census to construct time-invariant weights capturing what proportion of residents from each county have outmigrated to each province, and estimate an "destination shock" equal to the weighted average of predicted employment in each year in each destination province.[16] I then re-estimate equation (3) including both the own-county shock and the destination shock as independent variables. In order to render the magnitudes comparable, both shocks are standardized to have

[16]In this analysis, I exclude migrants who have migrated to other counties within the same province. In this case, it is not easy to separately identify the direct effect of neighboring county shocks on local economic outcomes through general economic spillovers, and the effect running through the channel of within-province migration.

mean zero and standard deviation one.

The results reported in Table A2 for GDP show that the estimated coefficients for the own-county shock remain generally consistent. Though the negative coefficient for primary GDP is now noisily estimated, we continue to observe large positive effects of own-county secondary labor demand shocks on secondary, tertiary and total GDP as well as GDP per capita. In addition, we observe that there is a significant decline in primary GDP, and a significant increase in tertiary GDP, given a positive shock at migration destinations. The magnitudes suggest a one standard deviation increase in predicted employment at historic migration destinations leads to a 33% decrease in primary GDP, while GDP in services approximately doubles. (There is also a large increase in per capita GDP following a positive shock at the migration destination, though this coefficient is noisily estimated.)

This evidence would be consistent with households substituting out of agriculture in response to growth at migration destinations, while remittances may be used to expand local businesses or increase demand for the tertiary sector. Importantly, while shocks at migration destinations are clearly relevant, there is no evidence that these shocks dominate the effect of local secondary shocks. The latter continue to appear economically and statistically significant.

Informativeness of Bartik shocks To evaluate the informativeness of the secondary labor demand shocks over the medium term, I perform a simple robustness check using the 2010 census. This subsequent wave did not report county-level data on employment in disaggregated secondary subsectors, and includes only aggregate data for industrial and mining employment. If observed secondary employment as measured in the 2010 census is regressed on the predicted employment shock in the same year, the coefficient is not significantly different from one and highly significant (t-statistic of nearly ten); note that unlike the previously reported results that reflect only counties that report employment data in provincial yearbooks, this robustness check includes all counties. This suggests that the constructed shocks are, on average, accurately capturing shifts in the secondary sector over this period.[17]

To summarize, the results suggest that there is a substantial response to secondary labor demand shocks at the county level. This is observed not only in increased employment, gross product and value added in the secondary sectors, but also a shrinking agricultural sector. Given that the increases in secondary (and tertiary) GDP are much larger than the decline in primary GDP, total product and GDP per capita also increase. Moreover, these patterns seem to be consistent across regions, and across counties with different initial characteristics in terms of the size and diversification of the local economy.

4.2 Agricultural shocks

The primary objective of this analysis is to analyze the interaction of non-agricultural and agricultural shocks, where the latter are proxied by rainfall. Is the response of county economic outcomes to a positive shock to the secondary sector amplified if this shock follows a boom in agriculture, or is the response dampened?

[17]Tabulations are not reported for concision, but available upon request.

In addition, the conceptual model predicts a very different pattern of interaction between non-agricultural and agricultural shocks for counties that are initially industrialized and those that are not. I define a dummy variable for initial industrialization I_{ifp}, using information about the fraction of the population engaged in non-primary employment as reported in the 2000 census. Any county for which this fraction is above the 25th percentile comparing across all counties at baseline (21%) is designated an industrialized county, $I_{ifp} = 1$; the remainder are designated less industrialized counties, or counties that are primarily agricultural. (I will subsequently demonstrate that the primary results of interest are robust to a range of different definitions of industrialized counties, using various percentile cutoffs.)

Accordingly, it is first useful to present some simpler results demonstrating heterogeneity in the direct effects of rainfall shocks comparing across initially industrialized and initially less industrialized counties. I estimate a specification parallel to equation (3) in which the dependent variables are current and lagged rainfall in the cultivation season, and current and lagged rainfall interacted with I_{ifp}.

$$\begin{aligned} Y_{ifpt} &= \beta_1 R_{ifpt} + \beta_2 R_{ifp,t-1} + \beta_3 R_{ifpt} \times I_{ifp} + \beta_4 R_{ifp,t-1} \times I_{ifp} \\ &= \sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt} + I_{ifp} \times (\sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt}) + \nu_i + \epsilon_{ifpt} \end{aligned} \tag{4}$$

The results are reported in Table 3; at the bottom of each panel, I report the linear combinations $\beta_1 + \beta_3$ and $\beta_2 + \beta_4$, capturing the net effect of an increase in both current and lagged rainfall in less industrialized and more industrialized counties, respectively. It is evident that there is a significant positive response of primary GDP, grain production (albeit noisily estimated), cash production, and primary value added to an increase in rainfall, but only in less industrialized counties. In all cases, a one standard deviation increase in rainfall, benchmarked employing the within-province standard deviation, leads to an increase in agricultural output or value added of between 1 and 2%. In counties that are initially industrialized, however, the interaction terms are generally negative and large in magnitude, and the net effect of rainfall is insignificant.[18]

There is also some evidence of a negative effect of rainfall shocks on GDP and value added in the secondary sector that is particularly important in initially more industrialized counties. While the coefficients of interest are generally not statistically significant for secondary or total GDP, there is a large negative effect of rainfall on secondary value added in initially more industrialized counties. This may reflect a direct negative effect of rainfall on productivity, or the reallocation of capital or time to agriculture given a positive productivity shock to this sector. These results are also consistent with other evidence that local climatic shocks can have a meaningful direct impact on industrial outcomes (Colmer, 2016; Dell et al., 2012; Hsiang, 2010; Santangelo, 2016).

To sum up, the results suggest that rainfall does lead to an increase in primary and total GDP, but only in less industrialized counties. This suggests that the pattern of interaction between non-agricultural and agricultural shocks will also be different in the two sets of counties. I explore this pattern in more detail in the next section.

[18]The only exception to this pattern is the increase in grain production, observed consistently across both initially industrialized and non-industrialized counties. Results for employment are not reported, but given the small sample, there is no significant heterogeneity in the effects of rainfall on employment comparing across less and more industrialized counties.

4.3 Interaction of non-agricultural and agricultural shocks

Again, the primary objective of this analysis is to analyze the interaction of non-agricultural and agricultural shocks, where the latter are proxied by rainfall. First, it is useful to examine some graphical evidence that captures the main intuition of the results. Figure 3 shows the coefficients from simple regressions of GDP on the secondary labor demand shock for counties that are and are not initially industrialized, in each quantile of lagged rainfall.[19] We observe that the response of GDP to a secondary labor demand shock is generally increasing by quantile of lagged rainfall for counties that are initially less industrialized ($I_{ifp} = 0$). However, there is no evidence of a comparable pattern for counties that are initially industrialized ($I_{ifp} = 1$); for these counties, the response of GDP to the secondary labor demand shocks is larger in magnitude on average, but unaffected by rainfall patterns.

This graphical evidence suggests that there may be some complementarity between non-agricultural and agricultural shocks in less industrialized counties: the increase in GDP in response to a secondary labor demand shock is larger when this shock coincides with or follows a boom in agriculture. There is no comparable complementarity in more industrialized counties. This difference will prove to be significant, large in magnitude, and evident across a range of economic outcomes.

The specification of interest thus analyzes the relationship between non-agricultural and agricultural shocks, as well as whether this relationship systematically varies across counties characterized by varying levels of initial industrialization. I regress outcomes of interest on predicted secondary employment, lagged rainfall, and the interaction of the two. I also control flexibly for the direct effect of rainfall by including dummy variables for each quantile of non-agricultural employment (the underlying continuous variable used to construct I_{ifp}) interacted with lagged rainfall, and differential trends for counties in each decile of average rainfall; these control variables are denoted Ω_{ifpt}.

$$
\begin{aligned}
Y_{ifpt} = {} & \beta_1 Pred_{ifpt} + \beta_2 Pred_{ifpt} \times I_{ifp} \\
& + \beta_3 Pred_{ifpt} \times R^{lag}_{ifpt} + \beta_4 Pred_{ifpt} \times R^{lag}_{ifpt} \times I_{ifp} \\
& + \Omega_{ifpt} + \sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt} + I_{ifp} \times \left(\sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt}\right) + \nu_i + \epsilon_{ifpt}
\end{aligned} \tag{5}
$$

The objective is to allow for lagged productivity shocks in agriculture to interact with shocks to the non-agricultural sector, and also for the slope of each response to be different comparing across initially more industrialized counties and initially less industrialized counties. Given that the rainfall measure employed is an average of the two preceding years, I denote this variable R^{lag}_{ifpt}. I will first present results employing a linear rainfall variable; however, given that the specifications of interest are all estimated using county fixed effects, I am effectively estimating the effect of an unusually high level of rainfall relative to the county average, and thus I describe this variable as a rainfall shock. I will subsequently demonstrate the results are robust to the use of alternate rainfall measures.

[19] The quantiles of rainfall are both defined employing the within-province distribution of the shock. The regressions also include province-year and county fixed effects, and differential trends for counties in each quantile of initial secondary employment.

Regression results are reported in Tables 4 through 6. The predicted employment shock and both rainfall shocks are de-meaned in order to facilitate interpretation of the interaction terms, and standard errors are again estimated employing two-way clustering at the county and year level. At the bottom of each table, I report the sums $\beta_1 + \beta_2$ and $\beta_3 + \beta_4$. These linear combinations capture the net effect of a predicted employment shock and a predicted employment shock in conjunction with lagged rainfall for initially industrialized counties. I also report the dependent variable means separately for counties for which $I_{ifp} = 0$ and $I_{ifp} = 1$.

The results for employment are reported in Table 4; in interpreting them, it is useful to note that total employment is around 41% higher in less industrialized counties, and the share of primary employment in these counties is 63%, compared to 47% in initially industrialized counties. First, consider the results for counties that are initially less industrialized. The estimated coefficients β_1 suggest that at the mean level of rainfall, a (within-prefecture) one standard deviation increase in the secondary labor demand shock results only in a decline in primary employment.

However, when this shock follows a positive rainfall shock, there is evidence of increased secondary and total employment. A one standard increase in lagged rainfall in conjunction with a one standard deviation increase in secondary labor demand leads to an increase in secondary employment of 6%, and an increase in total employment of 4%.

For initially industrialized counties, at the mean level of rainfall we can examine the response to a secondary labor demand shock by evaluating the linear combination $\beta_1+\beta_2$. The only response to a secondary labor demand shock is a 15% decline in primary employment and a 12% increase in secondary employment, though the latter is noisily estimated. Equally important, there is little evidence of any interaction between non-agricultural and agricultural shocks for initially industrialized counties: i.e., the linear combinations $\beta_3 + \beta_4$ are uniformly of small magnitude and insignificant. Rainfall shocks do not seem to significantly affect the response of employment to secondary labor demand shocks in these counties.

Table 5 reports the results for GDP. First, for counties that are initially less industrialized ($I_{ifp} = 0$), there is no significant response to a secondary labor demand shock at the mean level of rainfall. If this shock follows or coincides with an increase in rainfall, however, the effects are very different. More specifically, a one standard deviation increase in lagged rainfall in conjunction with the same shock to secondary labor demand leads to an 14 percentage points increase in secondary GDP, and a 8 percentage points increase in overall GDP. There is also some evidence that primary GDP is declining following these two shocks, though the estimated effects are not significant.

For counties that are initially industrialized ($I_{ifp} = 1$), we can evaluate the response of GDP to a secondary labor demand shock at the mean level of rainfall by analyzing $\beta_1 + \beta_2$. Here, there is a 10% decline in primary GDP in response to the shock, and large increases in secondary, tertiary, total, and per capita GDP. Again, however, there is no evidence of complementarity between lagged rainfall shocks and secondary labor demand shocks. The linear combinations $\beta_3 + \beta_4$ are uniformly insignificant.

Finally, Table 6 focuses on agricultural investment and value added. Here, for less industrialized counties, a positive secondary labor demand shock leads to a decline in sown area and cash production, but has no other significant impacts at the mean level of rainfall. By contrast, an increase in secondary value added is observed only when a positive shock to the secondary sector follows or coincides with an increase in rainfall. A one standard deviation increase in rainfall (current or lagged) in conjunction with a one standard deviation increase in predicted employment leads to an increase in value added of 11 percentage points.

For more industrialized counties, the effects of a positive secondary labor demand shock at the mean level of rainfall include a 6-8% decline in sown area and grain production, and a large increase in secondary value added. Again, there is no evidence of significant interactions between rainfall shocks and secondary labor demand shocks.

To sum up, at the mean level of rainfall the evidence generally suggests a more robust response to secondary labor demand shocks in counties that are already more industrialized compared to counties that are initially less industrialized: i.e., β_2 is significant and positive for dependent variables that capture growth and output in non-agricultural sectors. However, the relationship between non-agricultural and agricultural shocks is very different comparing across the two sets of counties.

In counties that are initially less industrialized, there is evidence of complementarity between secondary labor demand shocks and rainfall shocks. In other words, the substitution of productive factors into non-agricultural production and the associated increase in GDP and value added is faster when an secondary labor demand shock follows or coincides with an increase in rainfall. In counties that are initially more industrialized, there is no evidence of a comparable relationship; the two sets of shocks do not seem to interact significantly, consistent with the evidence that the direct effect of rainfall shocks is likewise small in magnitude in more industrialized counties. The channels for this observed pattern will be explored in greater detail in Section 6.

5 Robustness checks

In this section, I present a number of robustness checks on the primary results. I demonstrate that the observed pattern is robust to different characterizations of initially industrialized counties and different definitions of rainfall shocks. In addition, I explore whether endogeneity in the secondary labor demand shocks is a potential source of bias; present evidence that spatial spillovers of rainfall shocks are limited; and analyze the potential for manipulation of the administrative data employed.

5.1 Alternate definitions of initially industrialized counties

In the primary results, I analyze the differential impact of non-agricultural and agricultural shocks in counties that are and are not initially industrialized. Counties for which $I_{ifp} = 1$ are defined as those above the 25th percentile comparing across counties in the proportion of total employment constituted by non-primary employment in the 2000 census. In order to demonstrate that the observed results are not an artifact of this definition, I subsequently re-define initially industrialized counties using a range of cutoffs, ranging from the 10th to the 60th percentile, and re-estimate the primary specification, equation (5).

Figure 4 shows the normalized coefficients (coefficients divided by the standard error) that result from specifications estimated using the varying cutoffs, focusing on the two primary independent variables of interest: the interaction of a secondary labor demand shock and lagged rainfall, and the triple interaction of a secondary labor demand shock, lagged rainfall, and I_{ifp}. Each subfigure shows the same coefficients for a different dependent variable. For concision, I focus on secondary and total GDP.

It is evident that the same pattern is generally consistent across a range of cutoffs below the 50th percentile, but the primary effects are attenuated when the comparison employed focuses on variation in the upper part of the distribution of the initial proportion of non-primary employment. In addition, the estimated coefficients are generally more robust for secondary, as compared to total, GDP. However, the primary results are clearly not dependent on the precise definition of initially industrialized counties.

5.2 Alternate specifications of rainfall shocks

In addition to demonstrating that the primary results are robust to alternate definitions of industrialized counties, I can also demonstrate that the results are robust to alternate definitions of the rainfall shock. In particular, I explore two alternate specifications: the first employs dummy variables for extreme rainfall events, and the second adds additional controls for current, rather than lagged rainfall.

Table 7 reports the results employing dummy variables for rainfall above the 75th percentile and below the 25th percentile, where the cutoffs are calculated employing the within-prefecture distribution of rainfall. (The rainfall control variables Ω_{ifpt} now include dummy variables for each decile of non-agricultural employment interacted with both the dummy rainfall variable and a linear rainfall variable.) We can observe in Panel A the same pattern in which the coefficient β_3 capturing the interaction between non-agricultural and agricultural shocks is positive and significant for secondary and total GDP, while β_4 is negative and significant. In fact, the estimated coefficients have increased in precision: we now observe evidence of significant interactions between non-agricultural and agricultural shocks for GDP per capita.

In Panel B of the same table, I report results employing the dummy variable for rainfall below the 25th percentile. Here, the coefficients are opposite in sign as expected, but small in magnitude and insignificant. This asymmetry would be consistent with the hypothesis that the channel for the observed pattern of complementarity is positive agricultural shocks leading to an increase in the availability of local capital for re-investment. In this case, we would expect to observe that positive shocks have a significant and positive impact on investment in non-agricultural production as some threshold for a minimum indivisible investment is exceeded, but negative shocks have no comparable effects. I will return to this point in the discussion of channels in Section 6.

In Table 8, I add additional interactions between quantiles of non-agricultural employment and current rainfall, in addition to the interaction terms for lagged rainfall. The results here are consistent with the original specification. The coefficients on the interaction terms including current rainfall are generally parallel in sign to the coefficients on the lagged rainfall interactions, but smaller in magnitude and statistically significant only for GDP.

Finally, in Table A3 in the appendix, I estimate a placebo specification in which I employ rainfall shocks following the secondary labor demand shock as the proxy for rainfall. Here, the estimated coefficients are small in magnitude and insignificant. This suggests that the primary results do not reflect some systematic and persistent difference between counties exhibiting different patterns of rainfall in the period of interest.

5.3 Endogeneity of secondary labor demand shocks

One significant threat to the interpretation of these results would be reverse causality in the labor demand shocks. For example, if a particularly fast-growing industry is concentrated in a particular set of prefectures or provinces, then fast growth in this industry may be correlated with other region-specific trends. Accordingly, what we interpret as the effect of a positive secondary labor demand shock on local GDP may in fact reflect the impact of some unobserved local characteristics that are driving GDP growth, and simultaneously affect the secondary sector.

Alternatively, there could be a more generalized form of reverse causality in which counties with certain characteristics — e.g., forward-looking leaders, better-functioning governments, more robust agricultural sectors, etc. — substitute into industrial sectors that are ultimately faster-growing. In this case, the positive effect of demand shocks in the secondary sector could reflect a persistent effect of these initial characteristics.

To address this challenge, I implement several related robustness checks. First, I present evidence that in the period prior to widespread rural industrialization, there is no evidence of differential trends for counties that subsequently experienced more positive secondary labor demand shocks, examining a more limited set of variables available in earlier Chinese censuses. Second, I limit the sample employing various strategies to exclude regions that constitute a high fraction of national employment in a particular subsector, and thus may be driving growth in that sector. Third, I re-create the Bartik shocks excluding the largest two subsectors (by employment) in each county at baseline.

First, it is important to verify that counties concentrated in sectors experiencing varying shocks in the post-2000 period did not show significantly different trends in an earlier period of development. Examining outcomes immediately prior to the base year in 2000 is not particularly informative, given that there is a high degree of serial correlation in shocks to national secondary employment, and there is no reason to believe that a county benefiting from a positive shock to a particular secondary subsector in the period of interest was not benefiting from the same shock in the previous decade. However, it is informative to examine to trends in key outcomes prior to the widespread onset of rural industrialization; in order to do so, I exploit county-level data from the national censuses conducted in 1953, 1964, 1982, and 1990.

With the exception of the 1990 data, the number of variables reported in these earlier censuses is considerably reduced. However, I construct a panel of covariates observed in at least two of the four census rounds: county population, the percentage of the population that is non-agricultural, the percentage employed, the percentage illiterate, the percentage under 15, the percentage over 65, and the percentage 15-49 that is female.[20] The sample of counties is also somewhat reduced, as some counties do not match to earlier census data if they are newly created or have changed names.[21]

[20] The county population is reported in all four census rounds; the percentage of the population that is non-agricultural is reported in 1964, 1982 and 1990; and the remaining variables are reported in 1982 and 1990 only.

[21] More specifically, of the 2062 counties in the main sample, 959 match to the 1953 census, 1141 match

I then construct an artificial shock measure for the pre-period, using the average imputed growth rate in secondary employment calculated in the sample period to impute backwards and impute secondary employment in each census year. I regress the variables of interest on this placebo Bartik shock, to test whether fast-growing counties show evidence of significantly different trends in the pre-industrialization period. The specification of interest also includes county and province-year fixed effects, and province-year fixed effects interacted with the dummy for initial high industrialization I_{ifp}.

$$Y_{ifpt} = \beta Pred_{ifpt} + \gamma_{pt} + \gamma_{pt} \times I_{ifp} + \nu_i + \epsilon_{ifpt} \tag{6}$$

The results of estimating equation (6) are reported in Table 9, and show generally insignificant coefficients. There is evidence that population growth is faster in counties that subsequently experience more positive shocks, and there are also differential trends in the percentage over 65. In Column (8), I report a simple cross-sectional correlation between average output per capita, reported only once in the 1982 census, and the average post-2000 shock $\bar{G}_{ifp}$, and the coefficient is insignificant.[22] This suggests that counties concentrated in industries that are growing more rapidly in the period of interest are not characterized by systematically different trends in the pre-industrialization period.

Second, I analyze subsectors that are highly concentrated in certain regions. It is relatively rare for a county or even a prefecture to constitute a substantial proportion of national employment in a particular subsector. More specifically, only 3% of counties constitute more than 5% of national employment in any subsector, and only 1% constitute more than 10%; at the prefecture level, 17% of prefectures constitute more than 10% of national employment for a given subsector. At the provincial level, it is more common to observe provinces that are a significant contributor to employment in a particular subsector, but only around a quarter of the observed provinces constitute more than 75% of employment in any subsector.

In Tables A4 and A5 in the Appendix, I re-estimate the primary specification of interest, equation (5), for the dependent variables capturing GDP, employing two reduced samples: first, excluding all prefectures that constitute more than 7% of subsector employment for any subsector, and second, excluding all provinces that constitute more than 75% of subsector employment. (In each case, this drops the highest quantile of concentration in prefectures / provinces.) The primary results are consistent.

Third, I seek to address the challenge that counties that are concentrated in very different subsectors may share characteristics other than geography (i.e., differential quality institutions, natural resources, or agricultural sectors). I hypothesize that if counties with different baseline characteristics differentially select into fast-growing or slow-growing industrial subsectors, this selection process is presumably manifest mainly in the identity of the largest subsectors. For example, a fast-growing county may be primarily concentrated in the production of electrical equipment while a slow-growing county is concentrated in coal mining; in both counties, a number of other subsectors constitute smaller fractions of total employment. The objective of this exercise is to abstract from the variation introduced by the hypothetical emergence of electrical equipment vis-a-vis mining as the largest subsector — a difference that may reflect other unobserved differences between the counties — and focus on residual variation in secondary labor demand shocks.

to the 1962 census, 1620 match to the 1982 census, and 1644 match to the 1990 census.

[22]Interestingly, this variable is not even reported in the most recent census rounds of 2000 and 2010.

In order to test this hypothesis, I re-estimate the Bartik shocks excluding the two largest subsectors in each county, calculating predicted growth based on the shares of smaller subsectors at baseline and the growth in national employment observed for those subsectors over time. I then re-estimate the primary specification using these alternate shock measures.

The results are reported in Table A6 in the Appendix, and again are highly consistent; in fact, in this specification, there is evidence of faster substitution out of the primary sector, in addition to more rapid growth in the secondary sector, following the conjunction of agricultural and non-agricultural shocks in initially less industrialized counties. Some counties may be "betting on winners" (or losers), but excluding the largest subsectors that might plausibly reflect these unobserved policy or institutional differences does not affect the results.

5.4 Spillovers from one county to another

Up to this point, the analysis has not taken into account potential spatial spillovers of non-agricultural shocks. Local economic outcomes may respond to secondary labor demand shocks in adjacent counties via several channels. First, there may be flows of capital and labor from one county to another; in particular, a slow-growing county that is proximate to other, faster-growing counties may be characterized by net outward flows of capital and labor. Second, there could be positive spillovers of more rapid growth in a neighboring county if this shock generates increased demand for local output. Spillovers could also be a source of bias in the main results if the absence of a response to secondary labor demand shocks in less industrialized counties reflects primarily the outflow of capital and labor to more industrialized counties experiencing similar positive shocks.

I first focus on the prefecture as a unit capturing the local region; a prefecture encompasses around six counties on average, and has an average population of about three million people. This are clearly many ways to define a geographic area within which cross-country spillovers are relevant, and I will also demonstrate that these results are robust to alternate methods.

Second, I match each county to other counties in the same prefecture that are within certain specified distance ranges (0–25 kilometers, 25–50 kilometers, 50–75 kilometers, and 75–100 kilometers), where distances are calculated exploiting straight lines between the county centroids. If there are multiple other counties within the prefecture that fall within a particular distance range, I employ the closest such county as the match. (Counties in different prefectures are excluded in this analysis.) The sample of matched counties declines for the larger distances, as not all smaller prefectures encompass county pairs separated at a distance of more than 50 kilometers.

I then estimate a series of specifications regressing GDP on the own-county secondary labor demand shock, as well as the "neighboring county" shock, denoted $NPred_{ifpt}$, within the specified radius. The specification of interest can be written as follows. It is useful to note that, while the absolute correlation of own and neighboring shocks is high, conditional on county and province-year fixed effects the correlation is only around .05

and does not vary systematically with the distance between county pairs.

$$Y_{ifpt} = \beta_1 Pred_{ifpt} + \beta_2 NPred_{ifpt} + \sum_{i=1}^{4} \lambda^i_{nonag} \times t + \sum_{i=1}^{4} \lambda^i_{tot} \times t + \gamma_{pt} + \nu_i + \epsilon_{ifpt} \quad (7)$$

The results of estimating equation (7) are reported in Table 10. We can observe that the coefficients β_2 are positive and small in magnitude for counties within a 25-kilometer radius, but are almost uniformly insignificant for shocks of more distant neighbors. I also report in each panel the p-values corresponding to the test $\beta_1 = \beta_2$. The hypothesis that the effects of own-county and neighboring-county shocks are identical can generally be rejected, except for the specifications employing primary GDP as the dependent variable.

I also re-estimate the specification including the interaction of non-agricultural and agricultural shocks, equation (5), adding a variable capturing the shock experienced by the neighboring county within a 50-kilometer radius, and the interaction between the neighboring county shock and I_{ifp}. These results are reported in Table 11. The primary pattern of complementarity in non-agricultural and agricultural shocks only in less industrialized counties remains consistent, and the coefficients on the neighboring county shock are varying in sign and generally insignificant.

Arguably, focusing only on within-prefecture neighbors is incomplete. In Table A7 in the Appendix, I re-estimate the simpler specification capturing cross-county dependence on neighboring shocks, equation (7), employing an alternate definition of neighboring counties. Rather than restricting the geographic scope to the prefecture, I match each county with the closest county in the specified distance range in the same province. The results are entirely consistent, and the hypothesis that own-county and neighboring-county shocks have the same effect can again generally be rejected.

While informative, these results assume a symmetric effect of cross-dependence of shocks for both rapidly-growing and slowly-growing counties. This may not be a plausible assumption. Spillovers could be unidirectional: counties that are growing slowly experience an outflow of capital and labor in response to shocks in adjacent, rapidly-growing counties, but there is no response by these "leading counties" to shocks in "lagging counties." In order to test this hypothesis, I rank counties within each prefecture in terms of the magnitude of the shocks to secondary labor demand that they are experiencing in this period: more specifically, I rank them according to the average percentage change in secondary employment predicted using the Bartik shock. "High growth counties" are defined as counties above the 75th percentile of this within-prefecture distribution of predicted secondary employment growth.

I then define "High shock," $Pred^{high}_{fpt}$, as a variable that varies at the prefecture-year level, equal to the average shock experienced by high growth counties in that prefecture and year. The specification of interest regresses economic outcomes on the own-county shock and the high shock variable, for a sample of counties that are growing slowly: more specifically, counties below the 25th percentile of the within-prefecture distribution of predicted secondary employment growth. (Counties in the middle two quantiles of the within-prefecture distribution of demand shocks are not included in this analysis.) The specification can thus be written as follows.

$$\begin{aligned} Y_{ifpt} &= \beta_1 Pred_{ifpt} + \beta_2 Pred^{High}_{fpt} \\ &+ \sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt} + I_{ifp} \times (\sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt}) + \nu_i + \epsilon_{ifpt} \end{aligned} \tag{8}$$

I again focus only on GDP. The results are reported in Table A8 in the Appendix, and show coefficients on the high shock that are again small in magnitude and uniformly insignificant. The coefficients β_1 remain comparable to those originally estimated using the simple specification reported in Table 2; the only difference is that for this sample, the coefficient on the non-agricultural shock is positive for primary GDP, rather than negative.[23] There is no evidence of significant spillovers, positive or negative, from fast-growing counties to slow-growing counties within the same prefecture, suggesting this is not a source of bias in the main results.

5.5 Manipulation of statistical data

Given that this analysis is employing statistical data published by provincial governments, there may be a risk of bias introduced by selective manipulation of economic statistics by local or provincial statistical authorities. While this form of manipulation is not easy to eliminate, it is feasible to estimate its magnitude by comparing the provincial-level statistics generated by aggregating up county-level indicators to provincial-level indicators produced by the National Bureau of Statistics. I focus on gross domestic product, given that this is an economic indicator that may be particularly subject to misreporting.

Using this data, I construct an "overestimate fraction" at the province-year level that is equal to the ratio of the sum of the county-level data divided by the provincial statistic. (I also re-scale the county sum to take into account counties that are missing in the primary sample.) The mean overestimate fraction is 1.2, indicating that the county-level data suggests GDP 20% higher than the provincial statistics, but there is also considerable variation across provinces and years.

I then estimate a series of regressions at the province-year level regressing this overestimate fraction on the Bartik shocks, the lagged rainfall shocks, the interaction between the two, and the interaction of both with the initial high industrialization dummy. (The independent variables are all collapsed to the province-year level.) All specifications include province and year fixed effects. The simplest specification can be written as follows.

$$Frac_{pt} = \beta_1 Pred_{pt} + \mu_p + \eta_t + \epsilon_{pt} \tag{9}$$

The results are reported in Table 12, and show coefficients that are generally small in magnitude and insignificant. This suggests that while there may be some manipulation of the official GDP data, it does not appear to be correlated with the key shocks of interest, and thus should not be a significant source of bias.

[23] If the simple specification, excluding $High_{fpt}$, is estimated for this reduced sample, then the coefficient β_1 for primary GDP is positive but insignificant.

6 Channels

In this section, I return to the three postulated channels through which agricultural shocks can affect the local response to secondary labor demand shocks, and seek to provide additional evidence to identify the channel that is operative. The hypothesized factor demand effect suggested that positive shocks to agriculture would slow the substitution of factors into non-agricultural production, generating negative coefficients on the interaction terms of non-agricultural and agricultural shocks. There is no evidence of this pattern, suggesting this channel is not of first-order relevance.

There are, however, two channels that could be consistent with the observed pattern. The first is a local demand market effect in which rainfall shocks increase local income and thus generate a positive demand shock for local industries. The second is a capital channel in which rainfall shocks and the associated increase in income allow localities to make a certain minimum indivisible investment to enter non-agricultural production. I will present evidence about each channel in turn.

6.1 Local demand effects

One channel that could explain the observed pattern of interaction between non-agricultural and agricultural shocks is positive demand shocks for local industries. Positive agricultural productivity shocks generate an increase in income for rural households; if a non-trivial proportion of non-agricultural output in the county is consumed locally, then this may generate positive shocks for local industry. It is not obvious a priori why this would lead to complementarity between agricultural and secondary labor demand shocks. However, if there are some nondivisible costs to expanding production given positive price or productivity shocks at a national level, producers may be more likely to respond robustly if they also observe positive local demand shocks.

In order to test this hypothesis, I exploit the fact that there is variation across industries (and regions) in the extent to which the secondary sector is oriented toward local demand, demand in the broader domestic market, and export production. If the primary channel through which agricultural shocks interact with non-agricultural shocks is via increased consumer demand, counties in which secondary employment is concentrated in export-oriented industries, or industries broadly integrated into the national market, should show no significant interactions between the two sets of shocks. Needless to say, local demand shocks are irrelevant for producers that are not dependent on the local market. I will conduct two tests to examine this hypothesis; the first exploits variation in production for export, and the second exploits variation in interprovincial trade.

First, I employ data from the National Bureau of Statistics in China reporting the annual sales value and export value by industry for the same set of disaggregated subsectors employed to construct the Bartik shocks. While this trade data is not available contemporaneously (it is reported only post-2012), I compile for each industry the average share of sales value that is constituted by exports. This proportion ranges from virtually zero for mining and other extractive industries, to over 50% for electrical equipment. I calculate an estimated export fraction for each county using the export fractions by sector and the composition of baseline secondary employment by sector, and define an export dummy Exp_{ifp} equal to one if a county's fraction of output exported is above the 25th percentile comparing across all counties.

I then estimate a specification parallel to the main specification of interest, but adding a new set of interaction terms including Exp_{ifp}.

$$\begin{aligned} Y_{ifpt} = {} & \beta_1 Pred_{ifpt} + \beta_2 Pred_{ifpt} \times I_{ifp} + \beta_3 Pred_{ifpt} \times Exp_{ip} \\ & + \beta_4 Pred_{ifpt} \times R^{lag}_{ifpt} + \beta_5 Pred_{ifpt} \times R^{lag}_{ifpt} \times I_{ifp} + \beta_6 Pred_{ifpt} \times R_{ifp,t-1} \times Exp_{ifp} \\ & + \Omega_{ifpt} + \sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt} + I_{ifp} \times (\sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt}) + \nu_i + \epsilon_{ifpt} \end{aligned} \tag{10}$$

In addition to dummy variables for quantiles of lagged rainfall and the same variables interacted with I_{ifp}, Ω_{ifpt} now also includes interactions between each quantile of lagged rainfall and the high export dummy.

The primary coefficient of interest in this specification is β_6. If the main channel for the observed effect is via local demand shocks, then this coefficient should be significant and negative: there will be no complementarity between agricultural and non-agricultural shocks in counties heavily concentrated in export-oriented production.

For concision, I again focus only on the results for GDP, reported in Table 13. The results show the coefficients on the interaction terms of secondary labor demand shocks and rainfall and the triple interaction terms with the initially industrialized dummy (β_4 and β_5) remain comparable in magnitude and significance to the original, more parsimonious specification. The interaction term with the high export dummy is significant and negative only for secondary GDP (and narrowly so for tertiary GDP), but smaller in magnitude compared to the interaction term with I_{ifp}. This suggests that while local demand shocks may be relevant, they are not the primary channel for the observed effect. (The results are similar if a continuous variable for the fraction of county output exported is employed, rather than a dummy variable.)

Second, I want to analyze heterogeneity across regions in which production is primarily for local markets, compared to those where interprovincial trade is more salient. In order to do so, I draw on Poncet (2005), in which the author calculates domestic border effects by industry and by province. The border effect is interpretable as the tariff equivalent of crossing a provincial border for output in a certain industry or a certain province. I then employ these estimated coefficients to construct a border effect for each county, equal to the mean of the weighted average of the industrial border effects for the industries present in the county (utilizing the 2000 country-level employment shares as weights), and the province border effect for the province in which the county lies.

These county-level border effect estimates, denoted $Border_{ifp}$, are then interacted with the secondary labor demand shocks and their interactions with rainfall shocks in order to construct a specification parallel to equation (10). The results of estimating this specification are reported in Table 14. Here, the evidence suggest that there is no meaningful variation in the complementarity between non-agricultural and agricultural shocks comparing across counties that have varying levels of openness to domestic trade. The coefficients on the triple interaction of secondary labor demand shocks, rainfall shocks and the border effect are small in magnitude and insignificant, with the exception of the coefficient estimated for primary GDP. Again, this is consistent with the hypothesis that local demand effects are not the primary channel for the observed relationship between the two sets of shocks, given that local demand effects would be restricted to counties where neither production for export nor production for cross-border trade are important. A similar pattern is observed if a fully saturated specification is estimated including interactions with both Exp_{ifp} and $Border_{ifp}$; these results are reported in Table A9 in the Appendix.

It is also useful here to return to the evidence already presented around asymmetry in the effect of rainfall shocks. When dummy variables are constructed equal to one if rainfall is above or below the prefecture-specific 25th and 75th percentiles, the results suggest that unusually high rainfall leads to a significant increase in the response of secondary GDP to secondary labor demand shocks. However, there is no symmetric effect of unusually low rainfall, as reported in Table 7. This is consistent with the hypothesis that the primary channel is not local demand effects: negative rainfall shocks do generate a decline in local GDP and presumably, local demand for secondary output, but seemingly do not affect the response of local GDP to positive secondary labor demand shocks.

6.2 Capital channel

If the capital channel is relevant, positive rainfall shocks may lead to an increase in local capital available for re-investment in counties that may be initially constrained from making minimum indivisible investments to expand the non-agricultural sector. This channel can also be directly tested by re-estimating the simpler, linear specification including rainfall shocks and their interactions with I_{ifp}; this is equation (3), reproduced here for reference.

$$\begin{aligned} Y_{ifpt} &= \beta_1 R_{ifpt} + \beta_2 R_{ifp,t-1} + \beta_3 R_{ifpt} \times I_{ifp} + \beta_4 R_{ifp,t-1} \times I_{ifp} \\ &= \sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt} + I_{ifp} \times (\sum_{i=1}^{4} \lambda^i_{sec} \times t + \gamma_{pt}) + \nu_i + \epsilon_{ifpt} \end{aligned}$$

I estimate this equation employing as the dependent variables the two data series available that provide information about local investment: capital investment (reported only pre-2004) and local government expenditure. The results are reported in Table 15.

While there is no evidence of any significant pattern given the short time series of private investment data, government expenditure follows exactly the hypothesized pattern: it increases in response to rainfall in less industrialized counties, but is uncorrelated with rainfall in more industrialized counties. This is suggestive evidence that local investment in public goods or directly in industrial enterprises owned or subsidized by local governments may be a channel for the observed complementarity of non-agricultural and agricultural shocks in initially less industrialized counties.

7 Conclusion

This paper seeks to analyze evidence about patterns of industrialization in rural China between 2000 and 2010, focusing on three primary questions. First, what is the impact of a positive secondary labor demand shock on employment and production in non-agricultural and agricultural sectors? Second, is there evidence of complementarity between shocks to secondary labor demand and agricultural productivity shocks, where the latter are proxied by rainfall? Third, does this pattern vary in counties that are initially more industrialized compared to those that are less industrialized at baseline?

The results suggest that there is significant complementarity between non-agricultural and agricultural shocks, but only in counties that are initially not industrialized. An increase in rainfall in the years prior to a positive shock to secondary labor demand seems to amplify the response to the shock at the county level, leading to more rapid substitution of productive factors away from agriculture and into non-agricultural production, and more rapid increases in GDP. However, there is no evidence of a comparable effect in counties that are already relatively industrialized at baseline.

The most plausible channel for this pattern is via local capital markets. If counties have limited access to external capital and utilize income from agricultural production to invest in new productive sectors, they may be constrained from making certain minimum indivisible investments in non-agricultural production. A positive shock to agricultural production may then relax this constraint, leading to increased industrialization and more rapid overall growth.

References

Autor, David and Mark Duggan, "The rise in the disability rolls and the decline in unemployment," *Quarterly Journal of Economics*, 2002, *118* (1), 157–205.

Bartik, Timothy, *Who benefits from state and local economic development policies?*, W.E. Upjohn Institute for Employment Research, 1991.

Benjamin, Dwayne, Loren Brandt, and John Giles, "The evolution of income inequality in rural China," *Economic Development and Cultural Change*, 2005, *53* (4), 769–824A.

_ , _ , _ , and Sangui Wang, "Income inequality during China's economic transition," in Loren Brandt and Thomas Rawski, eds., *China's Great Economic Transformation*, New York: Cambridge University Press, 2008, pp. 729–775.

Blanchard, Olivier, Lawrence Katz, Robert Hall, and Barry Eichengreen, "Regional evolutions," *Brookings Papers on Economic Activities*, 1992, (1), 1–75.

Bound, John and Harry Holzer, "Demand shifts, population adjustments and labor market outcomes," *Journal of Labor Economics*, 2000, *18* (1), 20–54.

Brandt, Loren and Xiaodong Zhu, "Accounting for China's growth," 2010. IZA Working Paper.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural productivity and structural transformation: Evidence from Brazil," 2015.

Christiansen, Luc, Lionel Demery, and Jesper Kuhl, "The (evolving) role of agriculture in poverty reduction – an empirical perspective," *Journal of Development Economics*, 2011, *96* (2), 239–54.

Colmer, Jonathan, "Weather, labor reallocation and industrial production: Evidence from India," 2016.

de Janvry, Alain, Elisabeth Sadoulet, and Nong Zhu, "The role of non-farm incomes in reducing rural poverty and inequality in China," 2005.

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

_ , Benjamin Jones, and Benjamin Olken, "What do we learn from the weather? The new climate-economy literature," *Journal of Economic Literature*, 2014, *52* (3), 740–798.

Deng, Yongheng, Joseph Gyourko, and Jing Wu, "Land and house price measurement in China," 2001. NBER Working Paper No. 18403.

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

Gemmell, Norman, Timothy Lloyd, and M. Mathew, "Agricultural growth and intersectoral linkages in a developing economy," *Journal of Agricultural Economics*, 2000, *51* (3), 353–70.

Gollin, Douglas, David Lagakos, and Michael Waugh, "The agricultural productivity gap," *Quarterly Journal of Economics*, 2014, *129* (2), 939993.

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

Hsiang, Solomon, "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America," *Proceedings of the National Academy of Sciences*, 2010, *107* (35), 1536772.

Huang, Jikun, Keijiro Otsuka, and Scott Rozelle, "Agriculture in China's development: Past disappointments, recent successes and future challenges," in Loren Brandt and Thomas Rawski, eds., *China's Great Economic Transformation*, New York: Cambridge University Press, 2008, pp. 467–505.

Jedwab, Remi, "Cash crop windfalls and development: On cocoa and cities in Ivory Coast and Ghana," 2011.

Jenq, Christina and Junjian Yi, "Female education, strategic premarital investments and inequality: Investment in China," 2013.

Johnston, Bruce and John Mellor, "Agriculture in African development: Theories and strategies," *American Economic Review*, 1961, *51* (4), 56693.

Jones, Benjamin F. and Benjamin A. Olken, "Climate shocks and exports," *American Economic Review*, 2010, *100* (2), 454–59.

Lagakos, David and Michael Waugh, "Selection, agriculture, and cross-country productivity differences," *American Economic Review*, 2013, *103* (2), 948–80.

Li, Hongbin, Pak Wai Liu, Maoliang Ye, and Junsen Zhang, "Does money buy happiness? Evidence from twins in urban China," 2011.

Ligon, Ethan and Elisabeth Sadoulet, "Estimating the effects of aggregate agricultural growth on the distribution of expenditures," 2008. Background paper for the *World Development Report 2008: Agriculture for Development*, World Bank, Washington, D.C.

Liu, Zhiqiang and Qingyan Shang, "Individual well-being in urban China: The role of income expectations," *China Economic Review*, 2012, *4* (23), 833–849.

Luttmer, Erzo, "Neighbors as negatives: Relative earnings and well-being," *Quarterly Journal of Economics*, 2005, *120* (3), 963–1002.

Marden, Samuel, "The agricultural roots of industrial development: "Forward linkages" in reform era China," 2015.

Mukherjee, Amit and Xiaobo Zhang, "Rural industrialization in China and India: Role of policies and institutions," *World Development*, 2007, *35* (10), 1621–1634.

Poncet, Sandra, "A fragmented China: Measure and determinants of Chinese domestic market disintegration," *Review of International Economics*, 2005, *13* (3), 409–430.

Ravallion, Martin and Shaohua Chen, "China's (uneven) progress against poverty," *Journal of Development Economics*, 2007, *82* (1), 1–42.

Santangelo, Gabriella, "Firms and farms: The impact of agricultural productivity on the local Indian economy," 2016.

Tao, Fulu, Masayuki Yokozawa, Jiyuan Liu, and Zhao Zhang, "Climatecrop yield relationships at provincial scales in China and the impacts of recent climate trends," *Climate Research*, 2008, *38*, 8394.

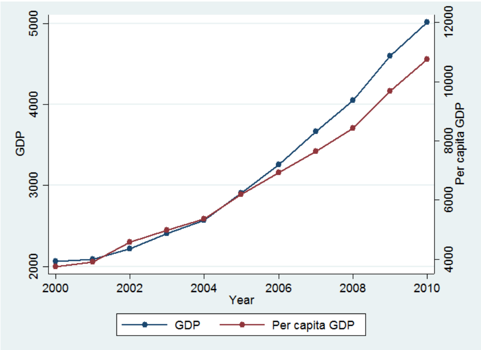
Tiffin, Richard and Xavier Irz, "Is agriculture the engine of growth?," *Agricultural Economics*, 2006, *35* (1), 79–89.

Zhu, Xiaodong, "Understanding China's growth: Past, present, and future," *Journal of Economic Perspectives*, 2012, *26* (4), 103–124.

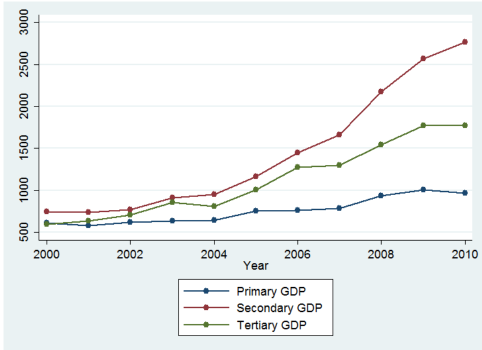
8 Figures and Tables

Figure 1: Trends in economic outcomes

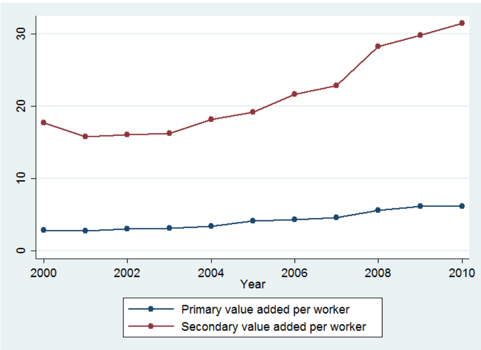
(a) GDP



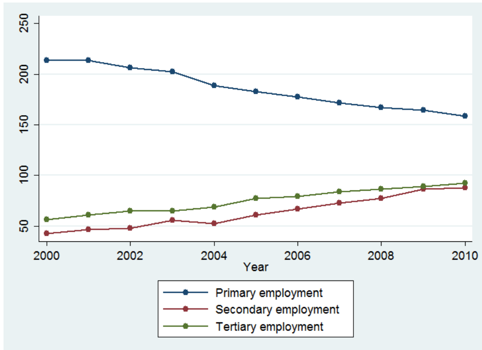
(b) GDP by sector



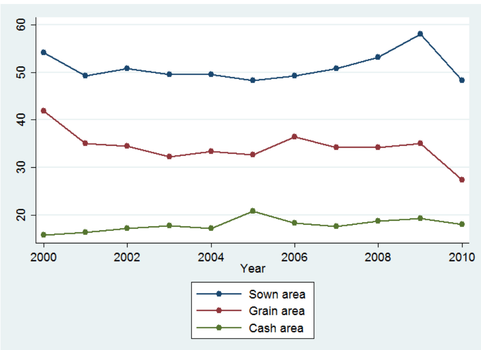
(c) Value added per worker by sector



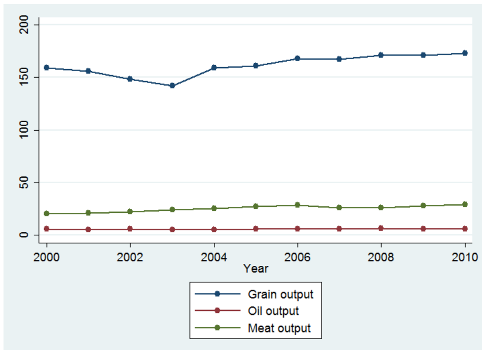
(d) Employment by sector



(e) Sown area

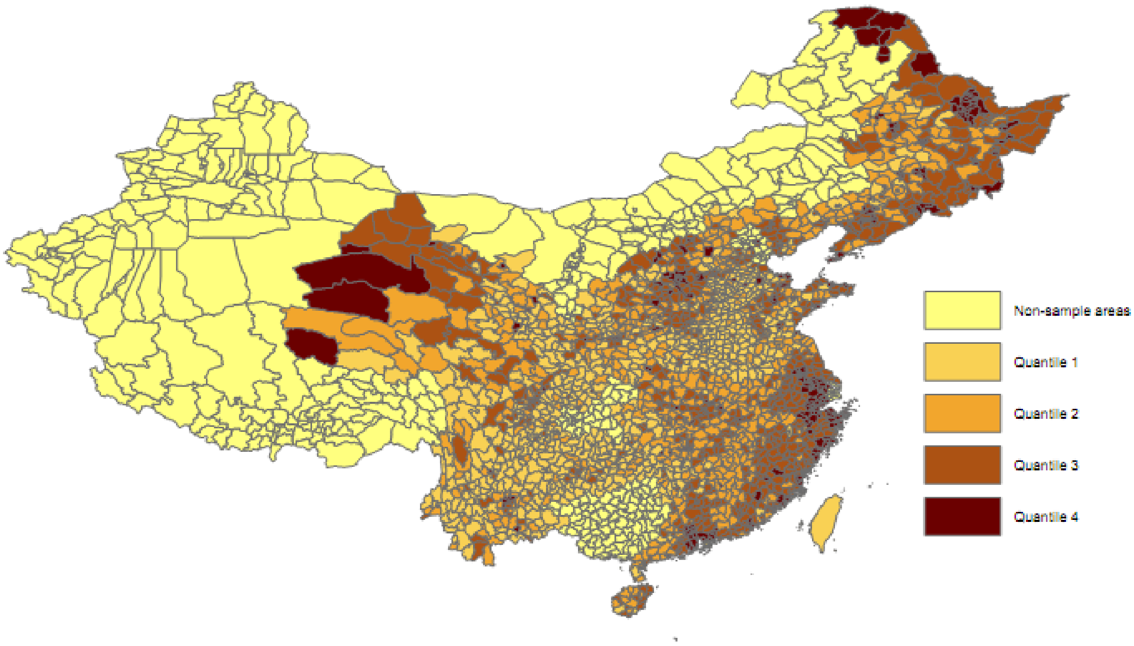


(f) Agricultural production



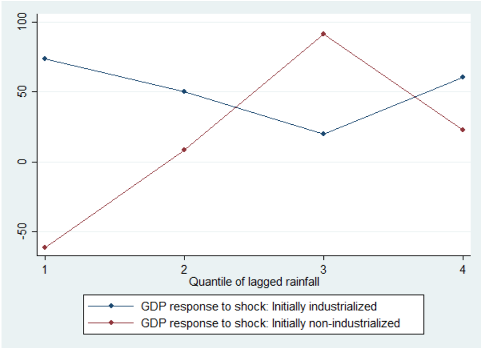
Notes: The graphs report the median of the specified measure by year in the sample counties; only counties that report each measure in at least eight of the ten sample years are included in order to avoid bias due to counties entering and exiting the sample.. GDP and GDP in the primary, secondary and tertiary sectors is reported in millions of yuan; GDP per capita is reported in yuan; employment is reported in thousands of person; sown area is reported in thousands of hectares; and agricultural production is reported in tons.

Figure 2: Map of counties by quantile of initial non-primary employment



Notes: This graph reports counties in each quantile of non-primary employment as a percentage of total employment, utilizing the employment data as reported in the 2000 census. Counties in autonomous regions that are not included in the sample are denoted by the lightest color.

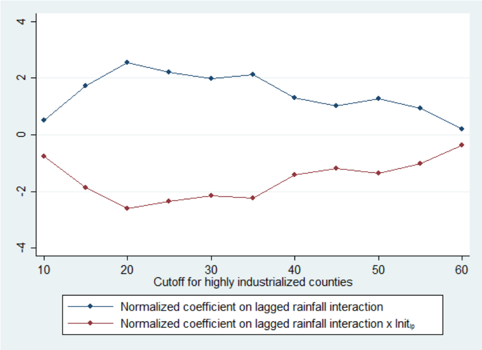
Figure 3: Response to secondary labor demand shock by quantile of lagged rainfall



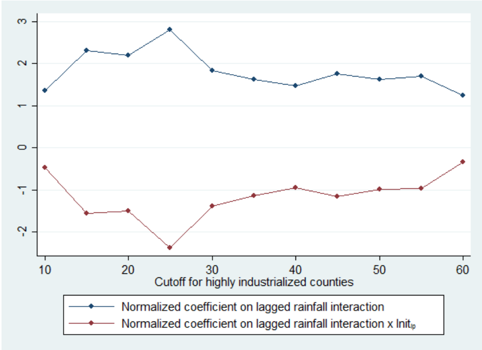
Notes: The graphs show the coefficients from regressions of GDP on a dummy variable for a secondary labor demand shock for counties that are and are not initially industrialized, in each quantile of lagged rainfall. The dummy variables for high secondary shocks and the quantiles of rainfall are both defined employing the within-province distribution of the shock. The regressions also include province-year and county fixed effects, and differential trends for counties in each quantile of initial secondary employment.

Figure 4: Estimated coefficients using varying definitions of industrialized counties

(a) Secondary GDP



(b) GDP



Notes: These graphs show the normalized coefficients (coefficient divided by standard error) on the interaction of a secondary labor demand shock and lagged rainfall, and the triple interaction of a secondary labor demand shock, lagged rainfall and I_{ifp}, that result from re-estimating the main specification, equation (5), for varying definitions of I_{ifp}. The cutoff in terms of the initial percentile of non-primary employment employed to construct I_{ifp} is displayed on the x-axis, and each subfigure is estimated for a separate dependent variable.

Table 1: Summary statistics

Variable	Mean	St. dev.	Min.	Max.	Obs.
Population	481.75	289.83	58.1	1328.2	16304
Employment	248.36	172.3	5.9	610.4	4335
Primary employment share	.54	.18	.01	.96	2713
Secondary employment share	.21	.13	0	.74	2832
Tertiary employment share	.28	.15	.04	.97	2813
GDP	5149.29	5662.32	296.01	45114.88	16996
GDP per capita	14168.73	72150.47	0	2096616	18394
Primary GDP share	.26	.13	0	.95	9993
Secondary GDP share	.41	.15	.02	.9	10202
Tertiary GDP share	.33	.09	.04	.93	10300
Sown area	58.1	47.34	1.85	215.92	6517
Grain area	36.77	31.03	.78	131.77	4891
Grain production	220.2	181.76	8.2	875.7	16080
Cash crop production	50.08	44.97	3.4	210.7	15629

Notes: The mean, standard deviation, minimum, maximum and number of observations for key variables are reported for the full sample for which the predicted employment shock variable can be constructed. Total employment is reported in thousands of person; the employment shares report the percentage of total employment constituted by employment in the specified sector. GDP is reported in millions of yuan and GDP per capita in yuan; the GDP shares report the percentage of total GDP constituted by the specified sector. Sown area and grain area are reported in thousands of hectares, and grain production and cash crop production are reported in thousands of tons.

Table 2: Secondary demand shocks and local economic outcomes

Panel A: Employment

	Primary (1)	Secondary (2)	Tertiary (3)	Total emp. (4)	
Pred. emp.	-1.505	.914	-.072	-1.213	
	(.732)**	(.549)*	(.356)	(.779)	
Mean dep. var.	189.42	77.92	80.12	346.54	
Obs.	2919	3051	2941	2744	

Panel B: GDP

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Pred. emp.	-5.469	40.885	115.562	122.111	239.813
	(2.573)**	(12.900)***	(26.250)***	(22.187)***	(82.269)***
Mean dep. var.	927.88	2064.32	2522.82	4849.13	10482.49
Obs.	9913	10052	10284	15255	15926

Panel C: Agricultural investment and value added

	Sown area (1)	Grain (2)	Cash (3)	Primary va (4)	Secondary va (5)
Pred. emp.	-.215	-.721	.048	.068	5.239
	(.095)**	(.372)*	(.119)	(.161)	(1.133)***
Mean dep. var.	60.39	224.17	50.85	89.47	197.47
Obs.	6147	14277	13828	14343	14443

Notes: The independent variable is predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector. All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}.

The dependent variables in Panel A are total employment and employment by sector, reported in thousands of persons; in Panel B, GDP and GDP by sector reported in millions of yuan, and GDP per capita reported in yuan; and in Panel C, sown area reported in thousands of hectares and production of grain and cash crops reported in thousands of tons, and value added reported in the primary and secondary sectors in thousands of yuan. All measures of GDP and value added are deflated employing the World Bank GDP deflator. Asterisks indicate significance at the ten, five and one percent level.

Table 3: Rainfall shocks and local economic outcomes

Panel A: GDP

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Current rainfall	11.871	-3.711	15.682	14.841	-42.681
	(10.484)	(10.899)	(7.806)**	(10.249)	(36.329)
Current rainfall x I_{ifp}	-4.690	-25.465	-9.102	-101.848	-28.487
	(11.636)	(26.837)	(55.221)	(53.379)*	(323.173)
Lagged rainfall	9.758	4.882	7.040	13.062	-35.127
	(7.018)	(8.040)	(8.431)	(7.453)*	(31.704)
Lagged rainfall x I_{ifp}	-13.288	-17.074	-37.377	-34.995	393.736
	(8.898)	(18.431)	(44.745)	(51.252)	(399.592)
$\beta_1 + \beta_3$	21.63	1.171	22.722	27.902	-77.807
	(11.541)*	(11.178)	(10.079)**	(18.092)	(48.912)
$\beta_2 + \beta_4$	3.652	-41.368	-23.756	-108.940	287.442
	(10.566)	(26.455)	(65.987)	(66.706)	(605.723)
Mean dep. var.: $I_{ifp} = 0$	913.03	1192.06	896.00	3080.55	5215.63
Mean dep. var.: $I_{ifp} = 1$	930.77	2400.43	3101.71	5391.61	12022.17
Obs.	10453	10597	10878	15761	17102

Panel B: Agricultural investment and value added

	Sown area (1)	Grain (2)	Cash (3)	Primary va (4)	Secondary va (5)
Current rainfall	-.047	2.148	.888	.915	-.056
	(.341)	(2.452)	(.619)	(.694)	(.469)
Current rainfall x I_{ifp}	.085	-2.702	-.819	-.921	-2.750
	(.343)	(.958)***	(.444)*	(.494)*	(1.565)*
Lagged rainfall	-.208	2.574	.453	.915	.130
	(.267)	(1.200)**	(.387)	(.384)**	(.251)
Lagged rainfall x I_{ifp}	.247	.036	-.242	-.731	-2.442
	(.394)	(1.551)	(.286)	(.348)**	(1.355)*
$\beta_1 + \beta_3$	-.255	4.722	1.341	1.83	.074
	(.440)	(3.61)	(.701)*	(.830)**	(.818)
$\beta_2 + \beta_4$	.076	2.055	.279	.178	-5.118
	(.395)	(2.618)	(.367)	(.524)	(1.784)***
Mean dep. var.: $I_{ifp} = 0$	84.29	248.44	55.46	86.62	123.72
Mean dep. var.: $I_{ifp} = 1$	51.22	212.12	48.17	89.32	218.53
Obs.	6488	15004	14505	15088	14879

Notes: The independent variables are current and lagged rainfall in the region-specific cultivation season interacted with I_{ifp}, a dummy variable for counties that are initially industrialized. All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. The linear combinations $\beta_1 + \beta_3$ and $\beta_2 + \beta_4$ report the net effect of an increase in lagged and current rainfall in initially less industrialized and initially industrialized counties, respectively.

The dependent variables in Panel A are GDP and GDP by sector reported in millions of yuan, and GDP per capita reported in yuan; and in Panel B, sown area reported in thousands of hectares and production of grain and cash crops reported in thousands of tons, and value added reported in the primary and secondary sectors in thousands of yuan. All measures of GDP and value added are deflated employing the World Bank GDP deflator. Asterisks indicate significance at the ten, five and one percent level.

Table 4: Interaction of non-agricultural and agricultural shocks: Employment

	Primary	Secondary	Tertiary	Total emp.
	(1)	(2)	(3)	(4)
Pred. emp.	-6.617	.030	-1.118	-4.982
	(3.913)*	(2.902)	(2.352)	(3.353)
Pred. emp. x I_{ifp}	5.504	.471	1.278	4.256
	(3.919)	(2.801)	(2.329)	(3.378)
Lagged rainfall int.	.126	.140	.014	.287
	(.099)	(.083)*	(.083)	(.109)***
Lagged rainfall int. x I_{ifp}	-.104	-.058	-.026	-.283
	(.105)	(.094)	(.080)	(.111)**
$\beta_1 + \beta_2$	-1.113	.501	.161	-.726
	(.614)*	(.497)	(.312)	(.738)
$\beta_3 + \beta_4$	.022	.083	-.012	.004
	(.055)	(.053)	(.025)	(.026)
Mean dep. var.: $I_{ifp} = 0$	266.04	73.94	87.06	419.96
Mean dep. var.: $I_{ifp} = 1$	139.03	80.59	75.85	296.71
Obs.	2919	3051	2941	2744

Notes: The dependent variables are total employment and employment by sector, reported in thousands of persons. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 5: Interaction of non-agricultural and agricultural shocks: GDP

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per-capita (5)
Pred. emp.	5.991	60.833	13.273	-44.983	-38.854
	(11.110)	(47.254)	(16.815)	(45.783)	(141.612)
Pred. emp. x I_{ifp}	-10.846	-21.746	99.375	167.685	269.668
	(11.350)	(46.569)	(32.941)***	(51.419)***	(164.732)
Lagged rainfall int.	-.857	5.044	.211	7.289	5.230
	(.600)	(2.445)**	(.728)	(2.793)***	(4.718)
Lagged rainfall int. x I_{ifp}	.663	-5.659	1.321	-5.380	-4.312
	(.599)	(2.581)**	(1.389)	(2.653)**	(4.117)
$\beta_1 + \beta_2$	-4.854	39.087	112.648	122.702	230.813
	(2.320)**	(12.580)***	(26.731)***	(22.670)***	(77.125)***
$\beta_3 + \beta_4$	-.193	-.615	1.532	1.909	.918
	(.123)	(.693)	(1.16)	(1.216)	(2.735)
Mean dep. var.: $I_{ifp} = 0$	913.03	1192.06	896	3080.55	5215.63
Mean dep. var.: $I_{ifp} = 1$	930.77	2400.43	3101.71	5391.61	1249.95
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 6: Interaction of non-agricultural and agricultural shocks: Agri. / value added

	Sown area (1)	Grain (2)	Cash (3)	Primary va (4)	Secondary va (5)
Pred. emp.	-1.800	2.109	-1.697	-.892	-.818
	(.535)***	(2.195)	(.656)***	(.723)	(3.433)
Pred. emp. x I_{ifp}	1.626	-2.909	1.812	.977	6.113
	(.559)***	(2.263)	(.652)***	(.724)	(3.514)*
Lagged rainfall int.	.023	.146	-.047	.004	.424
	(.038)	(.112)	(.042)	(.039)	(.225)*
Lagged rainfall int. x I_{ifp}	-.012	-.127	.049	-.008	-.423
	(.035)	(.104)	(.039)	(.042)	(.231)*
$\beta_1 + \beta_2$	-.174	-.800	.115	.085	5.295
	(.114)	(.374)**	(.114)	(.163)	(1.143)***
$\beta_3 + \beta_4$	.011	.018	.002	-.004	.000
	(.014)	(.026)	(.007)	(.011)	(.074)
Mean dep. var.: $I_{ifp} = 0$	84.29	248.44	55.46	86.62	123.72
Mean dep. var.: $I_{ifp} = 1$	51.22	212.12	48.17	89.32	218.53
Obs.	6147	14277	13828	14343	14443

Notes: The dependent variables are sown area reported in thousands of hectares and production of grain and cash crops reported in thousands of tons, and value added reported in the primary and secondary sectors in thousands of yuan. All measures of value added are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 7: Specifications including dummy variables for extreme rainfall events

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Panel A: Rainfall above the 75th percentile					
Pred. emp.	5.804	60.679	14.995	-49.276	-49.699
	(10.629)	(45.015)	(16.134)	(47.956)	(133.274)
Pred. emp. x I_{ifp}	-11.008	-21.766	105.189	171.403	293.138
	(10.925)	(44.062)	(33.503)***	(52.887)***	(158.237)*
Lagged rainfall int.	-1.781	21.308	1.206	21.149	34.687
	(1.943)	(9.070)**	(2.751)	(12.632)*	(10.205)***
Lagged rainfall int. x I_{ifp}	1.891	-23.334	-10.435	-16.974	-68.495
	(2.096)	(9.083)**	(7.394)	(11.410)	(17.820)***
$\beta_1 + \beta_2$	-5.204	38.913	120.184	122.127	243.439
	(2.424)**	(13.061)***	(28.007)***	(23.435)***	(78.680)***
$\beta_3 + \beta_4$	.110	-2.025	-9.23	4.175	-33.808
	(.746)	(2.381)	(6.378)	(6.238)	(16.041)
Panel B: Rainfall below the 25th percentile					
Pred. emp.	4.529	73.067	14.796	-34.492	-29.104
	(10.814)	(48.951)	(16.107)	(47.001)	(137.575)
Pred. emp. x I_{ifp}	-9.775	-34.296	98.860	161.344	261.671
	(11.166)	(48.211)	(33.985)***	(51.916)***	(160.881)
Lagged rainfall int.	2.546	-5.745	1.945	-5.435	3.784
	(1.812)	(7.213)	(2.948)	(7.377)	(12.154)
Lagged rainfall int. x I_{ifp}	-2.646	3.850	.031	-8.679	-20.426
	(1.730)	(7.218)	(5.781)	(8.307)	(15.203)
$\beta_1 + \beta_2$	-5.246	38.771	113.656	126.853	232.567
	(2.448)**	(12.528)***	(28.664)***	(22.376)***	(77.489)***
$\beta_3 + \beta_4$	-.100	-1.895	1.976	-14.114	-16.642
	(.395)	(1.821)	(4.961)	(3.772)***	(9.461)*
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the mean of the high rainfall dummy observed one and two years prior. The high rainfall dummy in Panel A is defined equal to one if rainfall exceeds the 75th percentile observed in the prefecture over time; in Panel B, it is defined equal to one if rainfall exceeds the 90th percentile observed in the prefecture over time. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall (both the linear variable and the dummy variable) interacted with dummy variables for each decile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 8: Specifications including controls for current rainfall: GDP

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	Per capita GDP (5)
Pred. emp.	5.158	54.844	7.957	-55.648	-39.185
	(11.227)	(48.530)	(16.507)	(45.071)	(145.136)
Pred. emp. x I_{ifp}	-9.921	-14.497	101.343	176.019	271.790
	(11.438)	(48.186)	(30.971)***	(50.436)***	(167.425)
Current rainfall int.	.112	1.628	.954	4.097	-1.324
	(.442)	(1.377)	(.658)	(1.139)***	(2.919)
Current rainfall int. x I_{ifp}	-.146	-2.229	.023	-3.091	-.426
	(.414)	(1.535)	(.714)	(.987)***	(3.051)
Lagged rainfall int.	-.782	5.819	.714	8.935	4.739
	(.742)	(2.882)**	(.888)	(3.508)**	(5.555)
Lagged rainfall int. x I_{ifp}	.581	-6.585	1.096	-6.791	-4.140
	(.757)	(3.130)**	(1.587)	(3.357)**	(5.249)
$\beta_1 + \beta_2$	-4.763	40.347	109.299	120.371	232.605
	(2.357)*	(11.945)***	(25.446)***	(22.034)***	(76.45)***
$\beta_3 + \beta_4$	-.201	-.765	1.81	2.144	.599
	(.125)	(.669)	(1.325)	(1.234)*	(2.914)
Mean dep. var.: $I_{ifp} = 0$	913.03	1192.06	896	3080.55	5215.63
Mean dep. var.: $I_{ifp} = 1$	930.77	2400.43	3101.71	5391.61	1249.95
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; the interaction of predicted employment with current cultivation season rainfall; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including current and lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 9: Pre-trends

	Pop.	Perc. non-ag	Perc. employed	Perc. illiterate	Perc. under 14	Perc. over 65	Perc. women 15-49	Output per capita
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pred. emp.	914.239	.001	.006	.011	.003	.002	-.0002	-.010
	(51.647)***	(.002)	(.004)	(.007)	(.003)	(.0005)***	(.001)	(.044)
Obs.	5009	4032	2967	2969	2973	2975	2973	1492

Notes: The dependent variables are county covariates as reported in the censuses of 1953, 1964, 1982 and 1992, including county population, the percentage of the population that is non-agricultural, the percentage employed, the percentage illiterate, the percentage under 15, the percentage over 65, the percentage 15-49 that is female, and average output per capita. The county population is reported in all four census rounds; the percentage of the population that is non-agricultural is reported in 1964, 1982 and 1990; the other demographic variables are reported in 1982 and 1990 only, with average output per capita reported only in 1990.

The independent variable in Columns (1) through (7) is predicted employment in the secondary sector in the year of interest, calculated using the average Bartik shock observed in the period 2000-2010 and baseline employment in 2000. In Columns (1) through (7), All regressions include province-year fixed effects, county fixed effects, and both set of fixed effects interacted with the dummy variable I_{ifp}. Standard errors are estimated employing clustering at the county level. Asterisks indicate significance at the ten, five and one percent level.

Table 10: Cross-dependence on neighboring county shocks

	Primary	Secondary	Tertiary	GDP	Per capita GDP
	(1)	(2)	(3)	(4)	(5)
Panel A: Neighboring counties within 25 kilometers					
Pred. emp.	-4.359	23.688	60.449	121.447	277.132
	(3.005)	(13.964)*	(23.478)**	(26.018)***	(94.480)***
Neighbor pred. emp.	-.619	7.006	12.588	8.633	33.995
	(.860)	(4.178)*	(6.764)*	(6.496)	(16.154)**
Test $\beta_1 = \beta_2$	.166	.237	.05	.000	.006
Obs.	7488	8186	8587	12495	13211
Panel B: Neighboring counties within 50 kilometers					
Pred. emp.	-1.399	17.277	53.514	110.068	115.764
	(1.360)	(11.355)	(19.893)***	(23.441)***	(59.428)*
Neighbor pred. emp.	.134	.634	-1.943	-2.215	-1.583
	(.377)	(2.163)	(3.989)	(3.548)	(6.077)
Test $\beta_1 = \beta_2$	.287	.164	.007	.000	.049
Obs.	7482	8276	8658	12940	13488
Panel C: Neighboring counties within 75 kilometers					
Pred. emp.	-2.246	31.289	75.296	142.446	139.490
	(1.673)	(12.939)**	(28.019)***	(30.738)***	(65.723)**
Neighbor pred. emp.	.081	-2.494	-20.497	-.266	-5.924
	(.343)	(2.312)	(14.367)	(5.356)	(12.035)
Test $\beta_1 = \beta_2$	.179	.01	.002	.000	.025
Obs.	6089	6670	6978	10741	11240
Panel D: Neighboring counties within 100 kilometers					
Pred. emp.	1.028	31.921	50.127	145.749	71.510
	(2.247)	(18.213)*	(27.782)*	(34.382)***	(82.652)
Neighbor pred. emp.	1.266	5.623	4.009	8.792	26.183
	(.539)**	(3.983)	(6.429)	(6.635)	(14.624)*
Test $\beta_1 = \beta_2$	.907	.106	.054	.000	.591
Obs.	4429	4761	4886	7778	7931

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and predicted employment in the closest neighboring county within the same prefecture in the specified distance range. Distances are calculated using straight-line distances between county centroids.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 11: Interaction specifications including neighboring county shocks

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	Per capita GDP (5)
Pred. emp.	-6.954	79.650	21.319	-15.441	-4.719
	(7.330)	(44.473)*	(16.513)	(47.447)	(14.062)
Pred. emp. x I_{ifp}	2.967	-57.750	36.860	123.814	20.167
	(7.744)	(43.575)	(26.776)	(51.546)**	(14.498)
Lagged rainfall int.	-.471	5.765	.176	6.430	.123
	(.353)	(2.766)**	(.755)	(2.762)**	(.444)
Lagged rainfall int. x I_{ifp}	.301	-6.800	-.790	-4.731	-.391
	(.303)	(2.707)**	(1.192)	(2.283)**	(.483)
Neighbor pred. emp.	2.185	1.168	-1.056	-2.017	-1.707
	(1.149)*	(3.154)	(2.444)	(2.581)	(.668)**
Neighbor pred. emp. x I_{ifp}	-2.233	.873	2.881	5.614	2.274
	(1.201)*	(3.516)	(3.531)	(4.033)	(1.109)**
Mean dep. var.: $I_{ifp} = 0$	712.08	1197.46	877.88	2938.07	5155.51
Mean dep. var.: $I_{ifp} = 1$	703.35	2280.81	2409.28	5033.29	11736.69
Obs.	8714	9606	10025	15024	13918

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The observed shock in the closest neighboring county within 50 kilometers, and the interaction of this shock with the dummy variable I_{ifp}, are also included. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 12: Overestimation of GDP at the province-year level

	Overestimate fraction					
	(1)	(2)	(3)	(4)	(5)	(6)
Pred. emp.	.013		.027		.009	.033
	(.042)		(.107)		(.042)	(.097)
Pred. emp. x I_{ifp}			-.016			-.027
			(.138)			(.127)
Lagged rainfall		-.025		-.068	-.015	-.124
		(.035)		(.061)	(.042)	(.107)
Lagged rainfall x I_{ifp}				.047		.121
				(.073)		(.131)
Lagged rainfall int.					.002	-.006
					(.002)	(.007)
Lagged rainfall int. x I_{ifp}						.009
						(.009)
Obs.	209	209	209	209	209	209

Notes: The dependent variable is the ratio of the sum of county-level GDP to the reported total GDP in a given province and year. The independent variables are calculated as means at the provincial-year level, and include predicted employment, average cultivation season rainfall observed one and two years prior, and the interaction of the two. All three variables are also interacted with the province-year mean of the a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. All specifications include province and year fixed effects and year fixed effects interacted with I_{ifp}. Asterisks indicate significance at the ten, five and one percent level.

Table 13: Evaluating heterogeneity with respect to export intensity

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Pred. emp.	6.832	76.332	-18.547	13.974	62.263
	(11.326)	(48.201)	(25.708)	(49.996)	(164.529)
Pred. emp. x I_{ifp}	-9.931	-1.424	70.067	174.791	269.191
	(11.528)	(47.888)	(35.752)*	(47.871)***	(173.117)
Pred. emp. x Exp_{ifp}	-1.974	-37.858	70.899	-78.188	-124.439
	(3.591)	(25.102)	(41.245)*	(38.963)**	(101.746)
Lagged rainfall int.	-.654	6.050	1.130	8.003	-6.574
	(.608)	(2.321)***	(1.426)	(3.033)***	(6.196)
Lagged rainfall int. x I_{ifp}	.540	-4.387	.827	-4.577	-6.479
	(.674)	(1.940)**	(1.511)	(2.442)*	(4.719)
Lagged rainfall int. x Exp_{ifp}	-.065	-3.113	-1.549	-2.047	17.288
	(.286)	(1.246)**	(2.626)	(2.249)	(7.974)**
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise, and a dummy variable Exp_{ifp}, equal to zero if a county is initially in the lowest quantile of the average export fraction in the secondary sector (exports as a percentage of total production) and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment, as well as the lagged rainfall - initial employment quantile - high export interactions; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 14: Evaluating heterogeneity with respect to within-country border effects

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	Per capita GDP (5)
Pred. emp.	5.817	60.584	11.397	-44.441	-57.453
	(10.967)	(47.147)	(17.977)	(45.965)	(142.767)
Pred. emp. x I_{ifp}	-11.196	-19.086	104.560	168.903	246.354
	(11.194)	(46.700)	(33.353)***	(51.129)***	(164.348)
Pred. emp. x $Border_{ifp}$	-1.242	-18.450	-79.150	-10.691	233.438
	(2.846)	(13.208)	(31.990)**	(23.428)	(103.379)**
Pred. emp. int.	-.734	4.735	.025	7.046	5.081
	(.590)	(2.471)*	(.738)	(2.898)**	(4.735)
Pred. emp. int. x I_{ifp}	.649	-5.525	1.465	-5.226	-4.754
	(.589)	(2.537)**	(1.441)	(2.651)**	(4.394)
Pred. emp. int. x $Border_{ifp}$	.310	-.789	-1.721	-.131	-.237
	(.088)***	(.521)	(1.191)	(.772)	(2.870)
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise, and $Border_{ifp}$, a variable capturing the estimated within-country border effect for the county of interest. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment, as well as the lagged rainfall - initial employment quantile - border effect interactions; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table 15: Local investment response to rainfall shocks

	Capital investment (1)	Gov. exp. (2)
Current rainfall	-.072 (.044)	.183 (.094)*
Current rainfall x I_{ifp}	-.151 (.056)***	-.255 (.089)***
Lagged rainfall	-.040 (.101)	.191 (.103)*
Lagged rainfall x I_{ifp}	-.065 (.170)	-.184 (.088)**
Mean dep. var.: $I_{ifp} = 0$	2.22	6.05
Mean dep. var.: $I_{ifp} = 1$	3.41	7.18
Obs.	4998	14288

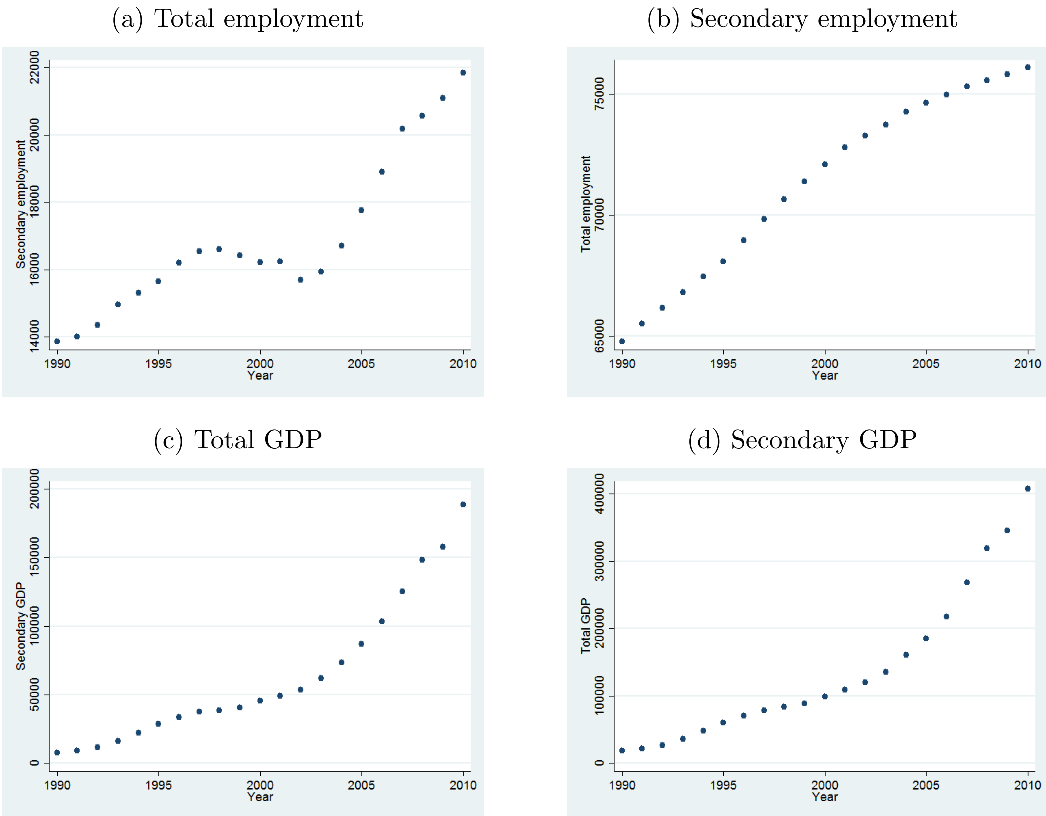
Notes: The independent variables are current and lagged rainfall in the region-specific cultivation season interacted with I_{ifp}, a dummy variable for counties that are initially industrialized. All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. The dependent variables are private capital investment in Column (1), and local government expenditure in Column (2). Asterisks indicate significance at the ten, five and one percent level.

A Appendix Figures and Tables

Figure 1: Aggregate series over time

(a) Total employment (b) Secondary employment

(c) Total GDP (d) Secondary GDP



Notes: These graphs show the trends over time for secondary employment, total employment, secondary GDP and total GDP for China as a whole' over the period 1990-2010.

Table A1: Heterogeneity by region and initial conditions

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita
Panel A: Heterogeneity by region					
Pred. emp.	-3.382	20.645	10.020	131.285	200.779
(North / Northeast)	(2.702)	(18.399)	(25.256)	(51.439)**	(105.046)*
Pred. emp.	-6.775	41.069	126.049	118.437	285.837
(South / East)	(2.841)**	(13.188)***	(30.405)***	(22.976)***	(93.314)***
Pred. emp.	7.059	53.531	95.737	130.127	-163.251
(Southwest / Northwest)	(3.539)**	(35.847)	(41.091)**	(54.271)**	(129.975)
Test $\beta_1 = \beta_2$	.352	.322	.001	.811	.496
Test $\beta_1 = \beta_3$	.007	.400	.025	.988	.022
Test $\beta_2 = \beta_3$	.001	.720	.532	.821	.001
Panel B: Heterogeneity by degree of initial specialization					
Pred. emp.	-3.740	48.268	158.767	135.704	185.465
	(3.329)	(17.421)***	(48.281)***	(29.817)***	(112.202)*
Pred. emp.	-2.941	-11.567	-74.434	-21.298	80.293
x init. highly spec.	(4.070)	(18.723)	(48.577)	(33.280)	(122.599)
Panel C: Heterogeneity by size of the local economy					
Pred. emp.	-6.590	39.943	299.430	120.871	523.681
	(5.495)	(23.232)*	(72.054)***	(35.749)***	(224.670)**
Pred. emp.	-.959	-2.855	-206.230	-12.354	-315.556
x init. high emp.	(5.676)	(24.174)	(71.922)***	(36.809)	(223.786)
Panel D: Heterogeneity by initial educational levels					
Pred. emp.	-3.161	24.448	18.188	57.297	-4.334
	(2.518)	(14.613)*	(27.050)	(21.674)***	(59.856)
Pred. emp.	-3.467	27.361	149.086	97.492	329.393
x initial high educ.	(2.976)	(15.864)*	(34.149)***	(23.937)***	(59.666)***
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variable is predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector. All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}.

In Panel A, the predicted employment variable is interacted with three dummy variables for different regions: the north and northeast region includes Hebei, Shanxi, Liaoning, Jilin and Heilongjiang; the south and east region includes Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan and Guangdong; and the southwest and northwest region includes Sichuan, Guizhou, Yunnan, Shaanxi, Gansu and Qinghai. In Panel B, the predicted employment variable is interacted with a dummy variable equal to one if the county is characterized by an initially high level of specialization (more specifically, if one or more secondary subsectors constitutes more than 20% of secondary employment). In Panel C, the predicted employment variable is interacted with a dummy variable equal to one if the county is characterized by initially high employment (above the median). In Panel D, the predicted employment variable is interacted with a dummy variable equal to one if the county is characterized by initially high proportion of population with education above primary school (proportion above the median). Asterisks indicate significance at the ten, five and one percent level.

Table A2: Secondary employment shocks at destination provinces

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	Per capita GDP (5)
Pred. emp.	-87.548	1008.819	2638.360	3251.866	5662.870
	(62.722)	(354.033)***	(703.183)***	(606.959)***	(1992.073)***
Destination shock	-328.827	493.785	2645.123	-296.004	3762.707
	(76.813)***	(415.991)	(879.033)***	(500.848)	(2371.763)
Mean dep. var.	927.88	2064.32	2522.82	4849.13	10482.49
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector, and the estimated predicted employment shock at the migration destination. The migration destination shock is calculated as the weighted average of the Bartik shock observed in each province reported as a migration destination for county residents in the 2000 census; the weights are estimated as the proportion of total outmigrants from that county who migrated to each province. (This exercise includes the average province shock in the home province, capturing migration to destinations outside the home county, but within the same province.)
All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Asterisks indicate significance at the ten, five and one percent level.

Table A3: Specifications including interaction terms for post rainfall shocks

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	Per capita GDP (5)
Pred. emp.	12.024	46.709	4.984	-40.684	-34.209
	(11.957)	(42.801)	(16.296)	(42.084)	(123.876)
Pred. emp. x I_{ifp}	-16.969	-18.457	96.295	153.707	251.118
	(12.383)	(43.070)	(34.003)***	(49.469)***	(161.719)
Post rainfall int.	.029	-.026	.008	.003	.017
	(.007)***	(.028)	(.013)	(.025)	(.059)
Post rainfall int. x I_{ifp}	-.028	.032	.012	.005	-.041
	(.007)***	(.030)	(.016)	(.024)	(.055)
Obs.	9039	9200	9379	13633	14302

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years after the shock. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including the post-period rainfall shock interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table A4: Sample excluding highly specialized prefectures

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Pred. emp.	19.503	81.956	25.496	-25.386	-6.220
	(12.244)	(60.323)	(19.151)	(55.305)	(16.064)
Pred. emp. x I_{ifp}	-24.935	-51.270	92.693	127.942	18.523
	(12.098)**	(60.377)	(39.245)**	(61.646)**	(16.757)
Lagged rainfall int.	-.972	5.208	-.308	5.936	-.036
	(.652)	(2.770)*	(.792)	(3.250)*	(.523)
Lagged rainfall int. x I_{ifp}	.742	-6.353	3.682	-5.191	-.104
	(.654)	(3.166)**	(1.789)**	(3.474)	(.640)
$\beta_1 + \beta_2$	-5.432	30.686	118.189	102.555	24.236
	(2.189)**	(14.699)**	(32.448)***	(25.978)***	(70.792)
$\beta_3 + \beta_4$	-.230	-1.145	3.374	.745	1.732
	(.161)	(.865)	(1.545)**	(1.529)	(2.955)
Mean dep. var.: $I_{ifp} = 0$	916.31	1133.75	878.45	3014.18	5410.03
Mean dep. var.: $I_{ifp} = 1$	952.87	2279.09	2974.17	5120.45	12074.34
Obs.	7986	8037	8162	12179	11268

Notes: The sample is restricted to prefectures in which no secondary subsector constitutes more than 10% of national employment in that subsector. The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table A5: Sample excluding highly specialized provinces

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Pred. emp.	5.291	71.217	9.900	-27.312	-8.602
	(10.824)	(48.577)	(16.985)	(47.455)	(15.542)
Pred. emp. x I_{ifp}	-6.642	-30.920	139.413	142.721	11.201
	(11.507)	(48.555)	(39.020)***	(53.858)***	(16.601)
Lagged rainfall int.	-.815	5.311	.576	7.440	.720
	(.600)	(2.746)*	(.855)	(3.864)*	(.635)
Lagged rainfall int. x I_{ifp}	.394	-5.152	3.325	-7.060	-1.105
	(.617)	(2.790)*	(2.158)	(4.100)*	(.733)
$\beta_1 + \beta_2$	-1.351	40.297	149.313	115.41	-40.282
	(2.457)	(18.508)**	(34.453)***	(26.264)***	(107.075)
$\beta_3 + \beta_4$	-.421	.159	3.901	.380	1.853
	(.168)**	(.638)	(1.917)**	(1.252)	(3.806)
Mean dep. var.: $I_{ifp} = 0$	901.83	1144.24	862.27	2772.47	5519.49
Mean dep. var.: $I_{ifp} = 1$	867.02	2210.4	2572.98	4668.79	12473.29
Obs.	7783	7794	7864	11063	10381

Notes: The sample is restricted to provinces in which no secondary subsector constitutes more than 75% of national employment in that subsector. The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table A6: Employing Bartik shocks estimated without the largest subsectors

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Pred. emp.	11.782	142.556	-3.642	22.503	49.188
	(21.207)	(92.203)	(32.231)	(29.609)	(41.302)
Pred. emp. x I_{ifp}	-12.279	-41.606	227.740	245.191	139.434
	(21.987)	(91.013)	(64.762)***	(56.488)***	(146.989)
Lagged rainfall int.	-1.065	5.112	.012	5.109	3.765
	(.627)*	(3.070)*	(.799)	(2.530)**	(3.261)
Lagged rainfall int. x I_{ifp}	.915	-5.167	1.403	-1.054	-1.089
	(.601)	(3.087)*	(1.407)	(2.211)	(3.412)
$\beta_1 + \beta_2$	-.497	100.95	224.098	267.694	188.623
	(4.514)	(30.837)***	(63.228)***	(52.474)***	(133.683)
$\beta_3 + \beta_4$	-.150	-.055	1.416	4.055	2.675
	(.145)	(.697)	(1.148)	(1.576)**	(3.346)
Mean dep. var.: $I_{ifp} = 0$	912.51	1189.67	895.74	3079.97	5495.82
Mean dep. var.: $I_{ifp} = 1$	926.38	2362.64	3042.53	5302.72	12240.71
Obs.	9858	10013	10222	15170	15840

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment; I also include differential trends for counties in each decile of overall average rainfall. Bartik shocks for this specification are constructed excluding the two largest subsectors (by employment) in each county at baseline.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table A7: Cross-dependence on within-province neighboring county shocks

	Primary	Secondary	Tertiary	GDP	GDP per capita
	(1)	(2)	(3)	(4)	(5)
Panel A: Neighboring counties within 25 kilometers					
Pred. emp.	-4.703	26.011	60.024	110.610	237.899
	(2.796)*	(11.806)**	(20.486)***	(22.857)***	(80.969)***
Neighbor pred. emp.	-.090	-.010	1.820	3.149	12.175
	(.424)	(1.985)	(2.646)	(2.635)	(7.433)
$\beta_1 = \beta_2$	.086	.031	.006	.000	.004
Obs.	8665	9546	9966	14891	15611
Panel B: Neighboring counties within 50 kilometers					
Pred. emp.	-4.738	24.652	59.441	108.741	239.385
	(2.828)*	(11.753)**	(20.256)***	(22.652)***	(82.727)***
Neighbor pred. emp.	-.004	2.213	1.865	2.459	1.830
	(.188)	(1.829)	(3.440)	(2.617)	(6.020)
$\beta_1 = \beta_2$	.095	.059	.005	.000	.004
Obs.	8714	9606	10025	15024	15745
Panel C: Neighboring counties within 75 kilometers					
Pred. emp.	-4.764	27.805	60.063	112.048	245.057
	(2.849)*	(11.644)**	(20.438)***	(22.915)***	(82.079)***
Neighbor pred. emp.	-.083	-1.305	3.089	-1.725	5.345
	(.214)	(1.422)	(1.595)*	(2.511)	(6.042)
$\beta_1 = \beta_2$	.100	.012	.005	.000	.003
Obs.	8652	9550	9975	14958	15678
Panel D: Neighboring counties within 100 kilometers					
Pred. emp.	-4.493	25.845	61.261	110.870	245.315
	(2.650)*	(11.905)**	(20.608)***	(22.874)***	(79.776)***
Neighbor pred. emp.	1.029	-1.873	1.491	-1.882	-39.165
	(.574)*	(1.238)	(2.264)	(1.859)	(17.916)**
$\beta_1 = \beta_2$	.067	.026	.003	.000	.002
Obs.	8668	9545	9963	14964	15684

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and predicted employment in the closest neighboring county within the same province in the specified distance range. Distances are calculated using straight-line distances between county centroids.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table A8: Spillovers from fast-growing to slow-growing counties

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	GDP per capita (5)
Pred. emp.	5.617	56.847	115.593	176.980	328.850
	(4.065)	(23.066)**	(67.520)*	(47.002)***	(138.301)**
High shock	2.050	-3.457	-12.230	-4.761	32.460
	(1.130)*	(5.273)	(13.317)	(7.377)	(25.859)
Obs.	3144	3247	3213	4930	5072

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the average predicted employment shock in the fastest-growest quantile of counties within the same prefecture. The sample is restricted to the slowest-growing quantile of counties within each prefecture.

All regressions include province-year fixed effects, county fixed effects, differential trends for counties in different quantiles of initial non-agricultural employment, differential trends for counties in different quantiles of initial (absolute) total employment, and differential trends for counties in different deciles of average rainfall; the province-year fixed effects and trends for counties at different levels of rainfall are also interacted with the dummy variable I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

Table A9: Heterogeneity with respect to internal and international trade

	Primary (1)	Secondary (2)	Tertiary (3)	GDP (4)	Per capita GDP (5)
Pred. emp.	6.574	75.803	-24.178	9.079	55.073
	(11.270)	(48.908)	(30.833)	(51.338)	(166.317)
Pred. emp. x I_{ifp}	-9.955	-2.566	66.003	180.877	261.559
	(11.492)	(48.742)	(39.526)*	(49.470)***	(175.525)
Pred. emp. x $Border_{ifp}$	-1.142	-15.193	-83.013	-4.560	240.057
	(2.831)	(12.264)	(32.141)***	(22.424)	(104.257)**
Pred. emp. x Exp_{ifp}	-2.110	-35.644	81.938	-79.249	-153.276
	(3.501)	(24.905)	(46.270)*	(40.390)**	(99.686)
Lagged rainfall int.	-.630	5.148	.198	6.865	-6.735
	(.570)	(2.421)**	(1.473)	(3.264)**	(6.032)
Lagged rainfall int. x I_{ifp}	.650	-4.875	1.114	-4.256	-4.171
	(.622)	(2.160)**	(1.654)	(2.512)*	(5.192)
Lagged rainfall int. x $Border_{ifp}$	.310	-.759	-1.638	-.098	-.546
	(.086)***	(.547)	(1.185)	(.790)	(2.964)
Lagged rainfall int. x Exp_{ifp}	-.113	-1.191	.071	-.622	13.355
	(.264)	(1.177)	(2.563)	(2.109)	(6.235)**
Obs.	9913	10052	10284	15255	15926

Notes: The dependent variables are GDP by sector and total GDP reported in millions of yuan, and GDP per capita reported in yuan. All measures of GDP are deflated employing the World Bank GDP deflator. The independent variables are predicted employment based on sectoral composition and the observed growth rates in employment by industrial sector; and the interaction of predicted employment with the average of cultivation season rainfall observed one and two years prior. Both of these variables are also interacted with a dummy variable I_{ifp}, equal to zero if a county is initially in the lowest quantile of non-agricultural employment as a proportion of total employment, and one otherwise; a dummy variable Exp_{ifp}, equal to zero if a county is initially in the lowest quantile of the average export fraction in the secondary sector (exports as a percentage of total production) and one otherwise; and a measure of the county-specific internal border effect, $Border_{ifp}$. The specification controls flexibly for direct effects of rainfall by including lagged rainfall interacted with dummy variables for each quantile of initial non-agricultural employment, as well as lagged rainfall interacted with both Exp_{ifp} and $Border_{ifp}$; I also include differential trends for counties in each decile of overall average rainfall.

All regressions include province-year fixed effects, county fixed effects, and differential trends for counties in different quantiles of initial secondary employment; the specification also includes the interaction between province-year fixed effects and secondary quantile trends and a dummy for initially industrialized counties, I_{ifp}. Standard errors are estimated employing two-way clustering at the county and year level. Asterisks indicate significance at the ten, five and one percent level.

B Matching the county census to county yearbooks

The 2000 county census included data on 2,873 counties; when restricted to the provinces of interest, there are 2,348 counties. There are two major types of anomalies that arise when merging the counties as reported in the census to the provincial yearbook county-level data. Both primarily, though not exclusively, relate to how data is reported for the county-level units of larger urban centers.

First, some county-level units reported in the census cannot be matched to provincial yearbooks. This sometimes reflects changes in county names, or omissions in yearbooks that do not follow a clear pattern. However, the most frequently observed case is county-level units that correspond to the urbanized sections of prefecture-level cities, for which information is not reported in provincial yearbooks in some provinces. This problem is most notable in Hebei, Heilongjiang, Henan, and Liaoning.[24]

Second, in some cases multiple county-level units in the census are matched to a single, higher-level geographic unit in the census. More specifically, this is typically a prefecture-level city for which each constituent county-level unit has separately reported data in the census, but the yearbook reports data only for the city as a whole. Accordingly, the data can be matched, but it is not a one-to-one match. This anomaly is primarily observed in Fujian, Guangdong, Hubei, Jilin, Jiangsu, and Zhejiang.[25]

In the whole sample, I am able to match 88% of the counties reported in the 2000 census to provincial yearbook data; if the provinces with the largest number of mismatches (again, Hebei, Heilongjiang, Henan, and Liaoning) are excluded, the match rate increases to 93%. Focusing on the sample of counties that are matched, 7% are part of a multiple match in which several census counties are matched to one prefecture-level city. These are all drawn from the six provinces specified in the previous paragraph; in those six provinces, 18% of county-level units are part of a multiple match. (For counties that are part of a multiple match, I treat the prefecture-level city as a unit for the purposes of clustering, and re-code all constituent counties as if they were part of a single unit.)

Table A10 reports the average number of census mismatches and multiple matches per province. Note these measures are defined with respect to the sample of counties that are observed in the 2000 census. There are also some counties for which information is reported in provincial yearbooks that cannot be matched to the census; I denote these "yearbook only" counties, and report the fraction of total counties observed in the yearbook that are not matched to the census.

In general, the fraction of counties reported only in the yearbook is higher than the fraction of counties reported only in the census, consistent with the intuition that the data in the census is generally higher-quality than the data in the provincial yearbooks. The number of counties only observed in the yearbook is also inflated by the fact that

[24] In all four of these provinces, the number of county-level geographic units for which information is reported in the provincial yearbooks is considerably lower than the total number of county-level units reported in the census; the missing counties are those that are part of the urbanized sections of large prefecture cities. There are certainly other provinces in which some counties observed in the census are missing in provincial yearbook; Henan, Jiangxi and Jiangsu all have a relatively large number of missing counties in the yearbooks. However, in these cases the missing counties do not follow a clear pattern, and do not seem to correspond to units of larger cities.

[25] The primary results are also robust to excluding the counties that cannot be matched to the census one to one.

the analysis employs data from provincial yearbooks over more than ten years. Thus if a given county's name is reported correctly in nine years, but reported incorrectly in one year, this error would be counted as one example of a "yearbook only" county, despite the fact that the observations from other years have been matched correctly. (Obvious spelling errors or variations in county names were taken into account prior to conducting this exercise.)

Table A11 reports the results from a series of simple specifications regressing county covariates as reported in the 2000 census (including total population, household size, the fraction of population that is non-agricultural, and employment in the primary, secondary and tertiary sectors) on a dummy for the county being missing in the provincial yearbook data; the dependent variables are normalized to have mean zero and standard deviation one. It is evident that the missing counties have lower total population, and are characterized by smaller household size and a greater concentration in non-agricultural activities. Primary employment is significantly lower, and secondary and tertiary employment are somewhat higher in absolute numbers. These results are also consistent with the previously discussed evidence that the missing counties are disproportionately urban, and are omitted primarily because some provinces opt not to report county-level information for county units that are part of larger cities.

Table A10: County merging by province

Province	Proportion mismatches	Proportion multiple matches	Proportion yearbook only counties
Anhui	.132	0	.317
Fujian	0	.310	0
Gansu	.023	0	.140
Guangdong	.040	.258	.108
Guizhou	0	0	.194
Hebei	.214	0	.062
Heilongjiang	.485	0	.058
Henan	.323	0	.050
Hubei	.119	.139	.100
Hunan	.033	0	.041
Jiangsu	.148	.704	.023
Jiangxi	.192	0	.050
Jilin	.050	.217	.055
Liaoning	.290	0	.057
Qinghai	0	0	.07
Shaanxi	.009	0	.264
Shandong	.014	0	.221
Shanxi	.042	0	.269
Sichuan	.078	0	.050
Yunnan	.031	0	.164
Zhejiang	.034	.213	.039

Notes: This table reports province-level statistics on the merge between the 2000 census data and the provincial yearbooks. "Proportion mismatches" reports the proportion of counties observed in the 2000 census that are not observed in provincial yearbooks. "Proportion multiple matches" reports the proportion of counties observed in the 2000 census for which there is a many to one merge in the provincial yearbooks; this is typically observed when multiple counties in the census match to data from one provincial-level city in the yearbooks. "Proportion yearbook only counties" reports the proportion of total counties observed in the provincial yearbooks not matched to the 2000 census.

Table A11: Characteristics of missing counties

	Total pop.	Hh size	Prop non-ag.	Primary employment	Secondary employment	Tertiary employment
	(1)	(2)	(3)	(4)	(5)	(6)
Missing dummy	-.429***	-.708***	1.709***	-.865***	.115	.276*
	(.137)	(.120)	(.173)	(.099)	(.077)	(.165)
Obs.	2348	2348	2348	2348	2348	2348

Notes: This table reports the results from a series of regressions including the specificed county covariate as reported in the 2000 census, regressed on a dummy variable equal to one if a county is not observed in the provincial yearbooks, and zero otherwise. Standard errors are clustered at the provincial level.

C Sample composition by variable

Table A12: Employment data: Provinces and years reported

Province	Primary		Secondary		Tertiary		Agricultural		Total	
Hebei	.	.	.	.	.	.	2003	2010	.	.
Shanxi	.	.	.	.	.	.	2001	2010	.	.
Liaoning	.	.	2003	2010	2003	2010	2001	2010	.	.
Jilin	.	.	.	.	.	.	2003	2010	.	.
Heilongjiang	.	.	.	.	.	.	2002	2010	.	.
Jiangsu	.	.	.	.	.	.	2001	2010	.	.
Zhejiang	2005	2010	2005	2010	2005	2010	2003	2010	2005	2010
Anhui	.	.	.	.	.	.	2003	2010	.	.
Fujian	.	.	.	.	.	.	2001	2010	.	.
Jiangxi	.	.	.	.	.	.	2002	2010	.	.
Shandong	.	.	.	.	.	.	2003	2010	.	.
Henan	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Hubei	2001	2006	2001	2006	2001	2006	2003	2010	2001	2006
Hunan	.	.	.	.	.	.	2003	2010	.	.
Guangdong	.	.	.	.	.	.	2010	2010	.	.
Sichuan	2001	2010	2001	2010	2001	2010	2004	2010	2001	2010
Guizhou	.	.	.	.	.	.	2003	2010	.	.
Yunnan	.	.	.	.	.	.	2001	2010	.	.
Shaanxi	.	.	.	.	.	.	2003	2010	.	.
Gansu	.	.	.	.	.	.	2001	2010	.	.
Qinghai	.	.	.	.	.	.	2001	2010	.	.

Notes: This table reports the first and last year each variable is reported for each province. Missing values indicate the variable is not reported at all for a given province.

Table A13: GDP data: Provinces and years reported

Province	Primary		Secondary		Tertiary		Total		Per capita	
Hebei	2006	2010	2006	2010	2006	2010	2001	2010	.	.
Shanxi	2001	2007	2001	2007	2001	2007	2001	2010	2001	2007
Liaoning	.	.	.	.	.	.	2001	2010	.	.
Jilin	2001	2009	2001	2009	2001	2009	2001	2010	2001	2009
Heilongjiang	2002	2007	2002	2007	2002	2007	2001	2010	2002	2007
Jiangsu	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Zhejiang	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Anhui	2002	2010	2002	2010	2002	2010	2001	2010	2002	2010
Fujian	2001	2010	2001	2010	2001	2010	2001	2010	2002	2010
Jiangxi	2002	2010	2002	2010	2002	2010	2002	2010	.	.
Shandong	.	.	.	.	.	.	2001	2010	.	.
Henan	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Hubei	2001	2005	2001	2005	2001	2005	2001	2010	.	.
Hunan	2002	2010	2002	2010	2002	2010	2001	2010	2002	2010
Guangdong	2001	2010	2001	2010	2001	2010	2001	2010	.	.
Sichuan	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Guizhou	2004	2009	2004	2009	2004	2009	2001	2010	2001	2003
Yunnan	2001	2007	2001	2007	2001	2007	2001	2010	.	.
Shaanxi	.	.	.	.	.	.	2001	2010	.	.
Gansu	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Qinghai	.	.	.	.	.	.	2001	2010	.	.

Notes: This table reports the first and last year each variable is reported for each province. Missing values indicate the variable is not reported at all for a given province.

Table A14: Value added and agricultural outcomes: Provinces and years reported

Province	Sown area		Grain		Cash		Value added Primary		Value added Secondary	
Hebei	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Shanxi	2001	2007	2001	2010	2001	2010	2001	2010	2001	2010
Liaoning	2003	2010	2001	2010	2001	2010	2003	2010	2001	2010
Jilin	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Heilongjiang	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Jiangsu	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Zhejiang	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Anhui	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Fujian	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Jiangxi	2002	2010	2002	2010	2002	2010	2002	2010	2002	2010
Shandong	.	.	2001	2004	2001	2010	2001	2010	2001	2010
Henan	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Hubei	2001	2001	2001	2010	2001	2010	2001	2010	2001	2010
Hunan	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Guangdong	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Sichuan	2001	2005	2001	2010	2001	2010	2001	2010	2001	2010
Guizhou	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Yunnan	.	.	2001	2010	2001	2010	2001	2010	2001	2010
Shaanxi	2008	2009	2001	2010	2001	2010	2001	2010	2001	2010
Gansu	2001	2010	2001	2010	2001	2010	2001	2010	2001	2010
Qinghai	2001	2008	2001	2010	2001	2010	2001	2010	2001	2010

Notes: This table reports the first and last year each variable is reported for each province. Missing values indicate the variable is not reported at all for a given province.