Positive Rainfall Shocks, Overoptimism, and Agricultural Inefficiency in China

Kaixing Huang[a*], Jingyuan Guo[b,c], Da Zhao[d]

a. China Center for Agricultural Policy, Peking University

b. Department of Economics, National University of Singapore

c. Department of Finance and IGIER, Bocconi University

d. School of Economics, Sichuan University

August 23, 2023

Abstract

This study identifies overoptimism of farmers as an important cause of factor misallocation and inefficiency in agriculture. Annual deviation of rainfall from the local normal is exogenous, unanticipated, and transitory, but we find that farmers substantially adjust labor and land allocation in response to a lagged positive rainfall shock. By examining the response of more than 10,000 farmers in China over 11 years, we show that a lagged positive rainfall shock significantly reallocates labor from high-income off-farm work to low-income farm work, reallocates farmland from high-productivity farmers to low-productivity farmers, and reduces the average rural income by 8.1 percent. We also found that these effects are primarily driven by the irrational responses of low-productivity farmers and that farms with good irrigation conditions are generally not damaged.

Keywords: Rainfall shocks, irrationality, overoptimism, factor misallocation, agricultural efficiency

JEL: Q54, Q1, Q15, J43

---

*Kaxing Huang (corresponding author) is at the China Center for Agricultural Policy, Peking University, Yiheyuan Road, Haidian District, Beijing 100871, China (Kaixing.huang@pku.edu.cn). Jingyuan Guo (joint first and joint corresponding author) is at the Department of Economics, National University of Singapore, 1 Arts Link AS2 117570, Singapore (e1144076@u.nus.edu). Da Zhao is at the School of Economics, Sichuan University, Chengdu 610065, China (13210680123@fudan.edu.cn).
We are grateful to the editor and two anonymous referees for their insightful comments and suggestions. All errors are ours. The work is funded by NSFC72273001, 2021YFC3200503, NSFC72348003, Mitigate+: Research for Low Emissions Food Systems, NSFC72104163, and 21XJC790016. Data: https://doi:10.7910/DVN/FBGGQWversion=DRAFT.

1 Introduction

Global climate change is accompanied by substantial increases in the variability of rainfall (O'Gorman & Schneider 2009, Trenberth 2011, Feng et al. 2013). Many studies have established the significant damage of negative rainfall shocks on agricultural outputs and examined the subsequent adverse effects on, for example, savings (Paxson 1992), consumption (Kazianga & Udry 2006), investment (Grabrucker & Grimm 2020), human capital (Colmer 2020), and conflicts (Gatti et al. 2021). Complementing these studies, this article investigates a novel and generally ignored channel for rainfall shocks to impact agricultural efficiency and rural welfare: the factor misallocation caused by the irrational and overoptimistic responses of farmers to lagged positive rainfall shocks.

A stylized fact frequently employed in the economic literature is that year-to-year variation in rainfall (i.e., rainfall shock) is exogenous, unanticipated, and transitory (e.g., Paxson 1992, Miguel et al. 2004, Kazianga & Udry 2006, Brückner & Gradstein 2013, Kaur 2019).[1] Rational farmers should recognize that future rainfall shocks cannot be predicted based on past rainfall shocks. Therefore, the behavior change of farmers in response to past rainfall shocks could reflect their irrationality. For example, a positive rainfall shock in the last year may lead irrational farmers to overestimate the chance of above-normal rainfall this year and thus adjust their labor and land inputs accordingly. By examining the response of farmers to lagged rainfall shocks, we can verify their irrationality and estimate their welfare loss from it.

To guide our empirical analysis, we developed a theoretical model to illustrate the factor misallocation and welfare loss caused by the irrational response of farmers to a lagged positive rainfall shock. The model assumes a representative Chinese village in which farmers are endowed with the same labor and farmland but with different productivity levels. In each period, each farmer determines his allocation of land and labor in the next period based on his predicted agricultural productivity, which, in turn, depends on his predicted rainfall, to maximize the total income from agriculture, off-farm work, and land rent. The model predicts that if farmers are at least partially irrational and have a tendency to be overoptimistic, a lagged positive rainfall shock shifts farmland from high-productivity farmers to low-productivity farmers, shifts labor from high-income off-farm work to low-income farm work, and reduces the village's aggregate income. The impact could be aggravated if farmers with lower agricultural productivity are more likely to be irrational.

We tested these theoretical predictions based on data from more than 10,000 house-

[1] See Dell et al. (2014) for a review.

holds in 151 Chinese villages during 2003–2013. We followed the literature to define the rainfall outcome as a positive (negative) rainfall shock if the rainfall in the first month of the growing season is 1-standard-deviation above (below) the long-term average. We found that there are substantial variation in the village-level rainfall shocks and that the shocks are uncorrelated over time and across our sample villages. Defining rainfall shocks based on the rainfall in the first month of the growing season, rather than considering the rainfall for the entire growing season, reduces concerns about a lagged rainfall shock directly impacting agricultural productivity through alterations in soil moisture in the following year (i.e., the transitory effect of the rainfall shock).

We estimated a regression model that uses, as explanatory variables, nine interactions between dummies of three types of rainfall shock (i.e., positive, none, or negative) in the last year and this year. The reference group omitted from the regression is the interaction of no shock in both years. The coefficient of the interaction of a positive shock last year and no shock this year most clearly captures the effect of a lagged positive rainfall shock. We estimated that a lagged positive rainfall shock would reduce the village's average income by 8.1 percent, and the estimate is robust to omitted variables, extreme rainfalls, and alternative definitions of rainfall shocks. This finding suggests that, to some extent, farmers are irrational and overoptimistic. We also show that a lagged *negative* rainfall shock has no effect on the village income, which suggests that farmers are not over-pessimistic when predicting future rainfalls.

We then examined the predicted factor misallocation. We first constructed a standard measure of household-level agricultural productivity. Based on this measure, we show that a lagged positive rainfall shock leads low-productivity farmers to significantly reduce the area of land rented out, increase the time allocated to farm work, and reduce the time allocated to off-farm work. As the random and transitory rainfall shock last year has no direct effect on the agricultural productivity this year, the factor reallocation necessarily leads to a deviation from the optimal factor allocation and reduces the aggregate income of the farmers. In addition, we find no significant effect on the factor allocation of the high-productivity farmers. Therefore, the loss from the lagged positive rainfall shock is concentrated on the low-productivity farmers, which suggests that low-productivity farmers are more likely to be irrational.

Finally, we verified our main findings by estimating the regression model separately for farmers with different irrigation conditions. If the estimated effects truly reflect the irrational response of farmers to lagged positive rainfall shocks, we should find that the effects are much smaller for farmers with good irrigation conditions. This is because the agricultural production of farmers with good irrigation conditions depends less on rainfall and thus should have a smaller response to an expected positive rainfall shock. Consistent with this hypothesis, significant effects of lagged positive rainfall shocks on income and factor allocation are found only for farmers with poor irrigation conditions.

The findings of this article have important policy implications. As in most developing countries, agricultural production in China is dominated by smallholder farmers. Their limited access to information and social capital indicates that they are more likely to be irrational when making production decisions (Duflo et al. 2011, Wuepper & Sauer 2016, Fan & Salas Garcia 2018). Our finding suggests that any efforts to improve the rational expectation of farmers about rainfall shocks may reduce the damage of lagged rainfall shocks on rural income.

This study is closely linked to studies that used the exogeneity of rainfall shocks to examine various issues that ranged from civil conflicts (Miguel et al. 2004, Sarsons 2015, Gatti et al. 2021), to consumption smoothing and savings (Paxson 1992, Kazianga & Udry 2006, Brückner & Gradstein 2013), to wages and incomes (Jayachandran 2006, Rosenzweig & Udry 2014), and to various other socioeconomic outcomes (Colmer 2020, Grabrucker & Grimm 2020, Nordman et al. 2022, Ciccone & Ismailov 2022). We followed these studies to assume that rainfall shocks are exogenous, unanticipated, and transitory. Among these studies, our empirical strategy is closest to that of Kaur (2019). In a similar empirical setting, Kaur estimated that nominal wages in Indian villages rise during positive rainfall shocks but do not fall during droughts, and this downward wage rigidity reduces rural employment by 9 percent. We extended Kaur's study to examine the effects of lagged positive rainfall shocks on the aggregate agricultural efficiency and misallocation of land and labor in Chinese villages.

This study contributes to uncover that the irrationality of farmers is an important cause of factor misallocation in agriculture in a developing country setting. Agricultural productivity is remarkably low in most developing countries (Caselli 2005, Restuccia et al. 2008, Gollin et al. 2014), and many studies have argued that factor misallocation is a major cause (e.g., Lagakos & Waugh 2013, Adamopoulos et al. 2022). Existing studies mainly focused on the factor misallocation caused by barriers and risks that reduce the use of modern intermediate inputs (Restuccia et al. 2008, Donovan 2016) or caused by policies and institutional arrangements that limit land consolidation (Adamopoulos & Restuccia 2014, Chari et al. 2021). To the best of our knowledge, this article is the first to examine the factor misallocation caused by the irrational response of farmers to rainfall shocks.

This study also contributes to the broad literature on the economic consequences of irrationality and overoptimism. Psychological studies have shown that most people are irrational and overoptimistic (Weinstein 1980, Taylor & Brown 1988, Malmendier & Taylor 2015). Systematic errors in probabilistic reasoning caused by irrationality can distort many economic decisions, including investments, starting a business, and searching for jobs (Dawson et al. 2014, Bénabou & Tirole 2016, Jehiel 2018, Beaudry & Willems 2022). More relevantly, several studies have shown that irrational response to adverse weather had biased purchase behaviors with respect to clothes (Conlin et al. 2007), cars (Busse et al. 2015), and land parcels (Pan et al. 2022). Our study contributes to such literature by showing that the irrationality and overoptimism of farmers regarding rainfall shocks substantially distort their production behavior.

The rest of this paper proceeds as follows. Section 2 presents the conceptual framework. Section 3 details the data sets. Section 4 lays out the empirical strategy. Section 5 presents the empirical results. Section 6 concludes this paper. The online Appendix contains all appendix figures and tables.

2 Conceptual Framework

We build a simple model to illustrate the impact of a lagged positive rainfall shock on over-optimistic farmers. In a stand-in village inhabited by V farmers, individual farmers $i \in [1, V]$ are ranked according to their agricultural total factor productivity (TFP). Each farmer is endowed with a unit of farmland and a unit of labor. The labor of each farmer can be freely allocated to either self-employed farm work or off-farm work (but not both).[2] The off-farm wage is exogenously determined and increases with the farmer's agricultural TFP.[3] Each farmer chooses the optimal allocation of labor and land to maximize the total income from farm and off-farm work. The agricultural output of farmer i in year t is

$$Y_{it} = A_{it}(c_t) T_{it}^{\alpha} L_{it}^{1-\alpha} \ ,$$

where T_{it} and L_{it} are land and labor inputs, and the parameter $\alpha \in (0, 1)$. The agricultural TFP A_{it} depends on the rainfall realization $c_t = \bar{c} + s_t$ that consists of a time-invariant (predictable) component $\bar{c}$ and a random (unpredictable) rainfall shock s_t. The rainfall shock is exogenous, unanticipated, and transitory. In addition, a pos-

[2] For simplicity, the model abstracts from within-village wage rigidity, which has been modeled explicitly by Kaur (2019) when examining the effect of positive rainfall shocks on hired labor in Indian villages, because only less than 10% of farmers in China hire labor.

[3] Farmers with higher agricultural TFP are likely to be those who are better-educated, younger, healthier, and have better access to information, and these farmers are more likely to find better-paid off-farm jobs.

itive rainfall shock increases A_{it}, but agricultural decisions have to be made before farmers observed the rainfall shock (see the details in Subsection 4.1).

If farmers are rational and believe that rainfall shocks are random, agricultural decisions will always be made based on the constant term $\bar{c}$. In this case, rainfall shocks do not cause factor misallocation. However, if farmers are overoptimistic, a positive rainfall shock in year $t-1$ leads them to expect an above normal rainfall in year t. This irrational expectation leads to factor misallocation and welfare loss. To show this, we first examine the equilibrium for rational farmers and then extend the model to irrational farmers.

2.1 Rational equilibrium

Since each farmer can allocate labor only to farm work or to off-farm work, his total income is

$$I_i = \begin{cases} A_i(\bar{c})T_i^{\alpha} - (T_i - 1)R & \text{for } L_i = 1 \\ (1 - T_i)R + W_i^* & \text{for } L_i = 0 \end{cases},$$

where R is the equilibrium land rent, and W_i^* is the exogenous off-farm wage. The agricultural price is normalized to unity. For those who have only farm work ($L_i = 1$), the total income equals the farm output minus the land rents paid. For those who have only off-farm work ($L_i = 0$), the total income equals the land rents received plus the off-farm wage. The condition for having only farm work is that the marginal returns to agricultural labor (i.e., the agricultural wage) are higher than the off-farm wage: $(1-\alpha)A_i(\bar{c})T_i^{\alpha} > W_i^*$.

We simplify the model by assuming that farmers with a lower agricultural TFP have a higher comparative advantage in off-farm work (we will relax this assumption later). Since farmers are ranked by their agricultural TFP, we can always find a threshold farmer $z^* \in (1, V)$ so that all farmers with $A_i < A_{z^*}$ ($A_i \geq A_{z^*}$) have only off-farm work (farm work) and rent land out (rent land in). The land market clearing condition is[4]

$$\sum_{i=z^*+1}^{V} T_i = V, \tag{1}$$

which indicates that the total land managed by those in agriculture equals the village

[4]This condition is simplified from $\sum_{i=z^*+1}^{V} (T_i - 1) R = z^* R$.

land endowment. The aggregate return to land is

$$I_T = \sum_{i=z^*+1}^{V} \alpha A_i(\bar{c}) \, (T_i)^{\alpha-1} \quad , \tag{2}$$

and the aggregate return to labor is

$$I_L = \sum_{i=1}^{z^*} W_i^* + \sum_{i=z^*+1}^{V} (1-\alpha) A_i(\bar{c}) \, (T_i)^{\alpha} \quad . \tag{3}$$

The village aggregate income is $I_V = I_T + I_L$.

2.2 Welfare loss from overoptimism

The impact of a lagged positive rainfall shock on overoptimistic farmers is visually presented in Figure 1. The green line presents the off-farm wage, and the solid red line presents the agricultural wage under normal rainfall $\bar{c}$.[5] The intersection of these two lines defines the threshold farmer z^*: farmers with $i < z^*$ ($i > z^*$) have only off-farm (farm) work. Now assume a positive rainfall shock in year $t-1$ and no rainfall shock in year t. The seasonality of agricultural production determines that farmers make agricultural decisions for year t based on the rainfall in year $t-1$. If farmers are overoptimistic and believe that the rainfall realization in year t will also be above normal, they will expect an agricultural wage (dashed red line) that is higher than the normal wage (solid red line). The equilibrium threshold farmer declines to z'. The labor misallocation reduces the aggregate labor income by ghj (the red triangle). The lower-productivity farmers z' to z^* no longer rent land out, and this land misallocation necessarily reduces the aggregate return to land.

[5]The relative position of the two lines reflects the assumptions that individuals with a higher agricultural TFP have a higher off-farm wage and that farmers with a lower agricultural TFP have a higher comparative advantage in off-farm work. We will discuss the consequence of relaxing these assumptions.

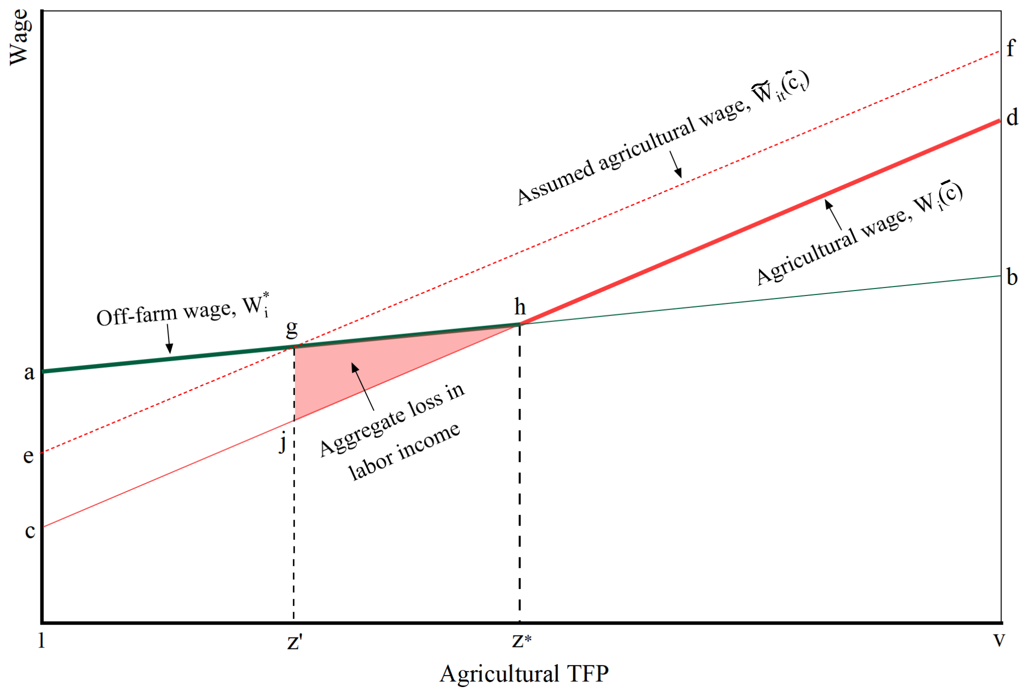


Figure 1: Factor misallocation and income loss from a lagged positive rainfall shock

Note: This is an illustrative figure based on the assumption that farmers with a lower agricultural TFP have a higher comparative advantage in off-farm work. The solid red (green) line shows the farm (off-farm) wages. The dashed red line shows the agricultural wage in year t *assumed* by overoptimistic farmers who observed a positive rainfall shock in year $t-1$.

Formally, we provide the following empirically testable propositions for overoptimistic farmers:

Proposition 1: A positive rainfall shock in year $t-1$ leads some overoptimistic farmers to misallocate labor from higher-wage off-farm work to lower-wage farm work in year t. As presented in Figure 1, farmers z' to z^* misallocate labor from off-farm work to farm work.

Proposition 2: A positive rainfall shock in year $t-1$ leads to the misallocation of farmland from higher-productivity farmers to lower-productivity farmers in year t. The land market clearing condition 1 implies that the decline of the threshold farmer from z^* to z' necessarily reduces the land managed by the high-productivity farmers ranked above z^*.

Proposition 3: A positive rainfall shock in year $t-1$ reduces the village aggregate income in year t. Equations (2) and (3) indicate that a decline of the threshold farmer reduces the aggregate incomes from land and labor, respectively. Therefore, the village aggregate income, which is the sum of the land and labor incomes, necessarily declines.[6] However, the impact on the village aggregate agricultural output ($\sum_{i=z+1}^{V} A_{it}(c_t) T_{it}^{\alpha} L_{it}^{1-\alpha}$) is ambiguous because average productivity declines while labor input increases.[7]

Note that these propositions do not depend on the strong assumptions. First, alternatively assuming that only some farmers are overoptimistic (but not all farmers) does not affect Propositions 1–3.[8] Moving on to the even more realistic assumption that farmers with a lower agricultural TFP are more likely to be irrational,[9] the model predicts a larger damage for farmers with a lower agricultural TFP. Second, we can alternatively assume that farmers with a higher agricultural TFP have a higher comparative advantage in off-farm work. Overoptimistic farmers still misallocate labor from higher-wage off-farm work to lower-wage farm work and the aggregate income still declines (Propositions 1 and 3 holds), but farmland is now reallocated from lower-productivity farmers to higher-productivity farmers (Proposition 2 fails). Finally, we can alternatively assume that the rainfall shock in year $t-1$ is positive or negative (instead of there being no shock). As long as the realized rainfall in $t-1$ is lower than the prediction of the overoptimistic farmers, all the qualitative predictions hold.

Note also that the conceptual framework adopts a strong definition of farmers' optimal factor allocation. Under the assumption that rainfall shocks are random and unpredictable by farmers, we implicitly define a farmer's factor allocation as optimal if they do not adjust it in response to a lagged positive rainfall shock. However, this definition is too stringent for the following empirical analysis since individual farmers may respond differently to lagged positive rainfall shocks due to the uncertainty of rainfall shocks and their varying access to (good or bad) information about them. Nonetheless, if farmers, on average, are rational and believe that rainfall shocks cannot be predicted based on available information, we should not expect to observe the average farmer adjusting their factor allocation in response to a lagged positive rainfall shock. This holds particularly true when examining a large number of farmers, such

[6]The equilibrium land rent ($R = \alpha A_i T_i^{\alpha-1}$) will increase due to the decline of T_i. However, this will not affect the village aggregate income because the rents paid equal the rents received within the village.

[7]More specifically, a lagged positive rainfall shock reduces the threshold from z^* to z', resulting in two offsetting effects on total agricultural output. First, it leads to the participation of more low-productivity farmers (i.e., those between z' and z^*) in agriculture, thereby reducing average agricultural productivity. Second, the number of farmers engaged in agriculture mechanically increases by $z^* - z'$. Since total agricultural output depends on both productivity and labor input, the net effect of a lagged positive rainfall shock on total agricultural output is ambiguous.

[8]The only noticeable difference is that land rent may not increase following a positive rainfall shock because the exogenous off-farm wage of rational farmers pins down the equilibrium land rent.

[9]This assumption is kind of self-fulfilling: irrational farmers are more likely to be damaged by lagged positive rainfall shocks and thus have a lower "observed" agricultural TFP.

as the more than 10 thousand farms analyzed in this study.

3 Data

3.1 National Fixed-Point Survey

Our primary data are from the National Fixed-Point Survey (NFPS). Beginning in 1986, this panel-structured survey was collected by the Research Center of Rural Economy of the Chinese Ministry of Agriculture. We use annual waves of data between 2003 and 2013, covering 10,086 households in 151 villages from 11 randomly selected Chinese provinces.[10] The 11 sample provinces are Shanxi, Jilin, Zhejiang, Fujian, Jiangxi, Henan, Hubei, Hunan, Guangdong, Sichuan, and Gansu, which cover different geographical areas and economic regions of China (see Appendix Figure 1). The NFPS villages were selected for representativeness based on region, income, population, cropping pattern, and non-farm activity. Within each village chosen, a random sample of households (ranging from 50 to 100 based on the size of the village) was drawn to be included in the survey. If the entire household moved permanently, the household was replaced by a similar household with a new household ID. After households with missing values and obvious errors were excluded, the remaining sample is 10,034 households.[11] The NFPS data contain detailed information on household agricultural production, income, and employment.

[10]The survey covered all of the 31 provincial districts in mainland China, but we only have access to the data from the 11 randomly selected provinces. The microdata before 2003 are not accessible, and the survey structure changed substantially after 2013.

[11]Our data set is an unbalanced panel, as some households moved out of the village permanently during the sample period. More than 90% of the sample households presented in the data for at least 9 years.

Table 1: Summary statistics of variables

	N	Mean	S. D.
Panel A: Household-year variables			
If hired labor for agricultural production (1=yes; 0=no)	89529	0.093	–
Daily wage for hired agricultural labor (yuan)	6184	51.8	51.5
Area of farmland cultivated (mu)	89529	6.2	9.0
If rent out land (1=yes; 0=no)	89529	0.168	–
Area of farmland rented out (mu)*	89529	1.547	3.541
If there is farm work (1=yes; 0=no)	89529	0.904	–
The number of farm-work family members	89529	1.894	1.193
If there is off-farm work (1=yes; 0=no)	89529	0.581	–
The number of off-farm work family members	89529	1.071	1.493
Irrigation expenditure (yuan)	89529	82.4	243.0
Panel B: Village-year variables#			
Average income per household (thousand yuan)	1488	40.2	41.0
Average agricultural income per household (thousand yuan)	1488	7.7	8.0
Average non-agricultural income per household (thousand yuan)	1488	32.5	40.8
Panel C: Household-crop-year variables			
Crop output (kg)	373747	1058	3383
Area of land harvested (mu)	373747	1.92	4.79
Labor inputs (days)	373747	29.9	61.6
Machine inputs (yuan)	373747	92.8	322
Other inputs (yuan)	373747	382	1266

Note: All variables are derived from the NFPS data over 2003—2013 for 10,034 households in 151 villages. All wages, incomes, and costs are measured in 2003 constant yuan, after adjusting for the national inflation rate.

* The area of farmland rented out is calculated as the flow of new land renting that occurred within a single year, rather than the total stock of land renting. We do not know the total amount of land rented out by each household as the land renting that took place before our sample period is not observed.

\# The village average income is calculated as the average of the household total income across all sample households within a village. The total income for each household is the sum of the family net agricultural income and the off-farm incomes of all family members. The family's net agricultural income does not account for the costs of family-endowed farmland and labor inputs from family members.

Benjamin et al. (2005) verified the high quality of the NFPS data and supported the representativeness of the data by comparing them with aggregate statistics from the Chinese Agricultural Census. The key advantages of the data are its panel structure and the detailed information on agricultural inputs and outputs at the household-crop-year level. The data allow us to generate productivity measures at the household level and thereby investigate the impact of lagged positive rainfall shocks on the behavior of households with different agricultural productivity levels.

The summary statistics for the key variables are presented in Table 1. For example, the data show that 9.3 percent of the households hired labor for agricultural production, with a daily wage of 51.8 yuan (in 2003 constant yuan). An average household cultivated 6.2 mu of farmland (0.41 hectare) and rented out 0.44 mu (see note of Table 1 for the interpretation). While 81.8 percent of the households participated in family-based agricultural production, 56.6 percent had at least one person with an off-farm job. The annual income of an average rural household was 40.2 thousand yuan (about 5.7 thousand US dollar). The village average agricultural income across the sample households was 7.7 thousand yuan, and the village average non-agricultural income was 32.5 thousand yuan.

3.2 Rainfall data

The rainfall data used in this study were collected from the latest state-of-the-art global reanalysis dataset, the Enhanced Global Dataset for the Land Component of the Fifth Generation of European ReAnalysis (ERA5-Land).[12] It is produced by the European Center for Medium-Range Weather Forecasts within the Copernicus Climate Change Service of the European Commission, spans from 1981 to the present, and covers key climatic variables with a resolution of 9×9 kilometers. We derived the monthly rainfall data from ERA5-Land for each of the 151 NFPS villages from 1981 to 2013. Because we do not have village boundary information, we simply use the rainfall of the climate grid from ERA5-Land that covers the center of each village as the rainfall of that village (instead of calculating the area-weighted average rainfall from all relevant climate grids).[13] We also derived the village-level temperature data from ERA5-Land using the same process.

3.3 Definition of rainfall shock

We define rainfall shock based on the total rainfall of the first month of the growing season of the most important crop in each village. This is because both the level and timing of rainfall are important for agricultural productivity. Existing studies found that crops are most sensitive to rainfall occurring at the beginning of the growing season (e.g., Kaur 2019, Gatti et al. 2021, Premand & Stoeffler 2022). We also estimated the effect of the rainfall shock from each of the 12 months (Appendix Figure 7) and

[12]Details of ERA5-Land can be found in Muñoz-Sabater et al. (2021).

[13]The size of a typical Chinese village is smaller than the ERA5-Land resolution of 9×9 km.

confirmed that the most significant effect is from the first month of the growing season. Another advantage of defining rainfall shock based on a single month (instead of on the entire growing season) is to ensure the credibility of the identification assumption that the effect of rainfall shock is transitory; see Subsection 4.1 for details.

We construct the village-year rainfall shock indicators in three steps. We first identify the most important crop in each village as that with the largest harvesting area during 2003–2013. We then identify the first month of the growing season of the most important crop based on the crop-level planting date records from 778 nationally representative agro-climatic monitoring sites.[14] We match each village with the nearest monitoring site. For most of the villages, the most important crop is wheat, rice, or corn, and the first month of the growing season is March, April, or May.[15] Finally, we follow Kaur (2019) in defining a village as subjected to a positive (negative) rainfall shock in a given year if the rainfall in the first month of the growing season is above the 85th percentile (below the 15th percentile) of the long-term rainfall distribution of the village for the same month. The remaining rainfall realizations are defined as no shock. The long-term rainfall distribution is constructed based on the data from 1981 to 2013.

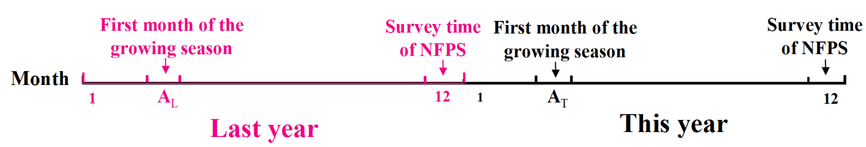


Figure 2: Timeline of the definition of rainfall shock

Note: This figure illustrates the timing of the NFPS survey in each year and the first month of the growing season, which is used to define the rainfall shock for that particular year.

The timeline presented in Figure 2 further clarifies our definition of rainfall shock. As detailed above, the rainfall shock is defined based on the rainfall outcome in the first month of the growing season, which could be March, April, or May, depending

[14]The data we used contain detailed site-level, crop-specific planting date records of 778 nationally representative monitoring sites over 1993–2013 stored in the data archive of the National Meteorological Information Center of China. Please refer to Cui & Xie (2022) for details of the data set and monitoring sites.

[15]In 21 villages the most important crop is a cash crop. It is difficult to identify the growing season of a cash crop, so we use instead the annual average of the monthly total rainfall to calculate the rainfall shock. The a robustness check in Table 4, we use instead the growing season of the most important food crop for the 21 villages and find comparable results.

on the major crop of the village. As shown in the figure, the rainfall shock from last year (i.e., the lagged rainfall shock) is defined at point A_L, and the rainfall shock for this year is defined at point A_T. Since each round of the NFPS data was collected in December of each year, farmers only need to recall events that occurred over the past 12 months. As each round of NFPS covers a complete growing season for the major crops defined in this study (usually from March to November), we are not concerned that recall bias could have caused a correlation between the rainfall shocks in last year and this year.

Figure 3 presents the village-level standardized deviation of rainfall in the first month of the growing season from the long-term average of the same month. Each red point in the figure is the total rainfall of the first month of the growing season minus the average over 1981–2013, divided by the standard deviation. Therefore, the y-axis measures the standardized deviations away from the long-term average. We mark with dashed green lines the 15th and 85th percentiles (i.e., 1.036 and -1.036 standardized deviations) to indicate the negative and positive shocks defined above. The figure also presents the cross-village distribution (the black curve) and quartiles (the whisker box) of the rainfall deviation. We find substantial and seemingly random rainfall shocks over the 11 sample years.

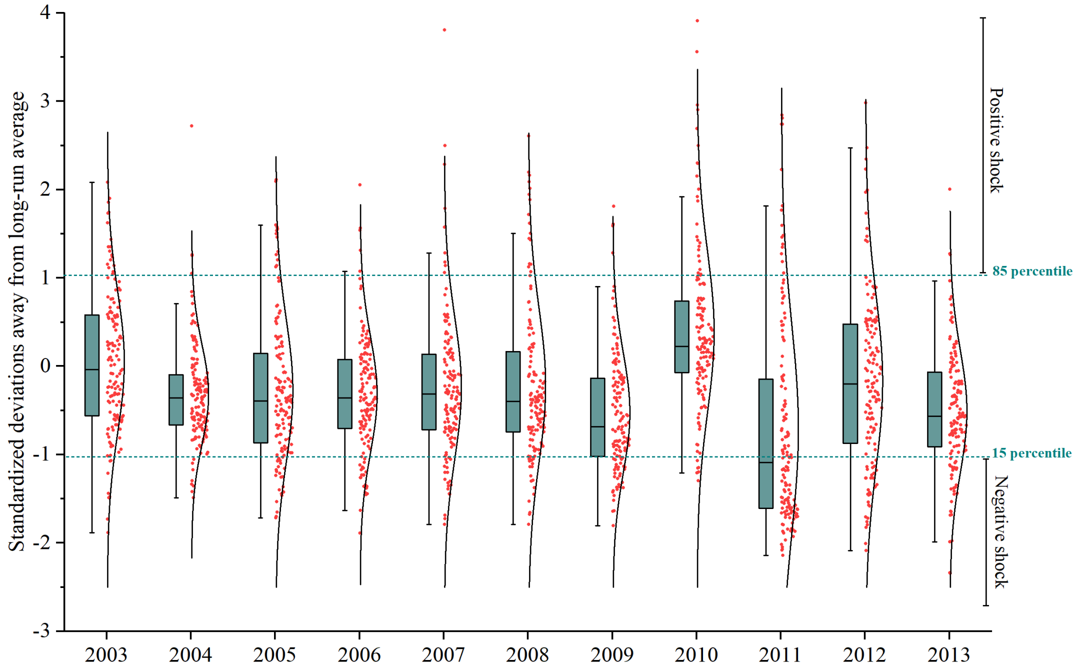


Figure 3: Village-level standardized deviation of rainfall from the long-term average

Note: Each red point in the figure represents a village's standardized rainfall deviation in the first month of the growing season from the long-term average. Specifically, denote $r_{v,1,t}$ as the rainfall of village v in the first month of the growing season and year t, and $\bar{r}_{v,1}$ and $SDr_{v,1}$ respectively as the average and the standard deviation of the rainfall over 1981–2013 for the same village and month, then each point in the figure is calculated as $\frac{r_{v,1,t}-\bar{r}_{v,1}}{SDr_{v,1}}$. Each box in the figure displays the first, second, and third quartiles of the rainfall deviation across villages in a given year, and the whisker defines outliers.

Table 2 classifies the village-level rainfall shocks into nine categories and presents the frequency of each category over 2003–2013. The largest category is no shock in both last year and this year, which accounts for 56.6% of the sample. The category of negative shock last year and no shock this year accounts for 12.6%, and the category of no shock last year and negative shock this year accounts for 12.4%. The category that we are most interested in—positive shock last year and no shock this year—accounts for 7.3% of the sample.

Table 2: Frequency of the village-year rainfall shocks

Last year's shock	This year's shock	Frequency	Proportion
Negative	Negative	26	1.75%
Negative	None	188	12.63%
Negative	Positive	18	1.21%
None	Negative	185	12.43%
None	None	842	56.59%
None	Positive	89	5.98%
Positive	Negative	21	1.41%
Positive	**None**	**109**	**7.33%**
Positive	Positive	10	0.67%

Note: We define a village-year as subjected to a positive (negative) rainfall shock if the total rainfall in the first month of the growing season is above the 85th percentile (below the 15th percentile) of the village's long-term rainfall distribution of the same month; other rainfall realizations are defined as no shock.

4 Identification Strategy

4.1 Stylized facts

The identification of this article is based on three stylized facts. First, rainfall shocks are exogenous, unanticipated, and transitory (Paxson 1992, Miguel et al. 2004, Brückner & Gradstein 2013). The randomness of rainfall shocks ensures that they are not determined by any other factors affecting incomes and factor allocation. Appendix Table 2 shows that the village-level rainfall shocks employed in this study are both serially and spatially uncorrelated.[16] The transitory of rainfall shocks ensures that rainfall shock this year will not directly affect agricultural output in the next year by altering soil moisture. The agronomy and meteorology literature show that the soil in regions where agriculture depends mainly on rainfall generally cannot retain rainfall water for a whole year (Charney et al. 1977, Bodner et al. 2015). Our study further ensures the transitory of rainfall shocks by defining the shock based on the rainfall in a single month.

The second stylized fact is that a positive rainfall shock generally increases agricultural productivity (e.g., Tao et al. 2008, Chen et al. 2013, Ciccone & Ismailov 2022). This fact explains why overoptimistic farmers increase their agricultural activity in re-

[16]We find no cross-village rainfall correlation presumably because our sample villages are distant from each other. Note that we use data from 151 villages randomly selected from the more than half a million villages in China.

sponse to a lagged positive rainfall shock. Appendix Table 4 presents the significantly positive correlation between rainfall shocks and crop yields in our sample villages. Note that the rainfall shocks defined in this study are different from rainfall extremes (i.e., floods and droughts) that could damage agricultural production. In a robustness check (Table 4), we exclude village-years with rainfall above the 5th percentile of the distribution and find a comparable result.

The third stylized fact is that due to the seasonality of agricultural production, most agricultural production decisions have to be made before farmers have observed the rainfall in the growing season. For example, farmers in China usually rent farmland in or out shortly after the growing season of the last lunar year so that they would have enough time to prepare the land and purchase inputs for the next growing season. Farmers can only adjust some of the flexible inputs, such as fertilizers and irrigation, after observing the rainfall in the growing season. This fact explains why irrational farmers depend on lagged rainfall shocks to make agricultural decisions.

4.2 The aggregate effect

Our theoretical model suggests that the effect of a lagged positive rainfall shock differs across farmers and income sources. To capture the aggregate effect, we first estimate a village-level regression model that uses a village's average household income as the dependent variable:[17]

$$\begin{aligned} \ln(Income_{v,t}) = \beta_0 &+ \beta_1 Neg_{v,t-1} Pos_{v,t} + \beta_2 Neg_{v,t-1} None_{v,t} \\ &+ \beta_3 Neg_{v,t-1} Neg_{v,t} + \beta_4 None_{v,t-1} Pos_{v,t} \\ &+ \beta_5 None_{v,t-1} Neg_{v,t} + \beta_6 Pos_{v,t-1} Pos_{v,t} \\ &+ \beta_7 Pos_{v,t-1} None_{v,t} + \beta_8 Pos_{v,t-1} Neg_{v,t} \\ &+ X_{v,t}\eta + \gamma_v + \gamma_{pt} + \epsilon_{v,t} \end{aligned} \tag{4}$$

where $Income_{v,t}$ is the household average income (from agriculture, off-farm work, or both) in village v and year t. The model includes a full set of interactions between the indicators of last year's rainfall shock and this year's rainfall shock. Specifically, the notations $Neg_{v,t-1}$ ($Neg_{v,t}$) , $None_{v,t-1}$ ($None_{v,t}$), and $Pos_{v,t-1}$ ($Pos_{v,t}$) are indicators that equal one if last year's (this year's) rainfall is a negative shock, no shock, and a positive shock, respectively, and equals zero otherwise. The interaction $None_{v,t-1} None_{v,t}$

[17] We use village-average household income instead of village total income as the measure of village aggregate income because the number of sample households differs across villages.

is excluded from the model, so the effect of each category is evaluated relative to this omitted category. Besides a set of time-varying control variables $X_{v,t}$,[18] the model also includes the village fixed effects (γ_v) to account for all village-specific time-invariant factors, and includes the prefecture-year fixed effects (γ_{pt}) to account for annual fluctuations that are common across villages.[19] Finally, $\epsilon_{v,t}$ is an error term.

The coefficient of interest is β_7, which captures the effect of a positive shock last year and no shock this year ($Pos_{v,t-1}None_{v,t}$). Since the reference category is no shock in both years ($None_{v,t-1}None_{v,t}$), the estimate of β_7 can be clearly interpreted as the effect of a lagged positive rainfall shock. In contrast, while the coefficients β_6 and β_8 also partially reflect the effect of last year's positive shock, they are difficult to interpret because they also capture the effect of this year's positive or negative shock.

Since we are only interested in the coefficient of $Pos_{v,t-1}None_{v,t}$, the regression model (4) can be simplified to

$$\begin{aligned} \ln(Income_{v,t}) = \alpha_0 &+ \alpha_1 Pos_{v,t-1}None_{v,t} + \alpha_2 Neg_{v,t-1}None_{v,t} \\ &+ \alpha_3 Pos_{v,t} + \alpha_4 Neg_{v,t} + X_{v,t}\eta + \gamma_v + \gamma_{pt} + \epsilon_{v,t} \quad . \end{aligned} \tag{5}$$

All the variables are defined as in model (4),[20] and the omitted category is still $None_{v,t-1}None_{v,t}$. The coefficient α_1 in model (5) is equivalent to the coefficient β_7 in model (4). Model (5) will be used as the baseline regression model throughout this article. According to the theoretical prediction, if farmers are overoptimistic, last year's positive rainfall shock could cause factor misallocation this year and thus reduce the village-average income. Therefore, a significantly negative estimate of α_1 can be taken as strong evidence of the overoptimism of farmers.[21]

[18] The control variables are the growing season mean temperature and its square, and three measures of village-level social capital: the share of households with party members, officials, and religious believers. We include a limited number of control variables because the randomness of rainfall shocks implies that correlated control variables are most likely to be the channel variables that should not be controlled.

[19] A prefecture is an administrative division above a county in China, and each prefecture contains about 10 county-level divisions on average. The prefecture-year fixed effects are preferred to the year fixed effects because they not only account for national common shocks but also account for common changes at the prefecture level.

[20] The dummy variable $Pos_{v,t}$ is the combination of $Neg_{v,t-1}Pos_{v,t}$, $None_{v,t-1}Pos_{v,t}$, and $Pos_{v,t-1}Pos_{v,t}$ from model (4), and the dummy variable $Neg_{v,t}$ is the combination of $Pos_{v,t-1}Neg_{v,t}$, $None_{v,t-1}Neg_{v,t}$, and $Neg_{v,t-1}Neg_{v,t}$.

[21] The coefficients α_3 and α_4 are difficult to interpret because they measure the effect of this year's positive or negative shock relative to the effect of no shock in *both* years. Someone may be also interested in the coefficient of α_2 because it examines whether last's negative shock affects income this year.

4.3 Factor misallocation

We estimate the effect of lagged positive rainfall shocks on factor allocation based on a version of model (5) specified at the household level:

$$\begin{aligned} H_{h,v,t} = \delta_0 + \delta_1 Pos_{v,t-1} None_{v,t} + \delta_2 Neg_{v,t-1} None_{v,t} \\ + \delta_3 Pos_{v,t} + \delta_4 Neg_{v,t} + X_{h,v,t}\eta' + \gamma_v + \gamma_{pt} + \epsilon'_{h,v,t}, \end{aligned} \tag{6}$$

where the dependent variable $H_{h,v,t}$ is one of the measures of labor or land allocation for household h from village v in year t. All the other variables are the same as defined before.[22] The dependent variables examined include *whether to participate in farm work*, *whether to participate in off-farm work*, *the number of family members participating in off-farm work*, *whether to subcontract out farmland*, and *the area of land subcontracted out.* We will also use this model to estimate the effect on household-level incomes. The coefficient δ_1 captures the effect of interest.

4.4 Measuring agricultural productivity

Our theoretical model predicts different effects on farmers with different agricultural productivity. To test this prediction, we construct a standard measure of household-level agricultural productivity. Assuming that the crop-specific production function is

$$\begin{aligned} \ln(Yield_{i,h,v,t}) = \lambda_A \ln(Area_{i,h,v,t}) + \lambda_L \ln(Labor_{i,h,v,t}) + \lambda_M \ln(Machine_{i,h,v,t}) \\ + \lambda_I \ln(Input_{i,h,v,t}) + \phi_{i,h,v,t}, \end{aligned} \tag{7}$$

where $Yield_{i,h,v,t}$ denotes the yield of crop i cultivated by household h in village v and year t, and $Area_{i,h,v,t}$, $Labor_{i,h,v,t}$, $Machine_{i,h,v,t}$, and $Input_{i,h,v,t}$ denote the farmland area, labor days, machinery costs, and all other input costs in the production, respectively. The logarithm of TFP is thus identified by $\phi_{i,h,v,t}$.

We decompose TFP into four components:

$$\phi_{i,h,v,t} = \phi_{i,h} + \phi_{h,t} + \phi_{i,v,t} + e_{i,h,v,t}. \tag{8}$$

where $\phi_{i,h}$ is a fixed household-crop component that captures the household's fixed ability to farm a given crop, $\phi_{h,t}$ is a household-year component, $\phi_{i,v,t}$ is a village-crop-

[22] A slight difference is that the three control variables for social capital are now defined at the household level.

year component, and $e_{i,h,v,t}$ is an TFP shock.

The household-year component, represented by $\varphi_{h,t}$, captures productivity shocks that affect all crops cultivated by a household in a given year, such as a health shock. Similarly, the village-crop-year component, denoted as $\varphi_{i,v,t}$, captures factors that are common to all households in the same village cultivating crop i in a given year. Therefore, the average of the fixed household-crop component $\phi_{i,h}$ across all crops cultivated by the household (denoted by ϕ_h) is a good measure of the household's agricultural ability. We estimate ϕ_h based on the crop-level input and output data from NFPS. Appendix Figure 2 presents the distribution of the estimated ϕ_h across the sample households.[23]

In robustness checks, we also measure a household's agricultural ability by its marginal land productivity (denoted by λ_h). This measure is also constructed in two steps. First, we calculate the household-crop level marginal productivity by multiplying the output per unit of land with the elasticity of output with respect to land [i.e., λ_A from Equation (7)]. Second, we calculate the average of the marginal productivity across all the crops to obtain the household-level marginal agricultural productivity λ_h.

5 Empirical Results

5.1 Aggregate income loss

Table 3 presents the estimated effect of a lagged positive rainfall shock on the village average income. All the estimations are based on model (5), which controls for prefecture-year fixed effects, village-fixed effects, and the five other control variables. Standard errors are clustered at the village level. The dependent variables examined are the village's average total household income (column 1), agricultural income (column 2), and off-farm income (column 3).[24] See Table 1 for the definitions of these variables. We are mainly interested in the coefficient of the indicator of a positive

[23]Households that never had agricultural production during the sample years (about 5% of the total sample) were excluded from the construction.

[24]We prefer to utilize village-level data when estimating the village-level total effect in Table 3. This choice is based on the theoretical model, which suggests that the impact of a lagged positive rainfall shock varies across farmers. Moreover, given that farmland markets in China generally operate at the village level, village-level estimates are better suited to capture the general equilibrium effect. However, it is worth noting that some concerns may arise regarding the fact that village is not a decision-making unit. To address these concerns, we replicate the estimation presented in Table 3 using household-level data. The results, as shown in Appendix Table 7, are very similar, although the estimated standard errors are smaller due to the larger sample size.

rainfall shock last year and no rainfall shock this year.

The estimate presented in column 1 suggests that a positive rainfall shock last year (and no shock this year) reduces the village average income this year by 8.1 percent (relative to the case of no shock in both years),[25] and this effect is statistically significant at the 5 percent level.[26] This finding supports our theoretical prediction that a lagged positive rainfall shock reduces the village aggregate income (measured by the household average income). If farmers are fully rational, the exogenous, unanticipated, and transitory rainfall shock that occurred in the last year should have no effect on their income this year. Therefore, this estimate provides strong evidence that lagged positive rainfall shocks lead farmers to make irrational and overoptimistic decisions.[27]

[25]The estimated effect of the lagged positive rainfall shock (8.1 percent) is comparable to that found by Kaur (2019), who estimated that wage rigidity resulting from the lagged positive rainfall shock reduces rural employment by 9 percent. It is worth noting that the estimated effect of the lagged positive rainfall shock is approximately 80 percent of the estimated effect of the current-year negative rainfall shock. However, the relatively small effect of the current-year negative rainfall shock may be attributed to the well-developed irrigation facilities in China, which have potentially mitigated a significant portion of the damage caused by negative rainfall shocks (Wang et al. 2020).

[26]The standard error of the estimate is large because we estimate the village average effect across households with very different responses to lagged positive rainfall shocks (see Figure 6).

[27]A crucial prediction of this study is that a lagged positive rainfall shock amplifies the rainfall expectations of irrational farmers. However, data on rainfall expectations are not obtainable from the NFPS (the data source employed in this study) or any other surveys conducted in rural China. Consequently, direct testing of this prediction is not feasible. To address this limitation, we conducted a phone-call survey involving 658 farmers. As detailed in Appendix B.6, our findings provide evidence that farmers who experienced above-average rainfall this year are more prone to anticipate above-average rainfall in the following year.

Table 3: The impact of a lagged positive rainfall shock on the village average income

		(1)	(2)	(3)
		Log village average income	Log village average agricultural income	Log village average off-farm income
Last year's shock	***This year's shock***			
Positive	None	-0.081**	-0.068	-0.110**
		(0.040)	(0.124)	(0.052)
Negative	None	-0.036	-0.025	-0.018
		(0.056)	(0.069)	(0.062)
Any	Positive	0.058	-0.098	-0.002
		(0.127)	(0.195)	(0.138)
Any	Negative	-0.105**	-0.088	-0.104**
		(0.040)	(0.077)	(0.051)
Prefecture-year fixed effects		Yes	Yes	Yes
Village fixed effects		Yes	Yes	Yes
Five other controls		Yes	Yes	Yes
Observations		1488	1488	1488
Adjusted R^2		0.930	0.978	0.915

Note: This table reports the estimates of model (5). The omitted rainfall shock category is no shock in both last year and this year. The dependent variables are the village average income, village average agricultural income, and village average off-farm income in columns 1–3, respectively. The standard errors in parentheses are clustered at the village level. The significance levels are *** $p<0.01$, ** $p<0.05$, and * $p<0.1$

We also briefly interpret other coefficients of the baseline regression. The coefficient of the indicator of a negative shock last year and no shock this year is negative but statistically insignificant, suggesting that farmers are not over-pessimistic. The coefficient of the indicator of any kind of shock last year and a positive shock this year is positive but statistically insignificant. This means that relative to the case of always no rainfall shock (i.e., the reference category), a positive rainfall shock does not significantly increase income. This finding reflects the fact that the gain from a positive rainfall shock is largely neutralized by the subsequent loss from it.[28] The coefficient of the indicator of any kind of shock last year and a negative shock this year is significantly negative, suggesting that a negative rainfall shock is always harmful.

Columns 2 and 3 of the table present the effects on the household average agricultural income and the household average off-farm income, respectively. The estimated damage to the average agricultural income is smaller (-6.8 percent) and is not statis-

[28] The income increase from a positive rainfall shock in the current year is 10.4% (column 3 of Table **??**) while the income loss from it in the next year is 8.1% (column 1 of Table 3).

tically significant at a conventional level. Note that our theoretical model predicts an ambiguous effect on the average agricultural income because a lagged positive rainfall shock increases the agricultural activity. The estimated damage to the average off-farm income is larger (-11.0 percent) and statistically significant at the 5 percent level. Therefore, the damage caused by a lagged positive rainfall shock comes mainly from the loss of off-farm income.

5.1.1 Robustness checks

We show that the baseline estimates presented in Table 3 are robust to the omitted variables, extreme rainfall, alternative definitions of rainfall shocks, and controlling for rainfall shocks lagged up to four years. All the robustness checks have the same model setting as the baseline estimation except that specified in each check. For simplicity, we report only the estimates of interest.

Omitted variable. The baseline estimation controls for three measures of social capital (i.e., the share of households with party members, officials, and religious believers) that may affect the aggregate income. It also controls for the growing season mean temperature and its square to address the concern that the rainfall measures may capture the effect of temperature to the extent that they are correlated. We include a limited number of control variables because the randomness of rainfall shocks implies that correlated control variables are most likely to be the mediating or moderating variables that should not be controlled for. Columns 1a, 1b, and 1c of Table 4 show that excluding these five control variables only slightly alters the estimated effects on the village-average total income, agricultural income, and off-farm income, respectively. We have also tried to control for additional time-varying factors (i.e., education, subsidy, and the dependent ratio) and alternative temperature measures (i.e., temperature bins, growing season degree-day, and the temperature of the first month of the growing season) and found comparable results; these estimates are available upon request.

Rainfall extremes. In our main analysis, we define a positive rainfall shock as the rainfall realization above the 85th percentile of the long-term distribution. Appendix Table 4 shows that a positive rainfall shock significantly increase the current-year agricultural yields and village income. A potential concern is that the positive rainfall shock includes extreme rainfall that might damage current-year agricultural production, which could make our findings difficult to interpret.[29] To address this concern, columns 2a, 2b, and 2c of Table 4 exclude village-years with rainfall above the top 5%

[29]If a positive rainfall shock reduces the current-year agricultural output, the response of farmers to this shock in the next year should not be interpreted as the consequence of overoptimism.

or below the bottom 5% of the long-term distribution. The resulting estimates have no statistically significant difference from the baseline estimates.

Defining rainfall shocks **based on food crops**. Our main analysis defines rainfall shock based on the rainfall in the first month of the growing season of the most important crop in each village. However, in 21 of the sample villages, the most important crops are cash crops, which have no clear growing season. For these villages, we alternatively define the rainfall shock based on the annual average monthly rainfall. Columns 3a, 3b, and 3c check the robustness to this choice by defining the rainfall shock based on the growing season of the most important *food* crops for these villages. The resulting estimates are comparable to the baseline estimates.

Spatial correlation. Although Appendix Table 3 demonstrates that rainfall shocks do not exhibit spatial correlation across our sample villages, it is possible that other shocks, such as non-weather economic shocks, may exhibit correlation. To address this concern, we employ Conley standard errors with a radius of 500 km and arbitrary serial correlation in columns 4a, 4b, and 4c (Abadie et al. 2022, Colella et al. 2020). Notably, the estimated standard errors in these columns are highly similar to those from the baseline estimations, reaffirming that spatial correlation does not significantly impact the findings of this study.

Table 4: Robustness checks of the impact on the village average income

		(1a)	(2a)	(3a)	(4a)	(1b)	(2b)	(3b)	(4b)	(1c)	(2c)	(3c)	(4c)
		Log village average income				Log village average agricultural income				Log village average off-farm income			
		Control variables	Rainfall extremes	Rainfall shock definition	Spatial correlation	Control variables	Rainfall extremes	Rainfall shock definition	Spatial correlation	Control variables	Rainfall extremes	Rainfall shock definition	Spatial correlation
Last year	***This year***												
Positive	None	-0.069**	-0.087**	-0.064**	-0.081**	-0.068	-0.066	-0.011	-0.068	-0.125**	-0.127**	-0.107**	-0.110**
		(0.031)	(0.043)	(0.026)	(0.041)	(0.137)	(0.128)	(0.049)	(0.097)	(0.049)	(0.061)	(0.052)	(0.045)
Controls for other shock categories		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture-year FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Five other controls		No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations		1488	1369	1488	1488	1488	1369	1488	1488	1488	1369	1488	1488
Adjusted R^2		0.931	0.926	0.925	0.930	0.978	0.977	0.982	0.978	0.903	0.918	0.899	0.915

***Note*s**: This table tests the robustness of the baseline estimates from Table 3 to the omitted variables (columns 1a, 1b, and 1c), extreme rainfall (columns 2a, 2b, and 2c), an alternative definition of rainfall shock (columns 3a, 3b, and 3c), and spatial correlation (columns 4a, 4b, and 4c). All robustness checks have the same model setting as the baseline estimation except for that specified in each check. In column 1–3, the standard errors in parentheses are clustered at the village level. In column 4, the standard errors in parentheses represent Conley spatial standard errors with a radius of 500 km. The significance levels are *** $p<0.01$, ** $p<0.05$, and * $p<0.1$.

Rainfall shocks in other months. Our main analysis follows the literature to define rainfall shock based on the rainfall in the first month of the growing season. This definition is some kind of arbitrary because rainfall shocks in other months are also likely to affect agricultural productivity. To show this, we alternatively define rainfall shock based on each of the months before and after the first month of the growing season. Appendix Figure 7 presents the estimated effect of a lagged positive rainfall shock in each month (relative to the first month of the growing season) on the village income. We find that the effects are negative and large for lagged positive rainfall shocks in most of the 12 months, especially in months that are close to the first month of the growing season. However, only the effect in the first month of the growing season is precisely estimated and statistically significant at the conventional level.[30]

Dynamic effects. An interesting question is how many years does the effect of a positive rainfall shock last? To answer this question, we extend the baseline model (5) to include additional lagged years of the positive rainfall shock dummy:

$$\begin{aligned}\ln(Income_{v,t}) = \alpha_0^l + \alpha_1^l Pos_{v,t-l}\prod_{j=0}^{l-1} None_{v,t-j} + \alpha_2^l Neg_{v,t-l}\prod_{j=0}^{l-1} None_{v,t-j} \\ + \sum_{j=0}^{l-1}\alpha_3^j Pos_{v,t-j} + \sum_{j=0}^{l-1}\alpha_4^j Neg_{v,t-j} + X_{v,t}\eta + \gamma_v + \gamma_{pt} + \epsilon_{v,t} \quad , \qquad (9)\end{aligned}$$

where l denotes the lagged years of the positive rainfall shock dummy. This model is identical to the baseline model if $l = 1$. Following the same logic, the coefficient of interest is α_1^l, which measures the effect of the interaction between an l-year lagged positive rainfall shock and no shock in all the following years (denoted by $\prod_{j=0}^{l-1} None_{v,t-j}$). The omitted reference group is no shock in all the relevant years. We estimate model (9) for $l = 1, 2, 3, 4$ and report our estimates of α_1^l in Appendix Figure 8. We find no significant effect of a positive rainfall shock that lagged by more than 1 year. This finding suggests that farmers can adjust back to the optimal factor allocation after realizing that they are overoptimistic.

[30] An interesting observation from Appendix Figure 7 is that the impacts during the three months preceding the first month of the growing season appear to be larger than the impacts observed in most months thereafter. However, it is worth noting that these impacts are not statistically significant at the conventional level. One potential explanation for this pattern is that rainfall during the three months preceding the growth season can significantly influence water availability during the planting season, and crops tend to be more sensitive to water availability during this crucial period (Kaur 2019, Gatti et al. 2021, Premand & Stoeffler 2022). On the other hand, the impacts for months following the first month of the growing season are relatively small, potentially due to the fact that most agricultural production decisions are made either before or at the beginning of the growing season.

5.2 Labor misallocation

We now turn to estimating the effect on factor allocation. All estimations are based on the household-level regression model (6). To verify the predicted differential effects across farmers, we estimate the model separately for households from different productivity quartiles.[31] Our main analyses indexed household agricultural productivity by the TFP measure of ϕ_h constructed in Subsection 4.4. Each estimation includes all the variables specified in model (6), but we report only the estimated coefficient of interest, $\hat{\delta}_1$, for simplicity.

Figure 4 presents the estimated effect of a lagged positive rainfall shock on four labor allocation measures: *whether the household has any farm work* (top left), *the number of family members who participate in farm work* (bottom left), *whether the household has any off-farm work* (top right), and *the number of family members who participate in off-farm work* (bottom right). Consistent with the theoretical prediction, we find that a lagged positive rainfall shock significantly shifts labor from off-farm work to farm work for low-productivity households but has no significant effect on high-productivity households. For example, the top-left panel of the figure shows that a lagged positive rainfall shock significantly increases the likelihood of participating in farm work by 9.7 percentage points and 5.6 percentage points for households in the first and second TFP quartiles, respectively, but has no effect on households in the third and fourth TFP quartiles.[32] We have also tried all the robustness checks used in the last subsection (i.e., omitted variables, rainfall extremes, alternative rainfall shock definitions, and dynamic effects) and found the same results.

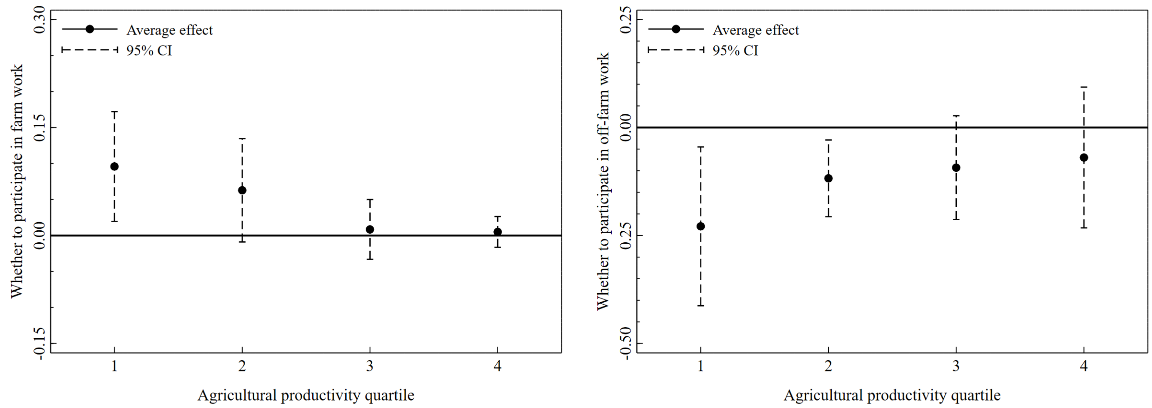
To enhance the interpretation of the estimates, we convert the percentage point effects into percentage effects for the first quartile of households. In the first quartile, the effect on the likelihood of farm work is 0.097, with a mean likelihood of farm work at 0.816. Consequently, a lagged positive rainfall shock increases farm work by 11.9 percent. Similarly, we can calculate that a lagged positive rainfall shock leads to a 18.6 percent increase in the number of farm work family members, a 41.2 percent reduction in the likelihood of off-farm work, and a 64.3 percent reduction in the number of off-farm work family members. The seemingly substantial percentage effect on off-farm work for the first quartile of households should be interpreted cautiously. The NFPS survey defines a rural resident as having off-farm work if they have any off-farm income

[31] The estimates for the whole sample are presented in Appendix Tables 8 and 9. As anticipated, the estimates for the entire sample primarily exhibit statistical insignificance, primarily because of the substantial variation in effects across quartiles.

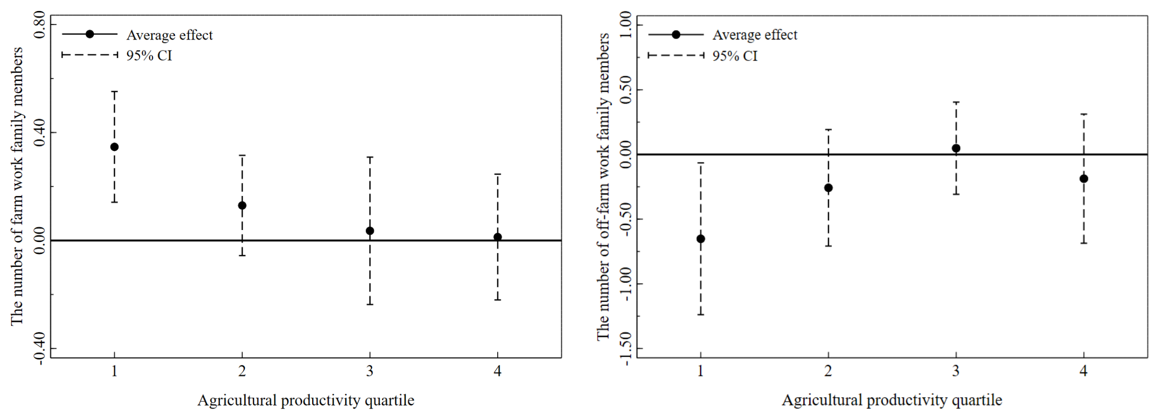
[32] In Appendix Table 1, we present the mean and standard deviation values for all the outcome variables across the TFP quartiles.

in the year of the survey. Considering the widespread existence of part-time off-farm work in rural China (Zhang et al. 2018), the estimated effect should be interpreted as that a lagged positive rainfall shock has a significantly negative effect on the off-farm work of 41.2 percent of rural residents in the first quartile.[33] The large percentage impact on the number of people with off-farm work may actually correspond to a much smaller impact on the total off-farm working time. For example, if an average rural resident with off-farm work spends only one-third of their working time on off-farm work, the percentage impact on off-farm working time would be two-thirds smaller.

[33]An important limitation of our dataset is the lack of information on the number of hours worked. This data would be valuable in assessing the magnitude of the impact of lagged positive rainfall shocks on off-farm work.



Mean	0.816	0.895	0.940	0.974	Mean	0.534	0.599	0.597	0.599



Mean	1.722	1.946	1.942	1.984	Mean	0.996	1.142	1.091	1.065
SD:	1.234	1.240	1.120	1.157	SD:	1.088	1.113	1.079	1.072

Figure 4: The effect of lagged positive rainfall shock on labor allocation

Note: This figure estimates versions of model (6) that use the dependent variables of *whether the household has any farm work* (left top), *the number of family members who participate in farm work* (left bottom), *whether the household has any off-farm work* (right top), and *the number of family members who participate in off-farm work* (right bottom). We estimate each model separately for households from different agricultural productivity quartiles (measured by ϕ_h, as constructed in Subsection 4.4). Each point reported in the figure is the estimated coefficient of the indicator of a positive shock last year and no shock this year, i.e., δ_1. The 95% confidence intervals are calculated based on standard errors clustered at the village level.

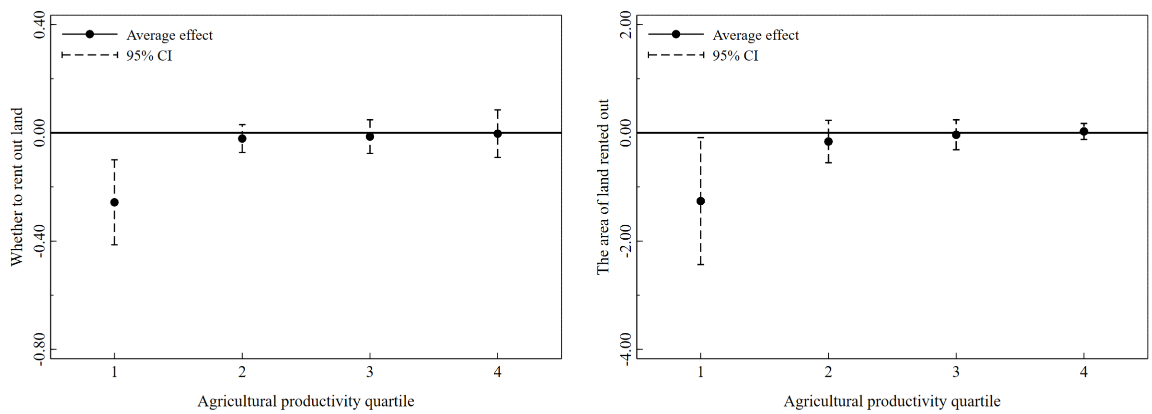
We address several potential concerns by constructing alternative measures of TFP. Firstly, we consider the concern that the TFP measure constructed using data from the entire sample period may be endogenous to outcome variables. To mitigate this concern, we estimate agricultural productivity using data from the first two years of the sample and then estimate the effects for each quartile using data from the remaining years. Appendix Figure 9 illustrates that the resulting estimates are highly similar. Secondly, we address the concern that the agricultural productivity measure employed in the main analysis may lack transparency. To alleviate this concern, we present Appendix Figures 11 and 12, which demonstrate that the estimated effects remain comparable when productivity is measured using the average crop yield per household and the marginal productivity of farmland at the household level (represented by λ_h, as constructed in Subsection 4.4), respectively.

Appendix Table 10 presents another piece of evidence of labor misallocation: a lagged positive rainfall shock significantly increases the wage of hired agricultural labor. This finding is consistent with that of Kaur (2019), who shows that a lagged positive rainfall shock increases agricultural wages and thus leads to inefficient labor allocation in Indian villages. However, our main analysis does not depend on this evidence because only 9.3 percent of households in China hire labor for agricultural production, which means that a conclusion can only be made based on 9.3 percent of the sample households. Nevertheless, we still find that the estimated effect on agricultural wages is robust to the control variables, rainfall extremes, and alternative rainfall shock definitions.

5.3 Land misallocation

Figure 5 presents the estimated effect on land allocation. The estimations are also based on model (6) and the dependent variables are *whether to rent out farmland* (left panel) and *the area of farmland rented out* (right panel). We still estimate the model separately for households from each productivity quartile (measured by ϕ_h) and report the estimate of δ_1 in the figure. Consistent with the theoretical prediction, we find that a lagged positive rainfall shock leads households with low agricultural productivity to significantly reduce the farmland that they rent out. Specifically, a lagged positive rainfall shock reduces the likelihood of renting out farmland by 26 percentage points and reduces the area of farmland rented out by 1.3 mu for households from the lowest productivity quartile, but has no significant effect on households from other three productivity quartiles.[34]

[34]The percentage impact on land renting appears to be significantly large in comparison to the mean. However, it is important to note that the reported mean represents the flow of new land renting that occurred within a single year, rather than the total stock of land renting. The actual stock of land renting could be much larger. Unfortunately, we cannot calculate the total amount of land rented out by each household as the land renting that took place before our sample period is not observed.



Mean	0.249	0.183	0.141	0.093
SD	–	–	–	–

Mean	2.442	1.394	1.192	1.075
SD	5.480	2.643	2.378	2.105

Figure 5: The effect of lagged positive rainfall shock on farmland allocation

Note: This figure estimates versions of model (6) that use *whether to rent out farmland* (left panel) and *the area of farmland rented out* (right panel) as the dependent variables. We estimate the model separately for households from each productivity quartile (measured by ϕ_h). Each point reported in the figure is the estimated coefficient of the indicator of a positive shock last year and no shock this year, i.e., δ_1. The 95% confidence intervals are calculated based on standard errors clustered at the village level.

A potential concern regarding the previous analysis is that it only demonstrates that low productivity farmers are less likely to rent their land out following a lagged positive rainfall shock. However, if a significant portion of land rental occurs within productivity quartiles, changes in agricultural activity may not necessarily lead to misallocation. To address this concern, Appendix Figure 3 illustrates that the amount of land rented by high productivity farmers also declines in response to a lagged positive rainfall shock.[35] Another piece of evidence that alleviates this concern is that a lagged positive rainfall shock increases the wage of hired agricultural labor, as shown in Appendix Table 10. We provide additional robustness checks, as discussed in the previous subsection, in Appendix Figures 12, 9, and 11.

[35] Due to the unavailability of a direct measure of farmland rented in within the NFPS dataset, we employ annual family land increase as a proxy for land rented in. However, we acknowledge that this proxy may not accurately capture land rented in and could also reflect the reclamation of land previously rented out in previous years. Consequently, we include these results solely in the appendix to acknowledge the limitations associated with using annual family land increase as a proxy for land rented in.

5.4 Alternative channels

The above analysis is based on the assumption that lagged positive rainfall shocks primarily impact factor allocation through farmers' irrational rainfall expectations. This assumption is supported by the nature of exogenous, unanticipated, and transitory rainfall shocks, which should not be influenced by any omitted factors related to factor allocation. However, it is important to address a potential concern that lagged positive rainfall shocks may also affect factor allocation through higher income, leading to consumption smoothing or relaxed budget constraints.

We explain why our findings are not primarily driven by consumption smoothing and relaxed budget constraints. Consumption smoothing suggests that higher income in the previous year should result in reduced income-generating activities in the current year. Therefore, if consumption smoothing is the dominant mechanism, we would expect to observe a decrease in working time among farmers following a lagged positive rainfall shock. Furthermore, since farm work is generally more physically demanding than off-farm work for smallholders in China, we would anticipate a greater reduction in farm work compared to off-farm work due to consumption smoothing. However, the observed increase in farm work following a lagged positive rainfall shock contradicts the expectations of consumption smoothing.

The impact of relaxed budget constraints on factor allocation is theoretically ambiguous and depends on whether agricultural production or off-farm work is subject to greater budget constraints. Existing evidence suggests that budget constraints significantly limit rural-urban migration in China (e.g., Bairoliya & Miller 2021, Garriga et al. 2023). Conversely, relaxed budget constraints could increase agricultural activity by enabling the adoption of costly inputs and technologies that enhance agricultural productivity and profitability. If low-productivity farmers are more economically disadvantaged and face greater budget constraints, relaxed budget constraints could motivate them to engage more in agriculture. However, our findings demonstrate that lagged positive rainfall shocks have no significant impact on agricultural physical capital and other crucial agricultural inputs such as seeds, fertilizer, and pesticides, as depicted in Appendix Figure 6. Moreover, controlling for these inputs, as well as agricultural physical capital, does not alter our main findings regarding the effects on factor allocation, as illustrated in Appendix Figures 4 and 5.[36] These results indicate that changes in budget constraints are unlikely to be the primary driving force behind the

[36]It should be noted that our main analysis does not include these variables as controls because lagged positive rainfall shocks could also influence other inputs and physical capital through farmers' irrational rainfall expectations.

estimated effects of lagged positive rainfall shocks on factor allocation.

5.5 Heterogeneous effects on income

The above findings on factor misallocation suggest that the damage to income should also be concentrated among low-productivity farmers. To confirm this, we estimate the effect of a lagged positive rainfall shock on household income based on model 6 and examine the effect separately for households from different productivity quantiles. The estimates of δ_1 are reported in Figure 6; the left panel measures agricultural productivity by ϕ_h, and the right panel measures it by λ_h. As expected, significant negative effects of a lagged positive rainfall shock on household income are found only for households in the lowest productivity quantile.[37] Similar results are obtained when using alternative income measures or applying other robustness checks used above.

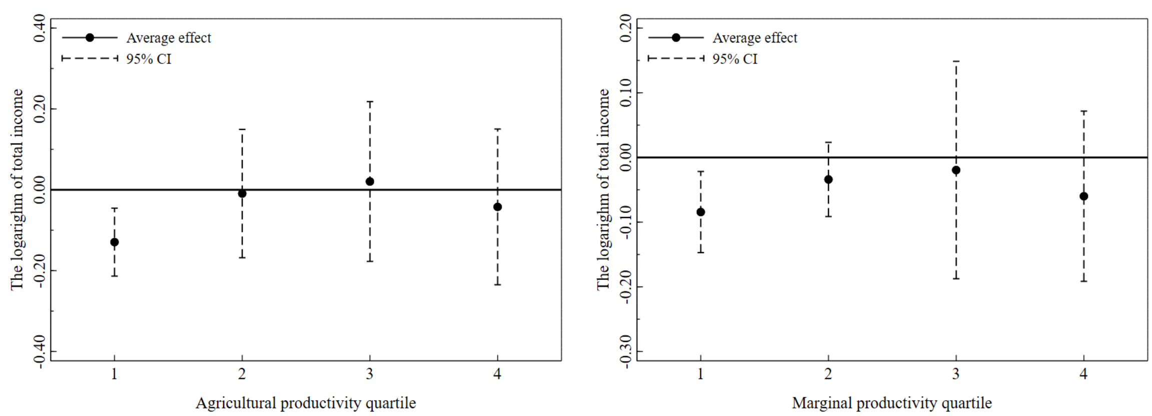


Figure 6: The effect of lagged positive rainfall shock on household-level income

Note: This figure estimates a version of model (6) that uses the *log household total income* as the dependent variable. We estimate the model separately for households from each productivity quartile, and we measure productivity by ϕ_h (left panel) and λ_h (right panel), constructed in Subsection 4.4. Each point reported in the figure is the estimated coefficient of the indicator of a positive shock last year and no shock this year, i.e., δ_1. The 95% confidence intervals are calculated based on standard errors clustered at the village level.

[37] The estimated effect for the lowest quantile is smaller when using λ_h as the productivity measure (than when using ϕ_h as the measure) because the effects on the second and fourth quantiles are also negative and large, suggesting that ϕ_h is indeed a better measure of productivity than λ_h.

5.6 Irrigation as a smoother

If the estimated factor misallocation truly reflects the irrational responses of farmers to lagged positive rainfall shocks, we should find a weaker impact on farmers with good irrigation conditions. This is because the agricultural production of farmers with good irrigation conditions depends less on rainfall and thus should have a weaker response to the expected positive rainfall shock. In other words, we can use the different responses of farmers with different irrigation conditions to verify our findings. To do so, we classified sample households into four equally-sized groups based on their average irrigation expenditure per mu over the sample period. We use irrigation expenditure as a measure of irrigation conditions because irrigation area data are not available at the household level.

We estimate model (6) separately for each group and report the estimates of δ_1 in Figure 7. The estimates confirm that only farmers with relatively poor irrigation conditions significantly increase their agricultural activity in response to a lagged positive rainfall shock. Specifically, a lagged positive rainfall shock only leads farmers from the first and second irrigation quantiles to significantly reduce the farmland they rented out (left panel) and to significantly increase their chance of agriculture production (right panel).[38]

[38]We also find that farmers with good irrigation conditions (those in the fourth quantile) tend to reduce their agricultural activity in response to a lagged positive rainfall shock. A potential explanation for this is that the within-village general equilibrium effects push up factor prices, and thus, crowd out these farmers.

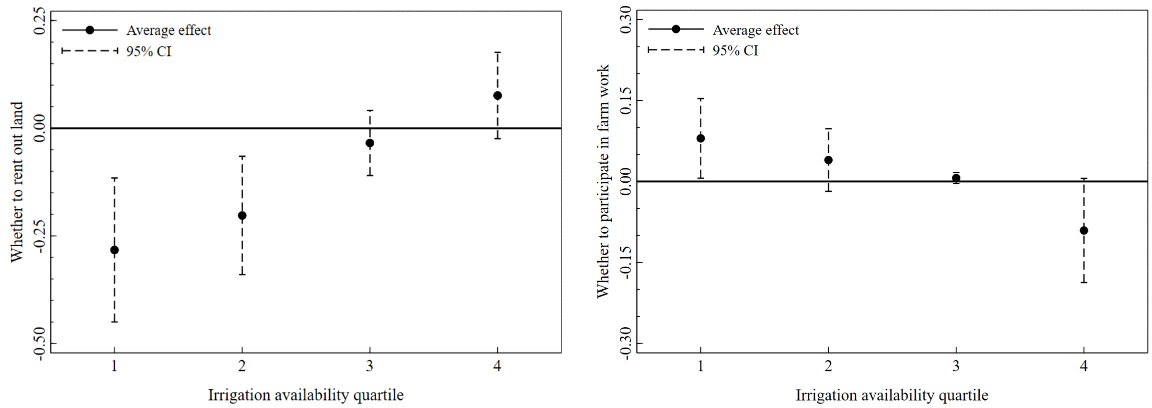


Figure 7: Irrigation condition and the effect of lagged positive rainfall shock on factor allocation

***Note*:**This figure estimates versions of model (6) that use *whether to rent out farmland* (left) and *whether to participate in farm work* (right) as the dependent variables. We classified sample households into four equally-sized groups based on their irrigation conditions, and we estimate the model separately for each group. Each point reported in the figure is the estimated coefficient of the indicator of a positive shock last year and no shock this year, i.e., δ_1. The 95% confidence intervals are calculated based on standard errors clustered at the village level.

6 Concluding Remarks

The random and transitory features of rainfall shock provide a unique opportunity to test the rationality of farmers in making agricultural decisions. We found that a positive rainfall shock last year significantly increases the agricultural activity of low-productivity farmers this year. This behavior change can be taken as evidence that farmers are not fully rational because the random and transitory rainfall shock last year should have no direct effect on agricultural productivity this year.

We developed a theoretical model to facilitate understanding of the welfare impact of the irrational response of farmers. In the model, individual heterogeneous farmers determine the labor they allocate to farm and off-farm work according to their own agricultural productivity and the predicted rainfall outcome. An overoptimistic prediction of rainfall motivates farmers with relatively low agricultural productivity to increase their agricultural production. This irrational behavior leads to aggregate welfare loss by distorting factor markets: the misallocation of labor from off-farm work to farm work reduces the aggregate returns to labor, and the misallocation of farmland to less-productivity farmers reduces the aggregate returns to farmland.

Empirical findings support these predictions. Based on panel data from more than 10,000 households, we estimated that a lagged positive rainfall shock reduces the household-average income by 8.1 percent. In addition, we estimated that a lagged positive rainfall shock significantly reallocates farmland from high-productivity farmers to low-productivity farmers and reallocates labor from high-income off-farm work to low-income farm work. These effects are primarily driven by the irrational response of low-productivity farmers. We also showed that farmers with good irrigation conditions are generally not damaged by lagged positive rainfall shocks because their agricultural production depends less on rainfall.

Bibliography

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. (2022), 'When should you adjust standard errors for clustering?', *The Quarterly Journal of Economics* **138**(1), 1–35.

Adamopoulos, T., Brandt, L., Leight, J. & Restuccia, D. (2022), 'Misallocation, selection, and productivity: A quantitative analysis with panel data from china', *Econometrica* **90**(3), 1261–1282.

Adamopoulos, T. & Restuccia, D. (2014), 'The size distribution of farms and international productivity differences', *American Economic Review* **104**(6), 1667–97.

Bairoliya, N. & Miller, R. (2021), 'Social insurance, demographics, and rural-urban migration in china', *Regional Science and Urban Economics* **91**, 103615.

Beaudry, P. & Willems, T. (2022), 'On the macroeconomic consequences of over-optimism', *American Economic Journal: Macroeconomics* **14**(1), 38–59.

Bénabou, R. & Tirole, J. (2016), 'Mindful economics: The production, consumption, and value of beliefs', *Journal of Economic Perspectives* **30**(3), 141–64.

Benjamin, D., Brandt, L. & Giles, J. (2005), 'The evolution of income inequality in rural china', *Economic Development and Cultural Change* **53**(4), 769–824.

Bodner, G., Nakhforoosh, A. & Kaul, H.-P. (2015), 'Management of crop water under drought: a review', *Agronomy for Sustainable Development* **35**(2), 401–442.

Brückner, M. & Gradstein, M. (2013), 'Effects of transitory shocks to aggregate output on consumption in poor countries', *Journal of International Economics* **91**(2), 343–357.

Busse, M. R., Pope, D. G., Pope, J. C. & Silva-Risso, J. (2015), 'The psychological effect of weather on car purchases', *The Quarterly Journal of Economics* **130**(1), 371–414.

Caselli, F. (2005), 'Accounting for cross-country income differences', *Handbook of economic growth* **1**, 679–741.

Chari, A., Liu, E. M., Wang, S.-Y., Wang, Y. & Schoenberg, U. (2021), 'Property rights, land misallocation, and agricultural efficiency in china', *The Review of Economic Studies* **88**(4), 1831–1862.

Charney, J., Quirk, W. J., Chow, S. & Kornfield, J. (1977), 'A comparative study of the effects of albedo change on drought in semiarid regions', *Journal of the atmospheric sciences* **34**(9), 1366–1385.

Chen, C., Baethgen, W. E. & Robertson, A. (2013), ‘Contributions of individual variation in temperature, solar radiation and precipitation to crop yield in the north china plain, 1961–2003’, *Climatic change* **116**(3), 767–788.

Ciccone, A. & Ismailov, A. (2022), ‘Rainfall, agricultural output and persistent democratization’, *Economica* **89**(354), 229–257.

Colella, F., Lalive, R., Sakalli, S. & Thoenig, M. (2020), ‘Acreg: Stata module to perform arbitrary correlation regression’.

Colmer, J. (2020), ‘Rainfall variability, child labor, and human capital accumulation in rural ethiopia’, *American Journal of Agricultural Economics* **103**(3), 858–877.

Conlin, M., O’Donoghue, T. & Vogelsang, T. J. (2007), ‘Projection bias in catalog orders’, *American Economic Review* **97**(4), 1217–1249.

Cui, X. & Xie, W. (2022), ‘Adapting agriculture to climate change through growing season adjustments: Evidence from corn in china’, *American Journal of Agricultural Economics* **104**(1), 249–272.

Dawson, C., de Meza, D., Henley, A. & Arabsheibani, G. R. (2014), ‘Entrepreneurship: cause and consequence of financial optimism’, *Journal of Economics & Management Strategy* **23**(4), 717–742.

Dell, M., Jones, B. F. & Olken, B. A. (2014), ‘What do we learn from the weather? the new climate-economy literature’, *Journal of Economic Literature* **52**(3), 740–98.

Donovan, K. (2016), ‘Agricultural risk, intermediate inputs, and cross-country productivity differences’, *Unpublished Working Paper, University of Notre Dame* .

Duflo, E., Kremer, M. & Robinson, J. (2011), ‘Nudging farmers to use fertilizer: Theory and experimental evidence from kenya’, *American Economic Review* **101**(6), 2350–2390.

Fan, Q. & Salas Garcia, V. B. (2018), ‘Information access and smallholder farmers’ market participation in peru’, *Journal of Agricultural Economics* **69**(2), 476–494.

Feng, X., Porporato, A. & Rodriguez-Iturbe, I. (2013), ‘Changes in rainfall seasonality in the tropics’, *Nature Climate Change* **3**(9), 811–815.

Garriga, C., Hedlund, A., Tang, Y. & Wang, P. (2023), ‘Rural-urban migration, structural transformation, and housing markets in china’, *American Economic Journal: Macroeconomics* **15**(2), 413–440.

Gatti, N., Baylis, K. & Crost, B. (2021), ‘Can irrigation infrastructure mitigate the effect of rainfall shocks on conflict? evidence from indonesia’, *American Journal of Agricultural Economics* **103**(1), 211–231.

Gollin, D., Lagakos, D. & Waugh, M. E. (2014), ‘Agricultural productivity differences across countries’, *American Economic Review* **104**(5), 165–70.

Grabrucker, K. & Grimm, M. (2020), 'Is there a rainbow after the rain? how do agricultural shocks affect non-farm enterprises? evidence from thailand', *American Journal of Agricultural Economics* **103**(5), 1612–1636.

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

Jehiel, P. (2018), 'Investment strategy and selection bias: An equilibrium perspective on overoptimism', *American Economic Review* **108**(6), 1582–97.

Kaur, S. (2019), 'Nominal wage rigidity in village labor markets', *American Economic Review* **109**(10), 3585–3616.

Kazianga, H. & Udry, C. (2006), 'Consumption smoothing? livestock, insurance and drought in rural burkina faso', *Journal of Development economics* **79**(2), 413–446.

Lagakos, D. & Waugh, M. E. (2013), 'Selection, agriculture, and cross-country productivity differences', *American Economic Review* **103**(2), 948–80.

Malmendier, U. & Taylor, T. (2015), 'On the verges of overconfidence', *Journal of Economic Perspectives* **29**(4), 3–8.

Miguel, E., Satyanath, S. & Sergenti, E. (2004), 'Economic shocks and civil conflict: An instrumental variables approach', *Journal of political Economy* **112**(4), 725–753.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C. & Thépaut, J.-N. (2021), 'Era5-land: a state-of-the-art global reanalysis dataset for land applications', *Earth System Science Data* **13**(9), 4349–4383.

Nordman, C. J., Sharma, S. & Sunder, N. (2022), 'Here comes the rain again: Productivity shocks, educational investments, and child work', *Economic Development and Cultural Change* **70**(3), 1041–1063.

O'Gorman, P. A. & Schneider, T. (2009), 'The physical basis for increases in precipitation extremes in simulations of 21st-century climate change', *Proceedings of the National Academy of Sciences* **106**(35), 14773–14777.

Pan, Y., Qin, Y., Zhang, F. & Zhu, H. (2022), 'Acquiring land in cold winter: Consequences and possible explanations', *Journal of Environmental Economics and Management* **116**, 102721.

Paxson, C. H. (1992), 'Using weather variability to estimate the response of savings to transitory income in thailand', *The American Economic Review* pp. 15–33.

Premand, P. & Stoeffler, Q. (2022), 'Cash transfers, climatic shocks and resilience in the sahel', *Journal of Environmental Economics and Management* **116**, 102744.

Restuccia, D., Yang, D. T. & Zhu, X. (2008), 'Agriculture and aggregate productivity:

A quantitative cross-country analysis', *Journal of monetary economics* **55**(2), 234–250.

Rosenzweig, M. R. & Udry, C. (2014), 'Rainfall forecasts, weather, and wages over the agricultural production cycle', *American Economic Review* **104**(5), 278–283.

Sarsons, H. (2015), 'Rainfall and conflict: A cautionary tale', *Journal of Development Economics* **115**, 62–72.

Tao, F., Yokozawa, M., Liu, J. & Zhang, Z. (2008), 'Climate–crop yield relationships at provincial scales in china and the impacts of recent climate trends', *Climate Research* **38**(1), 83–94.

Taylor, S. E. & Brown, J. D. (1988), 'Illusion and well-being: a social psychological perspective on mental health.', *Psychological bulletin* **103**(2), 193.

Trenberth, K. E. (2011), 'Changes in precipitation with climate change', *Climate research* **47**(1-2), 123–138.

Wang, J., Zhu, Y., Sun, T., Huang, J., Zhang, L., Guan, B. & Huang, Q. (2020), 'Forty years of irrigation development and reform in china', *Australian Journal of Agricultural and Resource Economics* **64**(1), 126–149.

Weinstein, N. D. (1980), 'Unrealistic optimism about future life events.', *Journal of personality and social psychology* **39**(5), 806.

Wuepper, D. & Sauer, J. (2016), 'Explaining the performance of contract farming in ghana: The role of self-efficacy and social capital', *Food Policy* **62**, 11–27.

Zhang, L., Dong, Y., Liu, C. & Bai, Y. (2018), 'Off-farm employment over the past four decades in rural china', *China Agricultural Economic Review* .