

NBER WORKING PAPER SERIES

MISALLOCATION, SELECTION AND PRODUCTIVITY:  
A QUANTITATIVE ANALYSIS WITH PANEL DATA FROM CHINA

Tasso Adamopoulos  
Loren Brandt  
Jessica Leight  
Diego Restuccia

Working Paper 23039  
<http://www.nber.org/papers/w23039>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
January 2017, Revised September 2021

We are grateful to Chaoran Chen for providing excellent research assistance. We also thank for useful comments the editor, three anonymous referees, Gueorgui Kambourov, Kjetil Storesletten, Carlos Urrutia, Daniel Xu, and seminar participants at Toronto, McGill, Ottawa, REDg Barcelona, NBER, IMF, Cornell, Michigan, UBC, Xiamen U., ASU, Tsinghua-UCSD, HKUST, Western, Tsinghua, EEA Geneva, Banco de Mexico, Montreal, Ohio State, Universitat de Barcelona, McMaster, CIREQ Montreal, AMES Hong Kong, SED Edinburgh, Oslo, ESEM Lisbon, Bank of Canada/Toronto, Duke, Georgetown, EIEF Rome, CEMFI, ENSAI Rennes, Bank of Chile, Bank of Italy, Simon Fraser, National University of Singapore, Singapore Management U, ITAM, Berkeley-ARE, and Guelph. Adamopoulos, Brandt, and Restuccia gratefully acknowledge the support from the Social Sciences and Humanities Research Council of Canada. Brandt thanks the support from the Noranda Chair for International Trade and Economics. Restuccia gratefully acknowledges the support from the Canada Research Chairs program and the Bank of Canada Fellowship program. The views expressed herein are not necessarily those of the Bank of Canada and are the author's alone. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Tasso Adamopoulos, Loren Brandt, Jessica Leight, and Diego Restuccia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China  
Tasso Adamopoulos, Loren Brandt, Jessica Leight, and Diego Restuccia  
NBER Working Paper No. 23039  
January 2017, Revised September 2021  
JEL No. O11,O14,O4

### ABSTRACT

We use household-level panel data from China and a quantitative framework to document the extent and consequences of factor misallocation in agriculture. We find that there are substantial within- village frictions in both the land and capital markets linked to land institutions in rural China that disproportionately constrain the more productive farmers. These frictions reduce aggregate agricultural productivity by affecting two key margins: (1) the allocation of resources across farmers (misallocation) and (2) the allocation of workers across sectors, in particular the type of farmers who operate in agriculture (selection). Selection substantially amplifies the productivity effect of distortionary policies by affecting occupational choices that worsen average ability in agriculture.

Tasso Adamopoulos  
York University  
Canada  
aadamo@yorku.ca

Jessica Leight  
International Food Policy Research Institute (IFPRI)  
Washington, DC  
j.leight@cgiar.org

Loren Brandt  
Department of Economics  
University of Toronto  
150 St. George Street  
Toronto, ON M5S 1A1 CANADA  
loren.brandt@utoronto.ca

Diego Restuccia  
Department of Economics  
University of Toronto  
150 St. George Street  
Toronto, ON M5S 3G7  
CANADA  
and NBER  
diego.restuccia@utoronto.ca

# 1 Introduction

A central theme in the study of economic growth and development is the large productivity differences in the agricultural sector across countries. Since labor in poor countries is primarily allocated to agriculture, understanding these differences is essential in accounting for aggregate income differences between rich and poor countries (Gollin et al., 2002; Restuccia et al., 2008). Productivity gaps in agriculture between developing and developed countries are also consistent with increasing evidence that resource misallocation across households that are heterogeneous in skill is more prevalent in developing countries.<sup>1</sup> Institutions and policies giving rise to misallocation are highly pervasive in agriculture in poor countries and can account for a large portion of the productivity differences across countries.<sup>2</sup> These institutions often diminish the efficiency of land and other complementary markets in directing resources to their most productive uses.

We use micro farm-level panel data from China and a quantitative framework to document the extent and consequences of factor misallocation in agriculture. We find that there are substantial within-village frictions in both the land and capital markets in rural China that disproportionately affect the more productive farmers. We argue that these distortions reduce aggregate agricultural productivity by affecting two key margins: (1) the allocation of resources across farmers (misallocation); and (2) the allocation of workers across sectors, in particular, the type of farmers who operate in agriculture (selection).

Empirically, we exploit the longitudinal nature of our micro data and estimate for each household permanent fixed effect farm-level productivity and distortions that have been adjusted for village-level factors and time effects. We show that these adjustments considerably limit the scope for measurement error in our farm-level measures that is typically associated with cross-sectional studies of misallocation. The focus on the family farm also skirts measurement reporting issues associated

---

<sup>1</sup>See Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Restuccia and Rogerson (2013, 2017) review the expanding literature on misallocation and productivity.

<sup>2</sup>See Adamopoulos and Restuccia (2014). Recent studies linking resource misallocation to land market institutions include: land reforms (Adamopoulos and Restuccia, 2020), the extent of marketed land across farm households (Restuccia and Santaaulalia-Llopis, 2015), and the role of land titling (Chen, 2017; Gottlieb and Grobovšek, 2019).

with attributing inputs to plots within farms.

Theoretically, the key insight of our paper is that there is an important interaction between selection and misallocation. Selection can amplify the misallocation effect of distortionary policies and influence the extent of measured misallocation through its effect on the distribution of productivity. Intuitively, institutions generating misallocation may have a particularly negative effect on more highly skilled farmers, who are then less likely to farm in agriculture, further reducing average agricultural productivity. A key conceptual novelty of our framework relative to the standard selection framework is that idiosyncratic frictions directly distort occupational choices even if there is no aggregate change or movements in aggregate relative prices. Further, we show that quantitatively this mechanism can have a large impact, especially when distortions are strongly positively correlated with productivity as is the case in China.

We focus on China for several reasons. First, China is a rapidly growing economy experiencing substantial reallocation within and across sectors. Yet, productivity growth in agriculture has been lacklustre, especially in the cropping sector, the focus of this paper. Second, the operational size of farm units in China is extremely small, only about 0.7 hectares on average, and has not increased over time. Third, institutionally, there is a lack of well-defined property rights over land, which can lead to both factor misallocation within agriculture and distortions of sectoral-occupational choices. And fourth, we have a unique panel dataset of households with detailed input and output information on all farm and non-farm activities over 1993-2002. The data allow us to construct precise real measures of value added and productivity at the farm-level, and to observe the incomes of the same households across sectors. Given a widespread shift into non-agricultural activities, these data offer a unique opportunity to examine the selection effect of distortionary policies.

To measure the extent of misallocation across farmers implied by the land market institutions in China, and to quantify its consequences for occupational choices and agricultural productivity, we proceed in two steps.<sup>3</sup> In the first step, we use a diagnostic tool from modern macroeconomics, a

---

<sup>3</sup>We do not study the effect of land market institutions on farm-level productivity. While insecurity over property rights may also affect the type of investments that households may make on their land (investments in irrigation and

heterogeneous firm-industry framework with minimal structure, to measure the deviations between marginal products and the overall extent of inefficiency. In this set-up, the weak property rights in China manifest themselves as “wedges” in marginal products, with the property that these wedges are larger for farmers with higher productivity, who are unable to accumulate additional land, and complementary factors, such as capital. To apply this framework, we exploit our longitudinal household data to estimate permanent fixed effect farm-level productivity and distortions. Our fixed effect panel estimates of TFP and distortions address an array of measurement issues typically associated with cross-sectional analyses of misallocation in both agricultural and manufacturing settings. To formally show the value of our approach in dealing with measurement error we apply the method in [Bils et al. \(2017\)](#), which assesses the extent of additive measurement error in measures of distortions. We find that our panel fixed-effect estimates of distortions not only reduce measurement error relative to a cross-sectional measure of distortions, but virtually eliminate it.

We find that the aggregate output (productivity) gains from reallocation are sizeable. Using our fixed effect estimates of productivity and distortions, reallocation of factors across existing farmers within villages to their efficient use increases agricultural TFP by 24.4 percent. In addition, allowing factor inputs to be allocated efficiently across villages, the gains more than double to 53.2 percent, with more than two-thirds of these gains accounted for by labor mobility across villages. This suggests that the land policy in China may be a contributor to the barriers of labor mobility across space ([De Janvry et al., 2015](#); [Ngai et al., 2019](#)) and their depressing impact on productivity.

In the second step, we embed the agricultural village framework into a two sector model in order to study the impact of misallocation in agriculture on the selection of individuals between agriculture and non-agriculture. We use the equilibrium properties of the model to calibrate the parameters to observed conditional moments and targets from estimates of household fixed effects using panel micro data for China. In particular, the substantial reallocation of households from agriculture to non-agriculture and their cross-sector income correlation provide discipline for the analysis. We

---

drainage, long-term soil fertility, etc.) and other related investments, we focus on the role that insecure land rights play for the operation of land markets.

conduct a counterfactual experiment to assess the quantitative importance of misallocation and its overall impact on agricultural productivity, accounting for distortions in sectoral occupational choices. In particular, we assess the effect of the land market institutions on village-level productivity by eliminating the within-village correlation of distortions with farm-level productivity. This results in an increase of agricultural productivity of almost 3-fold, an increase in aggregate agricultural TFP of 1.7-fold, and a substantial reallocation of labor across sectors, with the share of employment in agriculture falling from 46 percent to 16 percent. The total effect on agricultural productivity is substantially larger than the productivity effect of eliminating misallocation across existing farmers. The difference is due to the substantial amplification effect that distortions have on the selection of farmers in the model, which produces an additional increase in agricultural TFP of 1.5-fold. That is, selection quadruples the impact of reduced misallocation on agricultural productivity.

Our paper contributes to the broad literature on misallocation and productivity by addressing two essential issues emphasized in [Restuccia and Rogerson \(2017\)](#). First, we link misallocation to specific policies/institutions, in our context, land market institutions in China.<sup>4</sup> Second, we study the broader impact of misallocation, in particular, the effect of distortionary policies on the selection of skills across sectors, which substantially amplifies the productivity losses from factor misallocation. In this context, our paper relates to the role of selection highlighted in [Lagakos and Waugh \(2013\)](#). A key difference is that we empirically document the role of distortions in the agricultural sector as the key driver of low agricultural productivity and show that these distortions generate much larger effects on selection than equivalent changes in economy-wide TFP. That is, our selection results are distinct from [Lagakos and Waugh \(2013\)](#) in that they are driven by the direct impact of idiosyncratic distortions on occupational choices.

The paper proceeds as follows. In the next section, we describe the specifics of the land market

---

<sup>4</sup>Our paper also relates to earlier studies of the Chinese economy emphasizing the role of agriculture ([Lin, 1992](#); [Zhu, 2012](#)); the importance of misallocation across provinces and between the state and non-state sectors ([Brandt et al., 2013](#)); and growth in economic transition ([Song et al., 2011](#)). [Tombe and Zhu \(2019\)](#) and [Ngai et al. \(2019\)](#) analyze the impact of migration restrictions in the “hukou” system for reallocation and welfare in China.

institutions in China. Section 3 describes the panel data from China and the variables we use in our analysis. In section 4, we present the basic framework for identifying distortions and measuring the gains from reallocation and present the main results on misallocation in agriculture in China. Section 5 embeds the agricultural framework into a two-sector model and reports the main quantitative results. We conclude in Section 6. Additional details and results are provided in Supplemental Material Appendix.

## 2 Land Market Institutions in China

The Household Responsibility System (HRS), established in rural China in the late 1970s and early 1980s, dismantled the system of collective management set up under Mao and extended use rights over farmland to rural households. These reforms triggered a spurt in productivity growth in agriculture in the early 1980s that subsequently dissipated. This level effect is often attributed to the improved effort incentives for households as they became residual claimants in farming (McMillan et al., 1989; Lin, 1992). Ownership of agricultural land however remained vested with the collective, and in particular the village or small group, a unit below the village. Use rights to land were administratively allocated among rural households by village officials on a highly egalitarian basis that reflected household size. In principle, all individuals with “registration” (*hukou*), in the village were entitled to land.

The law governing the HRS provided secure use rights over cultivated land for 15 years (in the late 1990s use rights were extended to 30 years), however village officials often reallocated land among households before the 15-year period expired. Benjamin and Brandt (2002) document that in over two-thirds of all villages reallocations occurred at least once, and on average more than twice. Their survey data show that reallocations undertaken between 1983-1995 typically involved three-quarters of all households in the village, and most of village land. A primary motivation of the reallocations was to accommodate demographic changes within a village. In addition, village officials reallocated

land from households with family members working off the farm to households solely engaged in agriculture (Brandt et al., 2002; Kung and Liu, 1997).

In principle, households had the right to rent or transfer their use rights to other households (*zhuanbao*), however in practice these rights were abridged in a variety of ways, resulting in thin land rental markets. Brandt et al. (2002) document that in 1995, while 71.6 percent of villages reported no restrictions on land rental activity, households rented out less than 3 percent of their land, with most rentals occurring among family members or close relatives, hence not necessarily directing the land to the best uses. The limited scope for farm rental activity is frequently associated with perceived “use it or lose it” rules: households that did not use their land and either rented it to others or let it lie fallow risked losing the land during the next reallocation. As a result, households may have been deterred from renting out land because of fear that it may be viewed by village officials as a signal that the household did not need the land (see, for example, Yang, 1997). Finally, we note that lack of ownership of the land also meant that households could not use it as collateral for purposes of borrowing.

The difficulty in consolidating land either through land purchases or land rentals is one of the reasons that operational sizes of farms have been typically very low in China and have not changed much over time. According to the World Census of Agriculture of the Food and Agricultural Organization in 1997 average farm size in China was 0.7 hectares. Contrast this to the United States where in the same year average farm size was 187 hectares or to Belgium and the Netherlands—two developed countries with similar arable land per person as China—where average farm size is around 16-17 hectares. Moreover, in developed countries, farm size is growing over time.<sup>5</sup>

The administrative egalitarian allocation of land combined with the limited scope for land rentals implied that more able farmers or those that valued land more highly were not able to increase operational farm size. To the extent that village officials either do not observe farmer ability (unobserved heterogeneity) or do not make land allocation decisions based on ability (egalitarian

---

<sup>5</sup>Small operational farm scales are not unique to China, as average farm sizes among the poorest countries in the world are below 1 hectare and also reflect low productivity in agriculture (Adamopoulos and Restuccia, 2014).



concerns), reallocations were unhelpful in improving operational scale and productivity ([Benjamin and Brandt, 2002](#)). These frictions in the land market could generate allocative inefficiency or misallocation by distorting the allocation of land across farmers. Also, the inability to use land as collateral for borrowing purposes could result in the misallocation of other inputs such as capital.

### 3 Data

We use household survey data collected by the Research Center for the Rural Economy under the Ministry of Agriculture of China.<sup>6</sup> This is a nationally representative survey that covers all provinces. The survey has been carried out annually since 1986 with the exception of 1992 and 1994 when funding was an issue. An equal number of rich, medium and poor counties were selected in each province, and within each county a similar rule was applied in the selection of villages. Within villages, households were drawn in order to be representative. Important changes in survey design in 1993 expanded the information collected on agriculture, and enabled more accurate estimates of farm related variables.

We have data for ten provinces that span all the major regions of China, and use the data for the period between 1993 and 2002. The data are in the form of an unbalanced panel. In each year, we have information on approximately 8000 households drawn from 110 villages. For 104 villages, we have information for all 9 years. The average number of household observations per village-year is 80, or a quarter to a third of all households in a village. We have data for all 9 years for approximately 6000 households. Attrition from the sample is not a concern and is examined in detail in [Benjamin et al. \(2005\)](#). Much of the attrition is related to exit of entire villages from the survey. Household exit and entry into the sample is not systematically correlated with key variables of interest. During the period of our study, migration of entire households was severely restricted. Our main unit of observation is the household farm. The survey provides disaggregated information

---

<sup>6</sup>For a detailed description and analysis of the data see [Benjamin et al. \(2005\)](#).

on household income and labor supply by activity. For agriculture, we have data on total household land holdings, sown area and physical output by crop, and major farm inputs including labor, fertilizer, and farm machinery. Regarding non-agricultural activities, for family businesses we have information on revenues, expenditures, and net incomes from each type of household non-family business. We also know household wage earnings.

The richness of the data on crops, inputs, and prices allows us to construct accurate real measures of output (value added) and productivity at the farm-level. We use sample-wide average prices (unit values) of crops and intermediate inputs over 1993-2002 to aggregate output and compute value added for each farm. Hence, our real measures of value added are double real, constant prices over time and common prices across households. Supplemental Material Appendix A describes in more detail the measures of output and inputs that we use.

## 4 Measuring TFP and Misallocation in Agriculture

We describe an industry framework in order to assess the extent of misallocation in agriculture using the micro data from China. We use the framework and data to measure farm-level TFP and distortions in agriculture, providing important moments for the two-sector analysis. We also report the productivity gains from reducing farm-specific distortions.

### 4.1 Basic Framework

We consider a rural village economy indexed by  $v$  that at each date  $t$  produces a single good and is endowed with amounts of farm land  $L_{vt}$ , farm capital  $K_{vt}$ , and a finite number  $M_{vt}$  of farm operators indexed by  $i$ . Following [Adamopoulos and Restuccia \(2014\)](#), the production unit in the rural village economy is a family farm. A farm is a technology that requires the inputs of a farm operator (household), as well as the land and capital under the farmer's control. Farm operators are heterogeneous in their farming ability which we denote as  $s$ .

As in Lucas Jr (1978), the production technology available to farmer  $i$  in village  $v$  at time  $t$  with productivity  $s_{ivt}$  exhibits decreasing returns to scale in variable inputs and is given by the Cobb-Douglas function,

$$y_{ivt} = (A_a s_{ivt})^{1-\gamma} [\ell_{ivt}^\alpha k_{ivt}^{1-\alpha}]^\gamma, \quad (1)$$

where  $y$ ,  $\ell$ , and  $k$  denote real farm output, land, and capital. The parameter  $A_a$  is a common productivity term,  $\gamma < 1$  is the span-of-control parameter which governs the extent of returns to scale at the farm-level, and  $\alpha$  captures the relative importance of land in production.<sup>7</sup>

Our starting point is the efficient allocation of factors of production in the village economy with a given set of farmers, at any point in time, obtained from the solution to a simple planner’s problem that takes the distribution of productivities as given. We derive the efficient allocation that maximizes agricultural output given a set of inputs, and show that the efficient allocation involves allocating resources according to relative productivity, with more productive farms commanding more land and capital (see Supplemental Material Appendix B). We use this efficient allocation and the associated maximum aggregate agricultural output as a benchmark to contrast with the actual (distorted) allocations and the agricultural output in the Chinese economy.

Next, we estimate farm-specific distortions as implicit input and output wedges or taxes. These taxes are abstract representations that serve to rationalize as an equilibrium outcome the observed allocations in the Chinese economy. While this representation is not required for assessing the aggregate consequences of misallocation, as we can directly compare efficient allocations and output with actual data, it is useful in the two-sector analysis.<sup>8</sup> We construct the following summary

---

<sup>7</sup>For ease of exposition and tractability our framework abstracts from differences across farmers in the intensive margin of labor input. We deal with this by adjusting outputs and inputs in the data, obtaining a residual measure of farm TFP that is unaffected by this abstraction. We also abstract from intermediate inputs, and hence the corresponding variable for  $y$  in the data analysis is value added. We note however that since labor days and intermediate inputs may also be misallocated, our estimates of misallocation from this framework may be conservative.

<sup>8</sup>We derive the equilibrium of the distorted economy and describe our identification of the farm-input-specific distortions from the equilibrium conditions and data in Supplemental Material Appendix C.

measure of farm-specific distortions faced by farm  $i$  in village  $v$  at time  $t$ ,

$$TFPR_{ivt} = \frac{y_{ivt}}{\ell_{ivt}^\alpha k_{ivt}^{1-\alpha}}. \quad (2)$$

We note that  $TFPR$  corresponds to the concept of “revenue productivity” in [Hsieh and Klenow \(2009\)](#), and use this notation to make the analogy clear. Revenue productivity  $TFPR_{ivt}$  is proportional to a geometric average of the farm-specific land and capital distortions relative to the output distortion. With two inputs and one output we can separately identify only two of the three possible wedges, but this choice does not influence the magnitude of the overall farm-specific distortion.

We emphasize that  $TFPR$  is different from “physical productivity” or  $TFP$  which is,

$$TFP_{ivt} \equiv (A_a s_{ivt})^{1-\gamma} = \frac{y_{ivt}}{[\ell_{ivt}^\alpha k_{ivt}^{1-\alpha}]^\gamma}, \quad (3)$$

for farm  $i$  in village  $v$  at time  $t$ . Without distortions, farms with higher physical productivity  $TFP_{ivt}$  command more land  $\ell_{ivt}$  and capital  $k_{ivt}$ , and marginal products of each factor and  $TFPR$  equalize across farms; with idiosyncratic distortions this need not be the case.

Using the fact that total output is  $Y_{vt} = \sum_{i=1}^{M_v} y_{ivt}$ , it is straightforward to show that farm-level behavior in the presence of distortions aggregates up to a rural village-wide production function with aggregate land  $L_{vt}$ , capital  $K_{vt}$ , number of farmers  $M_{vt}$ , and aggregate (distorted) productivity  $TFP_{vt}$ . Under our identification of distortions from the data, allocations and aggregate distorted output  $Y_{vt}$  in the model coincide with actual data for China.

We measure aggregate agricultural output reallocation gains by comparing efficient output to actual output in the Chinese economy. Since aggregate factors  $K_v$ ,  $L_v$ , and  $M_v$  are held fixed, in this comparison, the output gains represent TFP gains.

## 4.2 Measuring Farm Productivity

Our measure of productivity at the farm-level is TFP, which we construct residually from the farm-level production function using equation (3) and data on operated land, capital, labor, and value added. In our framework, labor supply to agriculture is assumed to be the same across all households, however in the data households differ in the number of days worked on the farm. To make a consistent mapping of the data to model variables, we remove the variation in labor input by normalizing value added, land, and capital by total labor days.

Computing farm TFP requires values for the parameters  $\gamma$  and  $\alpha$ . The values we use are  $\gamma = 0.54$ , reflecting an income share of labor of 0.46, and  $\alpha = 2/3$ , implying a land income share of 0.36 and hence a capital income share of 0.18. These values are based on a variety of estimates for China. For instance, the average labor cost share over corn and rice crops estimated by [Jin et al. \(2010\)](#) over the period of our study 1993-2002 is 46 percent. Similar values are obtained for the labor share in developed economies such as the United States ([Adamopoulos and Restuccia, 2014](#)). Estimates for the land income share based on aggregate data agree on a value of 0.36 ([Cao and Birchenall, 2013](#); [Chow, 1993](#)). Given  $\gamma$ , the land share implies  $\alpha = 2/3$ . The implied capital share is 0.18, which is well within the range of estimates for the elasticity of capital found in the literature ([Cao and Birchenall, 2013](#); [Fan and Zhang, 2002](#); [Chow, 1993](#)). Our results are robust to reasonable ranges in these parameter values.

Based on the values of  $(\alpha, \gamma)$ , we compute TFP for each household, in each village, for each year. These TFP measures, however, may reflect measurement error, transitory output or input shocks, and unobserved village-specific characteristics, all of which can impact the dispersion of productivity, and the implied gains from reallocation. We address these issues by netting out the effect of time factors and other village characteristics to estimate permanent or farmer fixed effect levels of TFP. In particular, first we decompose the logarithm of farm-level TFP,

$$\log TFP_{ivt} = \mu_t^{TFP} + \mu_i^{TFP} + e_{ivt}^{TFP}, \quad (4)$$

where  $\mu_t^{TFP}$  is a year fixed effect component that captures time varying shocks to productivity that are common to all farmers;  $\mu_i^{TFP}$  is a farm-specific component that does not vary over time; and  $e_{ivt}^{TFP}$  captures idiosyncratic shocks specific to the farmer in a given year. We use fixed effect panel data methods to estimate equation (4) to extract the household fixed effects. Note that  $\mu_i^{TFP}$  is inclusive of village-specific effects that do not change over time, but differs across individuals in different villages. The village effects cannot be separately identified from (4) because they are collinear with household fixed effects. In a second step, we remove the village-specific effects by regressing the household fixed effect on village dummies ( $\mu_v$ ) and extracting the residual,

$$\mu_i^{TFP} = \mu_v^{TFP} + \zeta_i^{TFP}, \quad (5)$$

where  $\zeta_i^{TFP}$  is a fixed farm component that captures a farm's permanent productivity that is constant across years and purged of village-level factors. This procedure provides an estimate of the pure farm idiosyncratic fixed-effect (permanent) component  $\widehat{\zeta}_i^{TFP}$ .

Our baseline measure of farm TFP, which we denote as  $TFP_i$  (and is different from  $TFP_{ivt}$ ) is the exponential of the estimated household fixed effect  $\widehat{\zeta}_i^{TFP}$ . There are a few points to note. First, farm TFP is purged of changes over time and differences across villages. Second, the household fixed effect estimate does not include potential measurement error which is subsumed in the residual  $e_{ivt}^{TFP}$ . After partialling out the village fixed effect, dispersion is considerably lower: Measured by the standard deviation of the log, dispersion in the cross-sectional measures of farm TFP ranges between 0.68 and 0.75 over the ten year period and has a mean of 0.72. By contrast, the dispersion for our baseline fixed-effect measure of farm TFP is 0.35.<sup>9</sup> Focusing on misallocation associated with the permanent component of productivity yields a conservative estimate of misallocation as, for example, there could be additional misallocation effects from transitory shocks.

---

<sup>9</sup>With our baseline fixed effect measures, the 90/10 percentile ratio in farm TFP is 2.2-fold whereas the 75/25 percentile ratio is 1.5-fold (contrasted with the cross-sectional averages of 5.6 and 2.3 respectively).

**Land Quality.** An alternative interpretation of the differences in farm TFP is that they represent unobserved variation in land quality across households, a form of measurement error. Given that our micro data do not provide information on land quality at the farm-level, we use detailed land quality data from the Global Agro Ecological Zones (GAEZ) project of the Food and Agricultural Organization (FAO) to compute measures of land quality at the village level. Variation in this measure provides bounds on potential differences in land quality between households within villages. GAEZ provides data on a rich set of land quality characteristics relevant for agricultural production at the 5 arc-minute resolution (roughly cell size of  $10 \times 10$  kilometers) for the whole world. These data include soil attributes (e.g., fertility, depth), climate attributes (e.g., moisture, temperature), and terrain attributes (e.g., elevation, slope). Following [Adamopoulos and Restuccia \(2021\)](#), we use the detailed geographical data from GAEZ to construct a summary measure of land quality for each village. We find that the dispersion in this measure of land productivity is relatively small across our sample villages in China with a standard deviation of the log of 0.096.<sup>10</sup> By comparison, in the Philippines the dispersion of a similar measure of land quality across villages is 0.21 ([Adamopoulos and Restuccia, 2020](#)).

While the village dummies in our econometric procedure pick-up land quality differences across villages, land quality differences across households within a village may explain the variation in TFP across households we observe. This is unlikely to be the case for two reasons. First, our estimate of land quality differences across villages may be interpreted as an upper bound on the extent of across household variation in land quality. This is because we expect differences in geographical characteristics to be larger across villages than across households within a village. Second, consistent with the egalitarian nature of land allocations at the village level, in locations where there are significant differences in land quality, we know from first-hand interviews with village

---

<sup>10</sup>For each village in our dataset, we construct an aggregate potential yield using the potential yield by crop, which is the maximum amount of output of a given crop that can be produced given the climate and soil conditions of the locality and parameters of growing conditions. We use the coordinates of the village centers to spatially identify villages. We assume mixed level of inputs and include both rain-fed and irrigated land in the computation of the potential yield. We aggregate across crops within cells using (common) international crop prices from the FAO. See [Adamopoulos and Restuccia \(2021\)](#) for details.

officials and farmers that households are allocated “bundles” of land, and hold similar portfolios of land in terms of land quality, and distance among plots.

### 4.3 Distortions and Productivity

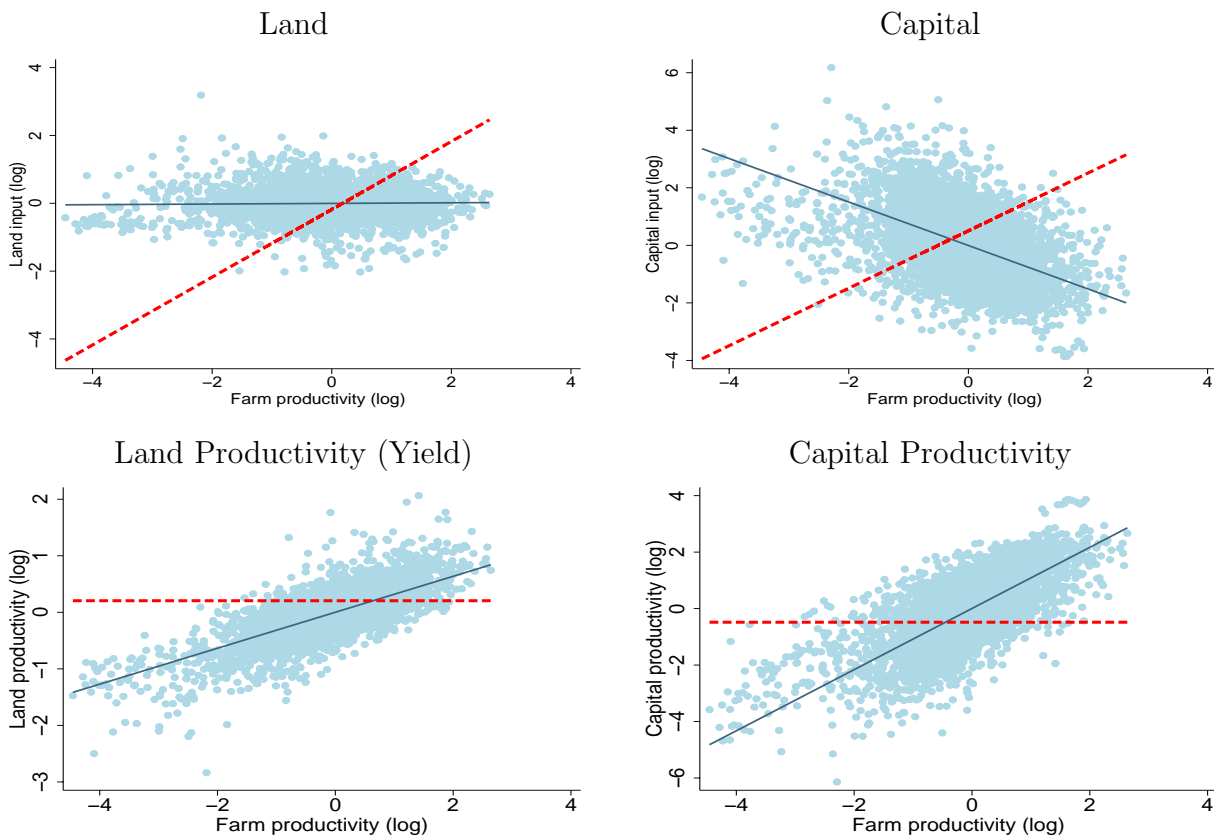
If land and capital were allocated across farms in a decentralized fashion through unhindered factor markets, the resulting allocations should resemble more closely the efficient allocations, with relatively more productive farmers operating at a larger scale with more land and capital. In this case, the relationship between farm input use and TFP would be strongly positive. In addition, we would expect marginal (and average) products of factors to be unrelated with farm TFP since in an efficient allocation these marginal products are equalized.

In the case of China we observe the exact opposite patterns. Figure 1 documents the allocation of land and capital across farms by farm-level productivity (in logs) using our baseline fixed-effect measures of farm productivity and similarly computed permanent fixed-effect measures of inputs and average products. Land use and capital use are not systematically positively correlated with farm productivity. In addition, the average productivity of land and capital inputs are systematically positively correlated with farm productivity across farms. These patterns are not consistent with an efficient allocation of resources across farmers in China (red dotted lines). They are however consistent with the institutional setting in China, most notably, the fairly uniform administrative allocation of land among members of the village. The lack of ownership over the allocated land, and hence inability to use land as collateral, can also partly rationalize the misallocation of capital. Overall, the land market institutions in China prevent the flow of resources to the most productive farmers. If anything, capital use appears to be negatively correlated with farm productivity. This slight negative correlation may be due to other frictions in the capital market, some of which are discussed in [Brandt et al. \(2013\)](#).

The input allocations across farmers in China indicate that there is substantial misallocation. In the context of our decentralized framework, this misallocation is manifested through farm-specific



Figure 1: Factor Allocations by Farm Productivity



Notes: The data on inputs and productivities refer to the permanent household fixed effect measures purged of time and village-level factors. Land and capital are measured relative to total labor days supplied to agriculture by the household. The solid blue line is the estimated relationship between inputs and farm productivity whereas the dashed red line is the efficient allocation associated with each level of farm productivity. Land productivity refers to value added per unit of land and capital productivity refers to value added per unit of capital, both of which are proportional to the marginal products of each factor in our framework.

distortions or “wedges,” measured as deviations of input allocations between actual and efficient. We summarize the distortions faced by a farm in both the land and capital markets by the measure of revenue productivity  $TFPR$  in equation (2). We follow the same procedure as with physical productivity to estimate permanent farm-specific distortions using the panel data on  $TFPR_{i,t}$ . In

particular, using panel methods we first estimate,<sup>11</sup>

$$\log TFPR_{iwt} = \mu_t^{TFPR} + \mu_i^{TFPR} + e_{iwt}^{TFPR}. \quad (6)$$

The interpretation of the regressors is the same as for equation (4). Then we remove village-specific factors from the estimated household fixed effects by regressing them on village dummies ( $\mu_v$ ) to extract the residual,

$$\mu_i^{TFPR} = \mu_v^{TFPR} + \zeta_i^{TFPR}. \quad (7)$$

This procedure provides an estimate of the permanent farm-specific components of distortions  $\hat{\zeta}_i^{TFPR}$ . We refer to farm- $TFPR_i$  as the exponential of the estimated household fixed effect  $\hat{\zeta}_i^{TFPR}$ .

In Figure 2 we plot the farm-specific distortions, as captured by TFPR, against farm TFP (in logs) for our baseline measures. There is a strong positive correlation between farm distortions (TFPR) and farm productivity (TFP), with the correlation equal to 0.91. The more productive farmers face higher farm-specific distortions. This relationship reflects the nature of the land market institutions in China. The administrative egalitarian allocation of land, along with the thin rental land markets, provide little scope for farmers to adjust the operational size of the farm they are “endowed” with. The farmers hurt the most from such an institutional setting are the more productive who would have optimally expanded the most in unfettered markets, operating more land and capital. In the context of our decentralized framework, this is reflected in higher farm-specific distortions on the more productive farmers.

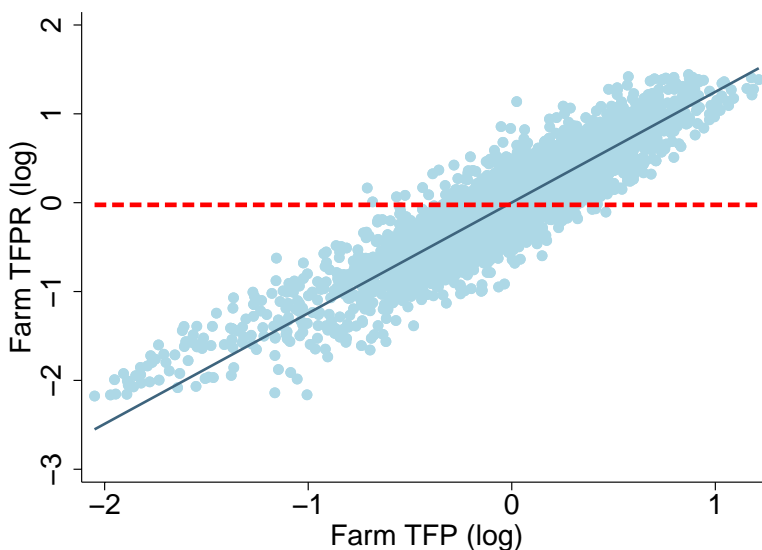
## 4.4 Mismeasurement

Typical analyses of misallocation with cross-sectional data are often criticized for potentially misinterpreting idiosyncratic transitory shocks or measurement error as permanent TFP differences.

---

<sup>11</sup>We use the same approach to estimate permanent measures of land input and then along with the permanent estimates of TFP and TFPR we back out all the other variables of interest, including output, using the structure of the model.

Figure 2: Farm-specific Distortions and Productivity



Notes:  $TFPR_i$  is a summary measure of distortions faced by each household. The standard deviation of  $\log TFPR_i$  is 0.48 and the correlation between  $\log TFP_i$  and  $\log TFPR_i$  is 0.91.

Our analysis is less subject to these critiques because we exploit the panel structure of the data to obtain fixed-effect estimates of farm productivity and farm distortions. We illustrate the value of our approach by applying the method of [Bils et al. \(2017\)](#) for inferring measurement error in TFPR when panel micro-data are available, and comparing statistics from our fixed effects estimates and cross-sectional data.

[Bils et al. \(2017\)](#) utilize changes in output relative to changes in inputs as an independent measure of an input's marginal product that explicitly exploits the time dimension of the panel. This measure of marginal responses of output to inputs is compared to the within-period, average product-based measure of TFPR that is commonly used in the misallocation literature. When the response of output to changes in inputs is larger for higher TFPR farms, average products better reflect true marginal products and measurement error is less of an issue. To what extent do production units, such as farms, plants, or establishments, with higher TFPR display larger output responses to input changes? [Bils et al. \(2017\)](#) show that this question can be addressed by regressing plant growth in measured output on growth in measured inputs and the interaction of the growth in inputs and the

level of measured TFPR.

We implement this approach using our panel data on farms for China. In particular, the econometric model we estimate in our context is,

$$\Delta \log (y_{it}) = \beta_1 \cdot \log (TFPR_{it}) + \beta_2 \cdot \Delta \log (I_{it}) + \beta_3 \cdot interaction_{it} + \mu_v + \mu_t + u_{it}, \quad (8)$$

where  $\Delta \log (y_{it})$  is the change in measured farm log-output;  $\Delta \log (I_{it})$  is the change in measured log-input bundle  $I = \ell^\alpha k^{1-\alpha}$ ;  $interaction_{it} = \Delta \log (I_{it}) \times \log (TFPR_{it})$  is the interaction term between input growth and TFPR, and  $\mu_v$  and  $\mu_t$  are village and time fixed effects. Intuitively, if the coefficient on the interaction term is negative then measured output changes tend to be less responsive to changes in measured inputs for higher TFPR farms. [Bils et al. \(2017\)](#) show that from the regression coefficients in equation (8) we can identify an estimate of the share of the dispersion in TFPR that is due to true variation in distortions,  $\lambda$ , as:  $\hat{\lambda} = 1 + \hat{\beta}_3 / \hat{\beta}_2$ . A  $\lambda$  close to 1 reflects minimal measurement error, implying that measured differences across farms in TFPR reflect largely differences in true distortions. Also, the lower  $\lambda$  is the lower the productivity gains from reallocation relative to those implied by measured TFPR. In the context of manufacturing plants in the United States and India, [Bils et al. \(2017\)](#) find that measurement error was substantial. Estimates for  $\lambda$  for India were stable over time around 0.5, while for the United States  $\lambda$  fell from 0.4 at the beginning of the period to 0.1 at the end.

We estimate equation (8) by OLS, clustering standard errors at the village level. We estimate it for three different measures of TFPR: (a) our baseline farm fixed effects TFPR measure that adjusts for village factors; (b) a farm fixed-effect TFPR measure, but without controlling for village factors (other than land quality); and (c) a raw unadjusted cross-sectional measure of TFPR, controlling only for village-level land quality. Changes in output and inputs are the same across the different specifications; what varies is the measure of TFPR used. The point estimates for  $\lambda$  along with standard errors and the corresponding 95% confidence intervals are reported in the last row of

Table 1. With the raw unadjusted measure of TFPR, we estimate  $\hat{\lambda} = 0.900$ , which implies that measurement error is 10%. For the fixed effect estimate of TFPR from which village factors have not been purged,  $\hat{\lambda} = 0.955$ , implying just under 5% measurement error; whereas for our baseline measure of permanent TFPR the estimated  $\lambda$  is 1.00, implying that there is little scope for the type of measurement error this method can capture with our baseline measure. Our findings suggest that the productivity gains from reallocation, implied by our baseline farm fixed effect estimates of TFP and TFPR, are fairly reflective of the true potential gains from such a reallocation.

Table 1: Mismeasurement in Productivity and Distortions

	Fixed Effect Estimates		Cross-section
	Household Farm	+ Village	average
Farm TFP:			
STD(log)	0.35	0.64	0.72
p90/p10	2.19	4.35	5.59
p75/p25	1.48	2.06	2.32
Farm TFPR:			
STD(log)	0.48	0.81	0.92
p90/p10	3.14	7.17	9.70
p75/p25	1.78	2.71	3.23
CORR (logTFP, logTFPR)	0.91	0.88	0.88
BKR $\hat{\lambda}$			
Standard error	(.026)	(.039)	(.024)
95% confidence interval	[0.95, 1.05]	[0.88, 1.03]	[0.85, 0.95]

Notes: The first two columns refer to statistics for fixed effect estimates of TFP and TFPR from panel regressions. The column “Household Farm” refers to the fixed effects estimates, after removing village specific factors. The column “+ Village” refers to the fixed effect estimates, inclusive of village-level factors but excluding land quality differences across villages. “Cross-section average” refers to statistics computed in each year and then averaged across years. “BKR  $\hat{\lambda}$ ” is the estimate of the fraction of TFPR variation that is not measurement error as in [Bils et al. \(2017\)](#). Each BKR  $\hat{\lambda}$  is based on an OLS regression of changes in log-output on changes in log-inputs, an interaction of changes in log-inputs and TFPR, with village and year dummies, and clustering of standard errors at the village level. Standard errors are in parentheses and 95% confidence intervals in square brackets.

A caveat of the method is that it can capture only certain types measurement error. In particular, it is suitable for assessing the extent of additive measurement error that is orthogonal to true farm

productivity. If, for example, measurement error is multiplicative it cannot be identified through this method. As a result, the lack of measurement error in our baseline farm fixed effects estimates may represent a lower bound on the extent of measurement error in the data.

Estimates reported in Table 1 also reveal a marked reduction in the dispersion of productivity and distortions as we move from the cross-section to the fixed effects estimates. Consistent with a reduction in mismeasurement, the standard deviation of log-distortions is reduced by 10 percent when measured as the household plus village fixed effects compared to the cross-section average. Dispersion in log-distortions is further reduced by 40 percent when measured as the household fixed effect only, but in this case the reduction is mostly due to removing village-level differences. The correlation of farm productivity and distortions is very similar (0.88) between the cross-section and the household plus village fixed effects cases. The correlation is strengthened marginally from 0.88 in the cross-section to 0.91 in the baseline household fixed effect case. The implication is that mismeasurement has virtually no effect on the systematic component of distortions. This is consistent with our description of the land institutions in China and the uniform allocation of village land across households independent of farming ability. The systematic component of distortions (correlated distortions) is a key element in the amplification effect via selection as we discuss in section 5.

## 4.5 Other Evidence

We contrast the land allocations observed in China to those in developed economies, and provide evidence on the effect of land rentals on land allocations within China.

**Comparison to developed countries.** In Figure 1 we contrasted actual allocations of land and capital in China against an efficient benchmark where more productive farms command more inputs. The question arises whether this is a reasonable benchmark. Unfortunately, we do not have access to micro level farm data for a developed country with secure property rights in land to

confirm the tight link between land use and farm productivity implied by the benchmark efficient allocation. Nevertheless, the evidence suggests a much stronger relationship between farm size and productivity in developed economies. We present two pieces of evidence. First, using the U.S. Census of Agriculture data in [Adamopoulos and Restuccia \(2014\)](#) with information on land, capital, value added, and number of farms across 12 farm land-size categories, we calculate the correlation between farm size (operated land input) and several measures of productivity. We find a high positive correlation among these variables: a correlation around 90 percent between farm size and labor productivity, and more than 80 percent between farm size and several alternative measures of farm TFP. This high correlation contrasts sharply with the land allocation in China documented earlier that results in no correlation between land input and farm productivity. The sharp difference between the United States and China is consistent with other recent empirical evidence for developed and developing countries. [Rada and Fuglie \(2019\)](#) summarize the evidence from micro studies suggesting a positive relationship between farm size and productivity in developed economies (such as Australia and the United States) and no systematic relationship in poor and developing countries. Second, the evidence among many developed economies and for the longer historical data in the United States and Canada is that average farm sizes have grown substantially along with high rates of agricultural productivity growth ([Adamopoulos and Restuccia, 2014](#)).

**Land rentals.** If actual allocations are more closely connected to farm productivity in environments with more exposure to land market rental activity, this can be interpreted as complementary evidence in favor of misallocation. While there are no explicit prohibitions of rentals in China, frequent administrative reallocations of land within villages likely lead households to fear loss of their use rights if they do not farm their land themselves. Indeed, land rental markets are very thin in China during our sample period, with rented-in land constituting less than 5 percent of cultivated land and exhibiting limited change over time. More systematic differences emerge between provinces with land rental as high as 12 percent in Zhejiang but virtually zero in Jiangsu. We examine the role of land rental by focusing on those provinces with land rental above the 5 percent

mean. Applying the same fixed-effects methodology to this sample, we find that the correlation of our permanent measures (in logs) of land input and farm productivity across households increases from 0.02 in the full sample to 0.13 for the provinces with more significant land rental activity.<sup>12</sup> This evidence indicates that rentals contribute to a land allocation that is more connected to farm productivity, however the effect is fairly small. In related work, [Chari et al. \(2021\)](#) examine the effects of a land tenancy reform that was announced in China in 2003, and was implemented after the period in our data set. Using a difference-in-differences approach, they find that provinces in China exposed to the reform saw increased land rental activity directed towards more productive farmers, leading to significant output and productivity gains.<sup>13</sup>

## 4.6 Efficiency Gains

We measure the efficiency gains associated with the misallocation of resources in agriculture by conducting a counterfactual exercise that asks, how much larger would aggregate output (and as a result TFP) be in agriculture if farm-specific distortions were eliminated holding constant aggregate resources? The efficiency gains from eliminating misallocation are given by the ratio of the efficient to the (distorted) observed total output minus one,  $Y^e/Y - 1$ . Extreme values of agricultural output, distortions, and non-agricultural income are winsorized at 0.5 percent tails.

Table 2 reports the efficiency gains from factor reallocation using our baseline measure of farm TFP, estimated as the fixed effect component from a panel regression, purged of village effects. Eliminating misallocation across household farms in China (within villages) increases agricultural output, and hence TFP by 24.4 percent (first row, first column). Factor misallocation is observed both across farm households with different levels of productivity (correlated) as well as among households with similar levels of productivity (uncorrelated). However, the bulk of the reallocation

---

<sup>12</sup>We also examine the effect of land rentals on the farm TFP gradient of land input. In the full sample, for all years and all provinces, controlling for a household’s percentage of rented-in land as well as year and province fixed effects, the effect of log-farm TFP on log-land input increases from 0.020 to 0.034.

<sup>13</sup>In a different context, [Chen et al. \(2021\)](#) also find that a land certification reform in Ethiopia had significant positive effects on rental markets and agricultural productivity.



Table 2: Efficiency Gains from Reallocation

	Output (TFP) gain (%)			
	Total	Across $s$ misallocation	Land distortion	Cross-section average
Eliminating misallocation				
across households:				
within villages	24.4	13.9	13.6	54.0
+ across villages	53.2	24.9	–	83.0

Notes: Output (TFP) gain from efficient reallocation in percent. “Total” refers to the fixed effects estimates from the panel regression. “Across  $s$  misallocation” refers to the reallocation gains only eliminating misallocation across farm households with different productivity. “Land distortion” refers to the efficiency gain arising from eliminating the output wedge arising from the land institution. “Cross-section average” refers to reallocation gains when farm TFP and distortions are computed for each cross section and reallocation gains calculated for each year in the panel and then averaged across years.

gains, about 60 percent ( $\log(1.139)/\log(1.244)$ ), is due to reallocating resources across farming households with different TFP, which increases agricultural output by 13.9 percent (first row, second column). If in addition, we allow for input reallocation across villages, i.e., we use household productivity measures without removing the village specific effects but controlling for village land quality, the reallocation gain increases to 53.2 percent (second row, first column). When only capital can be reallocated across villages, in addition to capital and land across existing farms within villages, the efficiency gains are 33.2 percent. These results imply that the gains from reallocation more than double when considering also the possibility of resource reallocation across villages, with about two-thirds accounted for by household mobility across space. This is consistent with the results of [Ngai et al. \(2019\)](#) who show that the limited ability to trade land for cultivation and access to social services tied to local registration constitute substantial barriers to population movements across sectors and space.

We also emphasize that if we consider instead the more conventional cross-sectional measures of farm TFP and distortions typically used in the literature, that do not control for time effects, transitory shocks, measurement error, and other factors, the average reallocation gains are 54 percent within villages and 83 percent when allowing for across village reallocation. This suggests that conventional

cross-sectional estimates may over-estimate reallocation gains.

In general, it is difficult to identify the sources of dispersion in marginal products when measuring input wedges indirectly. The identification issue is particularly acute when the underlying institution creating misallocation affects multiple inputs. In this case, the impact of the institution is better captured by an output wedge that affects dispersion in all marginal products. The land institution we emphasize for China also has effects on capital allocations. As a result, to decompose the impact of the land institution in the efficiency gains we have documented, we separate between an output wedge and a capital to land wedge. Our interpretation is that the output wedge is driven by the land institution whereas the capital to land wedge relates to other factors affecting the capital to land ratio across farmers, including potentially the use of different technologies with varying capital intensities. This is equivalent to interpreting the reported TFPR variation as output wedges, which leaves the capital to land ratio constant, and the residual variation as a capital to land wedge. Under this decomposition and for the estimated household fixed effect measures, the efficiency gain associated with the land institution is 13.6 percent which accounts for almost 60 percent of the efficiency gains (24.4) reported earlier, while the remaining relates to other factors affecting the capital to land ratio across farmers (see Table 2). We focus on the TFPR variation as an output wedge associated with the land institution in our quantitative sectoral analysis.

We also find that there is no substantial change in the magnitude of the misallocation in the rural sector, and thus the gains from reallocation over time in China during our sample period. For example, the standard deviation of log TFPR in the cross-section data is roughly constant over time, if anything slightly increasing, with an average of 0.92 across the years, and as low as 0.87 in 1993, as high as 0.95 in 1997, and 0.91 in 2002. Similarly, the reallocation gains associated with this misallocation appear roughly constant over time. This finding contrasts with the reduction in misallocation found for the manufacturing sector in China, documented in [Hsieh and Klenow \(2009\)](#) over a similar period. Our findings are consistent with the view of costly farm-specific distortions that are tied to land market institutions in China that have not changed much over the period we

study.

## 5 Misallocation and Selection across Sectors

We integrate our framework of agriculture into a two-sector general-equilibrium model of selection (Roy, 1951; Lagakos and Waugh, 2013) to assess: (1) how farm-specific distortions in agriculture alter the occupational choice between agriculture and non-agriculture; and (2) how selection affects measured misallocation and the productivity gains from reallocation. Individuals are endowed with productivities for each sector and make an occupational choice. Production in agriculture is undertaken by heterogeneous farms that differ not only in agricultural ability, but also with respect to farm-specific distortions. A key innovation is that farm distortions, such as those resulting from the land market institutions in China, directly distort occupational choices and amplify the productivity effects of misallocation.

**Environment.** We consider a representative closed village economy and for simplicity we drop village subscripts to focus on individual and sector differences. There are two sectors, agriculture ( $a$ ) and non-agriculture ( $n$ ). The economy is populated by a measure 1 of individuals indexed by  $i$ , who consume agricultural goods subject to a subsistence constraint, and non-agricultural goods. Individuals face an occupational choice, whereby they choose to become a farm operator in agriculture and produce according to production function (1) or a worker for the representative firm in the non-agricultural sector. Individuals are heterogeneous with respect to their abilities in agriculture and non-agriculture, and the farm-specific distortions they face in agriculture. In particular, each individual  $i$  is endowed with a pair of sector-specific abilities  $(s_{ai}, s_{ni})$  and an overall idiosyncratic farm distortion  $\varphi_i$ . The triplet  $(s_{ai}, \varphi_i, s_{ni})$  is drawn from a known population joint trivariate distribution of skills and distortions.

An individual  $i$  choosing to work for the non-agricultural sector earns  $I_{ni} = w_n s_{ni}$ , where  $w_n$  is the

non-agricultural wage per unit of skill. An individual that is a farm-operator in agriculture chooses inputs and output to maximize profits subject to the idiosyncratic distortion. Farm income is the value of output which includes not only the return to farm labor but also the land and capital incomes. We express an individual's income from agriculture as,

$$I_{ai} = w_a \varphi_i s_{ai}, \quad (9)$$

where  $\varphi_i$  captures the overall farm-specific distortion faced by farmer  $i$ . The variable  $w_a$  is the component of the farmer's income that is common to all farmers, and summarizes the effects of common relative prices.

An individual  $i$ 's occupational choice involves choosing the sector that provides the highest income, i.e.,  $\max \{I_{ai}, I_{ni}\}$ . An individual then chooses to operate a farm in agriculture if,

$$w_a \varphi_{ai} s_{ai} \geq w_n s_{ni} \quad (10)$$

The key insight of the occupational choice decision (10) is that, holding relative prices constant, farm-specific distortions directly distort occupational choices. For given common sectoral returns  $(w_a, w_n)$  and individual abilities  $(s_{ai}, s_{ni})$ , a lower  $\varphi$  (higher distortion) reduces the effective return in agriculture. In other words, what matters for the sectoral choice is not idiosyncratic sectoral abilities alone, as would be the case in the standard model, but effective abilities that in agriculture are inclusive of idiosyncratic distortions. To appreciate the weight of this implication, consider what it means through the lens of the type of land market institutions observed in China. Since farm-level distortions are strongly positively correlated with farm productivity, then  $\varphi_{ai}$  is low for productive farmers, reducing their effective ability in agriculture, and potentially affecting their sectoral choice.

The sectoral model has implications for the share of employment in agriculture, the pattern of selection, and sectoral and aggregate productivity. In addition, the model has micro-level implications

summarized by moments of sectoral incomes conditional on sectoral choices, as well as by moments on farm-level productivity and farm-level distortions for those operating in agriculture. We provide more details on the model, calibration, and additional results in Supplemental Material Appendix D, E, and F.

**Calibration.** Our strategy is to calibrate distortions, abilities, and sectoral selection in a Benchmark Economy (BE) to the panel household-level data from China. We assume a trivariate log-normal distribution for the joint distribution of agricultural ability, non-agricultural ability and farm-specific distortions. We then proceed in two steps. First, we infer the population moments on abilities and distortions from observed moments on sectoral incomes and estimated wedges. Second, given the population moments, we calibrate the remaining parameters from the general equilibrium equations of the model to match relevant data targets. To back out the population moments we: *(i)* construct in the model moments on sectoral incomes and farm distortions, conditional on sectoral choices as functions of the population moments; *(ii)* compute the counterparts to the conditional moments in our panel-data from China, using our estimated fixed effect permanent components of distortions, agricultural income, and non-agricultural income for each household; and *(iii)* solve a system of equations for the population moments.

A key empirical moment in our analysis is the covariance of log sectoral incomes. A typical limitation of empirical models of selection is that income is observed only for the chosen occupation. An advantage of our setting is that for the vast majority of households (around 96 percent), income is observed in both agricultural and non-agricultural activities; and many households switch from agriculture to non-agriculture under a variety of definitions of switchers in our panel data. The moment we use as our baseline is the contemporaneous covariance of log sectoral incomes across households, which implies a correlation of 0.034 in our micro-data.

The calibration of the remaining parameters ensures that at the aggregate level the share of employment in agriculture in China is 46 percent. The model reproduces well the macro and micro

statistics for China, with the log-normal assumption for the distributions of distortions and abilities, that affords substantial tractability, providing a good fit of the empirical distributions. In particular, whereas the productivity gain from eliminating distortions across existing farmers with different levels of productivity (correlated distortions) in the data is 13.9 percent (Table 2), it is 10 percent in the model with parametric distributions of distortions and abilities.

**The effect of farm-specific distortions.** We conduct a counterfactual experiment to assess the effects of the land market institution in China. We do so by eliminating distortions that are correlated with agricultural ability. We remove the systematic component of distortions simply by regressing distortions  $\log(\varphi_i)$  on log agricultural ability for all individuals in the benchmark economy and retaining the residual. We then compute the equilibrium with only residual distortions. Note that in this experiment there is still misallocation associated with the dispersion of distortions unrelated to farm productivity. Table 3 reports the results. Agricultural labor productivity increases by almost 3-fold. This increase arises from an increase in agricultural TFP of 67 percent and the associated reduction in the labor share in agriculture, from 46 percent to 16 percent. Note that the effect on agricultural TFP of eliminating correlated distortions, holding constant farmers in the benchmark economy, is 10 percent in our calibrated model. As a result correlated distortions contribute substantially to agricultural productivity via distorted selection in occupational choices. We note that the overall labor productivity gains from removing the land market friction are modest in absolute terms compared to the literature on agricultural labor productivity gaps (Restuccia et al., 2008).

For comparison purposes, we conduct an alternative counterfactual, whereby we eliminate all farm-specific distortions, i.e., when  $\varphi_i = 1$  for all  $i$ . This counterfactual implies an increase in agricultural labor productivity of 3.4-fold, an increase in agricultural TFP of 80 percent, and a reduction in the share of employment in agriculture to 14 percent (see Supplemental Material Appendix F). Comparing to the findings in Table 3, these results illustrate that the bulk of the selection effect

arise from correlated distortions associated with the land institution we emphasize.

Table 3: The Effects of Correlated Distortions

Statistic	Benchmark Economy (BE)	No Correlated Distortions
Aggregate Statistics		
Real Agricultural Productivity ( $Y_a/N_a$ )	1.00	2.96
Share of Employment in Agriculture ( $N_a$ ) (%)	0.46	0.16
TFP in Agriculture ( $TFP_a$ )	1.00	1.67
TFP in Agriculture, constant BE farms	1.00	1.10
Real Non-Agricultural Productivity ( $Y_n/N_n$ )	1.00	0.78
Average Ability in Agriculture ( $Z_a/N_a$ )	1.00	2.34
Average Ability in Non-Agriculture ( $Z_n/N_n$ )	1.00	0.78
Real GDP per Worker ( $Y/N$ )	1.00	1.18
Conditional Micro-level Statistics		
STD of log-farm TFP	0.56	0.39
STD of log-farm TFPR	0.48	0.14
CORR of log-(farm TFP, farm TFPR)	0.97	0.44
CORR of log-(agr. ability, non-agr. ability)	0.15	0.49
CORR of log-(agr. income, non-agr. income)	0.03	0.40

Notes: “No Correlated Distortions” eliminates the component of farm-level distortions  $\varphi_i$  associated with agricultural ability. Aggregate variables, except for the share of employment in agriculture, reported relative to the same statistic in the Benchmark Economy (BE). Micro-level statistics reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy. Real agricultural productivity is agricultural output divided by employment in agriculture.  $TFP_a$  is agricultural output divided by the bundle of inputs from the implied aggregate production function.

Eliminating correlated farm-specific distortions increases agricultural labor productivity directly by reducing misallocation across farms, but it also induces higher ability farmers to work in agriculture since they face no systematic restrictions on consolidating operational land. These two effects constitute the increase in agricultural TFP. At the same time, the associated increases in productivity via reduced misallocation and improved selection reduce the share of employment in agriculture, which further raises agricultural labor productivity by increasing the amount of land and capital available per farmer. In terms of selection, whereas in the benchmark economy with distortions employment in agriculture is roughly uniformly distributed across all agricultural ability types, removing distortions improves selection into agriculture of mostly high ability farmers. In

terms of TFP in agriculture, out of the overall increase by a factor of 1.67-fold, more than four-fifths ( $1 - \log(1.10)/\log(1.67)$ ) represent the effect of selection (an additional 50 percent). In other words, the amplification effect of selection on agricultural TFP quadruples the gains from reduced misallocation when keeping the set of farmers fixed (10 percent).

Average ability in non-agriculture, and as a result real labor productivity in non-agriculture, falls to 78 percent of their benchmark economy values, due to more workers in non-agriculture, not all of which are as productive in that sector. The nominal non-agricultural to agricultural productivity ratio falls from 3.95 to 3.13, when the share of labor in agriculture drops (from 46 to 16 percent). This finding is quantitatively consistent with the evidence for a wide range of countries, that lower employment shares in agriculture are associated with lower nominal productivity ratios, see for example Figure 2 and Table 7 in [Storesletten et al. \(2019\)](#). Despite the significant impact on agricultural productivity, real GDP per worker increases about 18 percent. The reason for the dampened effect on aggregate output is that there is a large shift of labor to the sector that sees a drop in its productivity relative to the benchmark. The dispersion of TFP in agriculture is lower, albeit with a higher mean.

Our main result of a substantial amplification effect on agricultural TFP from improved selection is robust to variations in model specification and the inferred correlation of abilities across sectors. For instance, if we impose a positive correlation of abilities and recalibrate the model, eliminating correlated distortions generates a stronger amplification effect on agricultural productivity, due to a stronger effect of selection. This however implies a counter-factually strong correlation of incomes across sectors (see Supplemental Material Appendix F). Similarly, introducing idiosyncratic preference or mobility barriers that are independent of ability results in changes in the values of the ability correlation and other model parameters required to match the targeted moments, but in this case the model produces a similar amplification effect of eliminating correlated distortions. While admittedly our analysis does not provide a strong identification of the “true” correlation of abilities, it does provide discipline on the magnitude of the amplification effect, which is substantial,



and robust to different estimates of the correlation of abilities across sectors.

## 6 Conclusions

Using a simple quantitative framework and micro panel data, we presented evidence that capital and land are severely misallocated across farmers within villages in China. Given the institutional framework, we argued that this factor misallocation reflects primarily restrictions in the land market, which also dampens access to credit for farmers. The administrative allocation of land-use rights on an egalitarian basis manifests itself as a larger idiosyncratic distortion on the more productive farmers. Over time, the resulting pattern of misallocation shows no systematic tendency to improve, consistent with the persistent nature of the institutional restrictions in the Chinese economy.

Using the idiosyncratic distortions we measure across farmers in China, we develop and estimate a two-sector general-equilibrium model of occupational selection. The panel data provide information on income in agriculture and non-agriculture, which we use to discipline the effect of selection on productivity. We find that measured distortions substantially affect the observed distribution of farm TFP in the Chinese data, and that eliminating the correlation of these distortions with farmer's ability improves aggregate agricultural productivity via reduced misallocation and improved selection of more able farmers into agriculture. This effect substantially contributes to structural change and growth.

Our analysis implies that implementing a system of secure property rights to facilitate a decentralized allocation of land would generate large productivity gains. To the extent that village officials do not observe farmer ability or do not make land allocation decisions based on ability, any administrative (re)allocation of land would be unable to channel land to farmers that value it the most or can make the most out of it. Developing a market allocation mechanism by extending fully transferable use rights over land to farmers would not only allow farmers to increase their operational scales through land consolidations, but also induce the best farmers to stay in agriculture, while

releasing labor to non-agriculture. The productivity and farm size increases due to a better allocation of factors of production among farmers and an improved selection of farmers in agriculture can arguably also induce changes in farm operations by incentivizing farmers to use modern inputs and better technologies. We leave this important extension of our framework for future research.

## References

- Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–97.
- Adamopoulos, T. and Restuccia, D. (2020). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics*, 12(3):1–39.
- Adamopoulos, T. and Restuccia, D. (2021). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. *Review of Economic Studies*.
- Benjamin, D. and Brandt, L. (2002). Property rights, labour markets, and efficiency in a transition economy: the case of rural china. *Canadian Journal of Economics*, pages 689–716.
- Benjamin, D., Brandt, L., and Giles, J. (2005). The evolution of income inequality in rural china. *Economic Development and Cultural Change*, 53(4):769–824.
- Bils, M., Klenow, P. J., and Ruane, C. (2017). Misallocation or mismeasurement? *Stanford Center for International Development*, (Working Paper No. 599).
- Brandt, L., Huang, J., Li, G., and Rozelle, S. (2002). Land rights in rural china: Facts, fictions and issues. *The China Journal*, (47):67–97.
- Brandt, L., Tombe, T., and Zhu, X. (2013). Factor market distortions across time, space and sectors in china. *Review of Economic Dynamics*, 16(1):39–58.
- Cao, K. H. and Birchenall, J. A. (2013). Agricultural productivity, structural change, and economic growth in post-reform china. *Journal of Development Economics*, 104:165–180.
- Chari, A., Liu, E. M., Wang, S.-Y., and Wang, Y. (2021). Property rights, land misallocation, and agricultural efficiency in china. *The Review of Economic Studies*, 88(4):1831–1862.
- Chen, C. (2017). Untitled land, occupational choice, and agricultural productivity. *American Economic Journal: Macroeconomics*, 9(4):91–121.
- Chen, C., Restuccia, D., and Santaeulàlia-Llopis, R. (2021). The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics*.

- Chow, G. C. (1993). Capital formation and economic growth in china. *The Quarterly Journal of Economics*, 108(3):809–842.
- De Janvry, A., Emerick, K., Gonzalez-Navarro, M., and Sadoulet, E. (2015). Delinking land rights from land use: Certification and migration in mexico. *American Economic Review*, 105(10):3125–49.
- Fan, S. and Zhang, X. (2002). Production and productivity growth in chinese agriculture: New national and regional measures. *Economic Development and Cultural Change*, 50(4):819–838.
- Gollin, D., Parente, S., and Rogerson, R. (2002). The Role of Agriculture in Development. *American Economic Review*, 92(2):160–164.
- Gottlieb, C. and Grobovšek, J. (2019). Communal land and agricultural productivity. *Journal of Development Economics*, 138:135–152.
- Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.
- Jin, S., Huang, J., and Rozelle, S. (2010). Agricultural productivity in china. *The shifting patterns of agricultural production and productivity worldwide*, pages 229–278.
- Kung, J. K.-s. and Liu, S. (1997). Farmers’ preferences regarding ownership and land tenure in post-mao china: unexpected evidence from eight counties. *The China Journal*, (38):33–63.
- Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *The American Economic Review*, 103(2):948–980.
- Lin, J. Y. (1992). Rural reforms and agricultural growth in china. *The American Economic Review*, pages 34–51.
- Lucas Jr, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523.
- McMillan, J., Whalley, J., and Zhu, L. (1989). The impact of china’s economic reforms on agricultural productivity growth. *The Journal of Political Economy*, pages 781–807.
- Ngai, L. R., Pissarides, C. A., and Wang, J. (2019). Chinas Mobility Barriers and Employment Allocations. *Journal of the European Economic Association*, 17(5):1617–1653.
- Rada, N. E. and Fuglie, K. O. (2019). New perspectives on farm size and productivity. *Food Policy*, 84:147–152.
- Restuccia, D. and Rogerson, R. (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Plants. *Review of Economic Dynamics*, 11(4):707–720.
- Restuccia, D. and Rogerson, R. (2013). Misallocation and Productivity. *Review of Economic Dynamics*, 16(1):1–10.

- Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–74.
- Restuccia, D. and Santaaulalia-Llopis, R. (2015). Land misallocation and productivity. Unpublished manuscript, University of Toronto.
- Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2):234–250.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.
- Song, Z., Storesletten, K., and Zilibotti, F. (2011). Growing like china. *The American Economic Review*, 101(1):196–233.
- Storesletten, K., Zhao, B., and Zilibotti, F. (2019). Business cycle during structural change: Arthur lewis’ theory from a neoclassical perspective. Technical report, National Bureau of Economic Research.
- Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, 109(5):1843–72.
- Yang, D. T. (1997). China’s land arrangements and rural labor mobility. *China Economic Review*, 8(2):101–115.
- Zhu, X. (2012). Understanding chinas growth: Past, present, and future (digest summary). *Journal of Economic Perspectives*, 26(4):103–124.

# Supplement to “Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China”

Tasso Adamopoulos  
York University

Loren Brandt  
University of Toronto

Jessica Leight  
IFPRI

Diego Restuccia  
University of Toronto and NBER

September 2021

## Appendix

### A Measures of Output and Inputs from Panel Data

**Value added agricultural output.** We utilize the detailed information on farm output by crop in physical terms to construct estimates of “real” gross farm output. Output of each crop is valued at a common set of prices, which are constructed as sample-wide averages (unit values) over 1993-2002 for each crop. Unit values are computed using information on market sales, and are exclusive of any “quota” sales at planned (below market) prices. In these calculations, a household’s own consumption is implicitly valued at market prices. Intermediate inputs such as fertilizers and pesticides are treated in an analogous way. We subtract expenditures on intermediate inputs from gross output to obtain our estimate of net income or value added for the cropping sector. In what follows, we use this measure of real value added at the farm-level when we refer to farm output.

**Land, capital, and labor.** The measure of land that we use in our analysis is cultivated land by the household, which corresponds to the concept of operated rather than “owned” farm size. The survey provides household-level information beginning in 1986 on the value at original purchase prices of farm machinery and equipment, larger hand tools, and draft animals used in agriculture. Assuming that accumulation began in 1978, the year the reforms of the agricultural system began, we utilize the perpetual inventory method to calculate the value of farm machinery in constant Renminbi (RMB). The survey does not capture household ownership of smaller farm tools and

implements, and so for just over a third of household-years, the estimated value of their capital stock is zero. To deal with these cases, we impute for all farm households a value equal to the amount of land operated by the household multiplied by ten percent of the median capital to land ratio by village-year. Robustness tests show that our results are not crucially sensitive to the adjustment factor we use. For the labor input, we have the total labor days supplied on agricultural activities by all members of the household and by hired labor.

## B Efficient Allocation in Basic Framework

The planner chooses how to allocate land and capital across farmers in the rural village economy to maximize agricultural output subject to resource constraints. Specifically, the problem of the planner is:

$$\max_{\{k_i, \ell_i\}_{i=1}^M} \sum_{i=1}^M y_i,$$

subject to

$$y_i = (A_a s_i)^{1-\gamma} (\ell_i^\alpha k_i^{1-\alpha})^\gamma, \quad i = 1, 2, \dots, M;$$

and the resource constraints,

$$\sum_{i=1}^M \ell_i = L; \quad \sum_{i=1}^M k_i = K. \tag{B.1}$$

Using the first-order conditions of this problem along with the rural village resource constraints in (B.1), the efficient allocation involves allocating total land and capital across farmers according to relative productivity,

$$\ell_i^e = \frac{s_i}{\sum_{j=1}^M s_j} L, \tag{B.2}$$

$$k_i^e = \frac{s_i}{\sum_{j=1}^M s_j} K, \tag{B.3}$$

where the superscript  $e$  denotes the efficient allocation. Equations (B.2) and (B.3) indicate that in the efficient allocation, more productive farmers are allocated more land  $\ell$  and capital  $k$ .

Using the definition of agricultural output  $Y = \sum_{i=1}^M y_i$  along with individual technologies and input allocations as derived above, we obtain a rural village-wide production function,

$$Y^e = A^e M^{1-\gamma} [L^\alpha K^{1-\alpha}]^\gamma,$$

where  $Y^e$  is agricultural output under the efficient allocation,  $A^e$  is agricultural TFP, given by

$A^e = (A_a \bar{S})^{1-\gamma}$ , where  $\bar{S} = \left(\sum_{i=1}^M s_i\right) / M$  is average farm productivity.

## C Equilibrium and Identification of Distortions

In the basic framework, we denote by  $\tau_i^\ell$  and  $\tau_i^k$  the land and capital input taxes, and by  $\tau_i^y$  the output tax faced by farm  $i$ . Tax revenues are equally distributed lump-sum across all households. We solve the farmer problem subject to all the farm-specific taxes and then show the identification issue that arises.

Given distortions, the profit maximization problem facing farm  $i$  is,

$$\max_{\ell_i, k_i} \{ \pi_i = (1 - \tau_i^y) y_i - (1 + \tau_i^k) r k_i - (1 + \tau_i^\ell) q \ell_i \},$$

where  $q$  and  $r$  are the rental prices of land and capital. In equilibrium, the land and capital markets for the rural village economy must clear as in equation (B.1).

We use this framework to identify the farm-specific distortions from the observed land and capital allocations across farmers. In our abstraction these distortions are induced by “taxes” but in practice they arise in part from China’s land market institutions. In particular, the first-order conditions with respect to land and capital for farm  $i$  imply:

$$\frac{MRPL_i}{\alpha\gamma} = \frac{y_i}{\ell_i} = \frac{q(1 + \tau_i^\ell)}{\alpha\gamma(1 - \tau_i^y)} \propto \frac{(1 + \tau_i^\ell)}{(1 - \tau_i^y)}, \quad (\text{C.4})$$

$$\frac{MRPK_i}{(1 - \alpha)\gamma} = \frac{y_i}{k_i} = \frac{r(1 + \tau_i^k)}{(1 - \alpha)\gamma(1 - \tau_i^y)} \propto \frac{(1 + \tau_i^k)}{(1 - \tau_i^y)}, \quad (\text{C.5})$$

where  $MRPL$  and  $MRPK$  are the marginal revenue products of land and capital, respectively. Given that we normalize the price of agricultural goods to one,  $MRPL$  and  $MRPK$  are also the marginal products of the respective factors. Equations (C.4) and (C.5) show that in the presence of farm-specific distortions, average products and marginal products of land and capital are not equalized across farms, but rather vary in proportion to the idiosyncratic distortion faced by each factor relative to the output distortion.

Equations (C.4)-(C.5) imply two things. First, only two of the three taxes can be separately identified. Second, farm-specific distortions can be identified up to a scalar from the average product of each factor. The scalar for the land input common to all farms is  $\frac{q}{\alpha\gamma}$ , while the scalar for the capital input is  $\frac{r}{(1-\alpha)\gamma}$ .

We construct the following summary measure of distortions faced by farm  $i$ ,

$$TFPR_i = \frac{y_i}{\ell_i^\alpha k_i^{1-\alpha}} = \widetilde{TFPR} \frac{(1 + \tau_i^\ell)^\alpha (1 + \tau_i^k)^{1-\alpha}}{(1 - \tau_i^y)}, \quad (\text{C.6})$$

where  $\widetilde{TFPR} \equiv \left(\frac{q}{\alpha\gamma}\right)^\alpha \left(\frac{r}{(1-\alpha)\gamma}\right)^{1-\alpha}$  is the common component across all farms.

Using the fact that total output is  $Y = \sum_{i=1}^M y_i$ , we can derive the rural village-wide production function,

$$Y = TFP \cdot M^{1-\gamma} [L^\alpha K^{1-\alpha}]^\gamma, \quad (\text{C.7})$$

where  $(L, K)$  are total land and capital, and  $TFP$  is rural village-wide TFP,

$$TFP = \left[ \frac{A_a \sum_{i=1}^M s_i \left(\frac{\overline{TFPR}}{TFPR_i}\right)^{\frac{\gamma}{1-\gamma}}}{M} \right]^{1-\gamma}, \quad (\text{C.8})$$

with average revenue productivity  $\overline{TFPR}$  given by

$$\overline{TFPR} = \frac{\widetilde{TFPR}}{\left[ \sum_{i=1}^M \frac{y_i}{Y} \frac{(1-\tau_i^y)}{(1+\tau_i^\ell)} \right]^\alpha \left[ \sum_{i=1}^M \frac{y_i}{Y} \frac{(1-\tau_i^y)}{(1+\tau_i^k)} \right]^{1-\alpha}}.$$

Equation (C.8) makes clear that with no dispersion in  $TFPR_i$  across farm households, the equilibrium allocations, aggregate output and TFP coincide with the corresponding efficient statistics.

## D Model of Misallocation and Selection across Sectors

This section provides a more detailed description of the model in Section 5 of the paper.

### D.1 Environment

We consider a representative closed village economy and for simplicity we drop village subscripts to focus on individual and sector differences. At each date there are two goods produced, agricultural ( $a$ ) and non-agricultural ( $n$ ). The non-agricultural good is the numeraire and we denote the relative price of the agricultural good by  $p_a$ . The economy is populated by a measure 1 of individuals indexed by  $i$ .



**Preferences** Each individual  $i$  has preferences over the consumption of the two goods given by,

$$U_i = \omega \log(c_{ai} - \bar{a}) + (1 - \omega) \log(c_{ni}),$$

where  $c_a$  and  $c_n$  denote the consumption of the agricultural and non-agricultural good,  $\bar{a}$  is a minimum subsistence requirement for the agricultural good, and  $\omega$  is the preference weight on agricultural goods. The subsistence constraint implies that when income is low a disproportionate amount is allocated to the agricultural good. Individual  $i$  faces the following budget constraint,

$$p_a c_{ai} + c_{ni} = I_i + T,$$

where  $I_i$  is the individual's income, and  $T$  the transfer to be specified below.

Working in the agricultural sector involves operating a farm and is subject to idiosyncratic distortions, captured by  $\varphi_i$ , which we define more fully below. Income from working in the non-agricultural sector is subject to a tax  $\eta$ , common to all individuals.  $\eta$  operates as a barrier to labor mobility from agriculture to non-agriculture and is meant to capture the factors that restrict access to off-farm opportunities for all farmers. Quantitatively, this parameter allows us to fit the ratio of agricultural to non-agricultural labor productivity, but otherwise plays no significant role in our analysis.

**Individual abilities and distortions.** Individuals are heterogeneous with respect to their abilities in agriculture and non-agriculture, and the farm-specific distortions they face in agriculture. In particular, each individual  $i$  is endowed with a pair of sector-specific abilities  $(s_{ai}, s_{ni})$  and an idiosyncratic farm distortion  $\varphi_i$ . The triplet  $(s_{ai}, \varphi_i, s_{ni})$  is drawn from a known population joint trivariate distribution of skills and distortions with density  $f(s_{ai}, \varphi_i, s_{ni})$  and cdf  $F(s_{ai}, \varphi_i, s_{ni})$ . We allow for the possibility that skills are correlated across sectors, and that agricultural skills (but not non-agricultural skills) are correlated with farm-specific distortions. In particular, we assume a trivariate log-normal distribution for  $(s_{ai}, \varphi_i, s_{ni})$  with mean  $(\mu_a, \mu_\varphi, \mu_n)$  and variance,

$$\Sigma = \begin{pmatrix} \sigma_a^2 & \sigma_{a\varphi} & \sigma_{an} \\ \sigma_{a\varphi} & \sigma_\varphi^2 & 0 \\ \sigma_{an} & 0 & \sigma_n^2 \end{pmatrix}.$$

We denote the correlation coefficient for abilities across sectors by  $\rho_{an} = \sigma_{an}/(\sigma_n\sigma_a)$ , and the correlation coefficient between agricultural ability and farm-specific distortions by  $\rho_{\varphi a} = \sigma_{\varphi a}/(\sigma_\varphi\sigma_a)$ . Individuals face two choices: (a) a consumption choice, the allocation of total income (including transfers) between consumption of agricultural and non-agricultural goods; and (b) an occupational

choice, whether to work in the non-agricultural sector or the agricultural sector. We denote the income an individual  $i$  would earn in agriculture as  $I_{ai}$  and that in non-agriculture as  $I_{ni}$ . The individual chooses the sector with the highest income. We denote by  $H_n$  and  $H_a$ , the sets of  $(s_{ai}, \varphi_i, s_{ni})$  values for which agents choose each sector  $H_n = \{(s_{ai}, \varphi_i, s_{ni}) : I_{ai} < I_{ni}\}$ , and  $H_a = \{(s_{ai}, \varphi_i, s_{ni}) : I_{ai} \geq I_{ni}\}$ .

**Consumption allocation.** To determine the allocation of income between agricultural and non-agricultural goods individuals maximize utility subject to their budget constraint, given their income  $I_i + T$ , and the relative price of the agricultural good  $p_a$ . The first order conditions to individual  $i$ 's utility maximization problem imply the following consumption choices,

$$c_{ai} = \bar{a} + \frac{\omega}{p_a} (I_i + T - p_a \bar{a}), \quad c_{ni} = (1 - \omega) (I_i + T - p_a \bar{a}).$$

**Production in non-agriculture.** The non-agricultural good is produced by a stand-in firm with access to a constant returns to scale technology that requires only effective labor as an input,

$$Y_n = A_n Z_n,$$

where  $Y_n$  is the total amount of non-agricultural output produced,  $A_n$  is non-agricultural productivity (TFP), and  $Z_n$  is the total amount of labor input measured in efficiency units, i.e., accounting for the ability of workers  $Z_n = \int_{i \in H_n} s_{ni} di$ . The total number of workers employed in non-agriculture is,

$$N_n = \int_{i \in H_n} di.$$

The representative firm in the non-agricultural sector chooses how many efficiency units of labor to hire in order to maximize profits. The first order condition from the representative firm's problem in non-agriculture implies  $w_n = A_n$ .

**Production in agriculture.** The production unit in the agricultural sector is a farm. A farm is a technology that requires the inputs of a farm operator with ability  $s_{ai}$  as well as land (which also defines the size of the farm) and capital under the farmer's control. The farm technology exhibits decreasing returns to scale and takes the form,

$$y_{ai} = (A_a s_{ai})^{1-\gamma} (\ell_i^\alpha k_i^{1-\alpha})^\gamma, \tag{D.9}$$

where  $y_a$  is real farm output,  $\ell$  is the land input, and  $k$  is the capital input.  $A_a$  is an agriculture-specific TFP parameter, common across all farms. An individual that chooses to operate a farm faces an overall farm-specific tax on output  $\tau_i$ . Note that in the data  $(1 - \tau_i)$  is constructed as a summary of the distortions faced by each farm, as identified in Section 4 of the paper. Tax revenues are redistributed equally to the  $N$  workers independently of occupation, and equal to  $T$  per individual.

The profit maximization problem for farm  $i$  is given by,

$$\max_{\ell_i, k_i} \{ \pi_i = p_a (1 - \tau_i) y_{ai} - r k_i - q \ell_i \}, \quad (\text{D.10})$$

where  $(q, r)$  are the rental prices of land and capital. The first-order conditions to farm operator  $i$ 's problem imply that farm size, demand for capital input, output supply, and profits depend not only on productivity but also on farm-specific distortions,

$$\ell_i = A_a (\gamma p_a)^{\frac{1}{1-\gamma}} \left( \frac{1-\alpha}{r} \right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left( \frac{\alpha}{q} \right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}} (1 - \tau_i)^{\frac{1}{1-\gamma}} s_{ai}, \quad (\text{D.11})$$

$$k_i = A_a (\gamma p_a)^{\frac{1}{1-\gamma}} \left( \frac{1-\alpha}{r} \right)^{\frac{1-\alpha\gamma}{1-\gamma}} \left( \frac{\alpha}{q} \right)^{\frac{\alpha\gamma}{1-\gamma}} (1 - \tau_i)^{\frac{1}{1-\gamma}} s_{ai}, \quad (\text{D.12})$$

$$y_{ai} = A_a (\gamma p_a)^{\frac{\gamma}{1-\gamma}} \left( \frac{1-\alpha}{r} \right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left( \frac{\alpha}{q} \right)^{\frac{\alpha\gamma}{1-\gamma}} (1 - \tau_i)^{\frac{\gamma}{1-\gamma}} s_{ai}, \quad (\text{D.13})$$

$$\pi_i = A_a (1 - \gamma) p_a^{\frac{1}{1-\gamma}} \gamma^{\frac{\gamma}{1-\gamma}} \left( \frac{1-\alpha}{r} \right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left( \frac{\alpha}{q} \right)^{\frac{\alpha\gamma}{1-\gamma}} (1 - \tau_i)^{\frac{1}{1-\gamma}} s_{ai}. \quad (\text{D.14})$$

The income of a farmer is the (after-tax) value of their output  $I_{ai} = p_a (1 - \tau_i) y_{ai}$ . As a result farmer income includes not only the return to the farmer's labor input  $\pi$  but also the land and capital incomes. We can re-write an individual's income from agriculture as,

$$I_{ai} = w_a \varphi_i s_{ai}, \quad (\text{D.15})$$

where  $\varphi_i \equiv (1 - \tau_i)^{\frac{1}{1-\gamma}}$  captures the overall farm-specific distortion faced by farmer  $i$ , and  $w_a$  is the component of the farmer's income that is common to all farmers,

$$w_a \equiv p_a^{\frac{1}{1-\gamma}} A_a \gamma^{\frac{\gamma}{1-\gamma}} \left( \frac{1-\alpha}{r} \right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left( \frac{\alpha}{q} \right)^{\frac{\alpha\gamma}{1-\gamma}}. \quad (\text{D.16})$$

Note that  $w_a$  summarizes the effects of relative prices as it is a function of the endogenous relative price of agriculture  $p_a$ , the rental price of land  $q$ , and the rental price of capital  $r$ .

Similarly, we can re-write land input demand, capital input demand, output supply and profits for farmer  $i$  in terms of their agricultural ability and farm-specific distortions,

$$\ell_i = \bar{\ell}\varphi_i s_{ai}; \quad k_i = \bar{k}\varphi_i s_{ai}; \quad y_{ai} = \bar{y}_a \varphi_i^\gamma s_{ai}; \quad \pi_i = \bar{\pi}\varphi_i s_{ai},$$

where the terms in bars denote the components that are common across all farmers:  $\bar{\ell} = w_a \alpha \gamma / q$ ;  $\bar{k} = (1 - \alpha) \gamma w_a / r$ ;  $\bar{y}_a = w_a / p_a$ ;  $\bar{\pi} = (1 - \gamma) w_a$ .

**Occupational choice.** Individuals choose to operate a farm in the agricultural sector or be workers in the non agricultural sector. If individual  $i$  chooses to operate a farm, their income is given by (D.15), while if they choose to work in non-agriculture their income is,

$$I_{ni} = (1 - \eta) w_n s_{ni}.$$

We note that incomes are net of the transfer  $T$ , which is common to all individuals and hence does not affect occupational choices. Individual  $i$  chooses the sector that provides the highest possible income, given the individual's triplet  $(s_{ai}, \varphi_i, s_{ni})$ . Individual  $i$  chooses agriculture, i.e.  $i \in H_a$ , if  $I_{ai} \geq I_{ni}$  and non-agriculture otherwise. As a result individual  $i$ 's income is given by,

$$I_i = \max \{I_{ai}, I_{ni}\}.$$

Note that income in agriculture depends not only on the individual's agricultural ability  $s_{ai}$  but also on the individual's farm distortion  $\varphi_i$ . We can define an individual's effective ability as the product of the two,  $\hat{s}_{ai} \equiv s_{ai}\varphi_i$ . An individual then chooses to operate a farm if  $w_a \hat{s}_{ai} \geq (1 - \eta) w_n s_{ni}$ . We note that, holding relative prices constant, farm-specific taxes directly distort the occupational choices. For given common sectoral returns  $(w_a, w_n)$ , barrier  $\eta$ , and individual abilities  $(s_{ai}, s_{ni})$ , a lower  $\varphi$  (higher tax) reduces the effective return in agriculture. We denote the occupational choice of an individual  $i$  facing triplet  $(s_{ai}, \varphi_i, s_{ni})$  by an indicator function  $o(s_{ai}, \varphi_i, s_{ni})$  that takes the value of 1 if  $I_{ai} \geq I_{ni}$  and 0 otherwise.

**Definition of equilibrium.** A competitive equilibrium is a set of prices  $\{p_a, r, q\}$ , an allocation for each farm operator  $\{\ell_i, k_i, y_{ai}\}$ , and allocation for the non-agricultural firm  $\{Y_n, N_n\}$ , an occupational choice  $\{o(s_{ai}, \varphi_i, s_{ni})\}$  for each individual  $i$  faced with triplet  $(s_{ai}, \varphi_i, s_{ni})$ , a per capita transfer  $T$ , a consumption allocation  $\{c_{ai}, c_{ni}\}$  for each individual  $i$ , such that: (a) the consumption allocation

for each individual  $\{c_{ai}, c_{ni}\}$  maximizes their utility subject to their budget constraint, given prices, abilities, distortions, and transfers; (b) the production allocation for each farm operator  $\{\ell_i, k_i, y_{ai}\}$  maximizes profits given prices, agricultural ability, and distortions; (c) the non-agricultural production allocation  $\{Y_n, N_n\}$  maximizes the profits of the non-agricultural representative firm, given prices; (d) occupational choices  $\{o(s_{ai}, \varphi_i, s_{ni})\}$  maximize income for each individual given relative prices, abilities, distortions, transfers, and barrier to labor mobility; (e) the markets for labor, capital, land, agricultural goods, and non-agricultural goods clear; and (f) the government budget constraint from the tax-transfer scheme is satisfied.

## D.2 Analysis

We exploit the properties of the multivariate log-normal distribution over  $(s_{ai}, \varphi_i, s_{ni})$  in order to provide analytical results. We show that when  $(s_{ai}, \varphi_i, s_{ni})$  are drawn from a multi-variate log-normal distribution the share of employment in agriculture is given by,

$$N_a = \Phi(b), \quad (\text{D.17})$$

where  $\Phi(\cdot)$  is the standard normal cdf and,

$$b \equiv \frac{b_a - b_n}{\sigma}, \quad b_a \equiv \log(w_a) + \mu_\varphi + \mu_a, \quad b_n \equiv \log(w_n) + \log(1 - \eta) + \mu_n, \quad (\text{D.18})$$

where  $\sigma$  is the variance of relative effective abilities between non-agriculture and agriculture.

Define deviations of log draws from means,

$$u_{ai} = \log(s_{ai}) - \mu_a, \quad u_{ni} = \log(s_{ni}) - \mu_n, \quad u_{\varphi i} = \log(\varphi_i) - \mu_\varphi.$$

Define the deviation of log effective agricultural ability from mean,

$$\hat{u}_{ai} = \log(\varphi_i) + \log(s_{ai}) - \mu_\varphi - \mu_a = u_{\varphi i} + u_{ai}.$$

Note that  $u_{ni}$  is normally distributed with mean  $E(u_{ni}) = 0$  and variance  $VAR(u_{ni}) = E(u_{ni}^2) = \sigma_n^2$ .

In turn,  $\hat{u}_{ai}$  is also normally distributed with mean  $E(\hat{u}_{ai}) = E(u_{\varphi i}) + E(u_{ai}) = 0$  and variance,

$$VAR(\hat{u}_{ai}) = \sigma_\varphi^2 + \sigma_a^2 + 2\sigma_{a\varphi} \equiv \hat{\sigma}_a^2.$$

Since  $s_n$  and  $\varphi$  are uncorrelated, the covariance of  $\widehat{u}_{ai}$  and  $u_{ni}$  is given by,

$$COV(\widehat{u}_{ai}, u_{ni}) = E[(u_{ai} + u_{\varphi i}) u_{ni}] = \sigma_{an}.$$

Finally note that  $(u_n - \widehat{u}_a)$  has mean  $E(u_{ni} - \widehat{u}_{ai}) = 0$  and variance given by,

$$VAR(u_{ni} - \widehat{u}_{ai}) = \widehat{\sigma}_a^2 + \sigma_n^2 - 2\sigma_{an} \equiv \sigma^2.$$

The log-incomes of individual  $i$  from agriculture and non-agriculture respectively are,

$$\log(I_{ai}) = \log(w_a) + \log(\varphi_i) + \log(s_{ai}),$$

$$\log(I_{ni}) = \log(w_n) + \log(1 - \eta) + \log(s_{ni}).$$

We can re-write agricultural and non-agricultural incomes as the sums of constants and log mean deviations,

$$\log(I_{ai}) = b_a + u_{\varphi i} + u_{ai} = b_a + \widehat{u}_{ai}, \quad (\text{D.19})$$

$$\log(I_{ni}) = b_n + u_{ni}, \quad (\text{D.20})$$

where  $b_a \equiv \log(w_a) + \mu_\varphi + \mu_a$  and  $b_n \equiv \log(w_n) + \log(1 - \eta) + \mu_n$ .

The probability an individual chooses to be a farm operator in agriculture,

$$\begin{aligned} n_a &= Pr\{\log(I_{ai}) > \log(I_{ni})\} = Pr(b_a + \widehat{u}_{ai} > b_n + u_{ni}) = \\ &= Pr(b_a - b_n > u_{ni} - \widehat{u}_{ai}) = Pr\left(\frac{b_a - b_n}{\sigma} > \frac{u_{ni} - \widehat{u}_{ai}}{\sigma}\right). \end{aligned}$$

Let  $b \equiv \frac{b_a - b_n}{\sigma}$  and note that  $\xi_i \equiv \frac{u_{ni} - \widehat{u}_{ai}}{\sigma}$  is a standard normal random variable. Then,  $n_a = \Phi(b)$ , where  $\Phi(\cdot)$  is the standard normal cdf. Given that we have a continuum of individuals of measure 1,  $n_a$  is also the fraction of individuals that choose agriculture, i.e.,  $N_a = n_a$ . Similarly, we can show that the probability an individual chooses to work in non-agriculture (and therefore the fraction of individuals that choose non-agriculture) is  $N_n = 1 - \Phi(b)$ .

We use the conditional averages of log-effective sectoral abilities to illustrate the possible patterns of sorting of individuals across sectors and the average quality of those that choose to work in each sector relative to the population. The average log-effective ability in agriculture among those that choose to work in agriculture is,

$$E\{\log(\widehat{s}_{ai}) | i \in H_a\} = \widehat{\mu}_a + \frac{\sigma_{an} - \widehat{\sigma}_a^2}{\sigma} \lambda^l(b), \quad (\text{D.21})$$

while the average log-ability in non-agriculture among those choosing non-agriculture is,

$$E \{ \log (s_{ni}) | i \in H_n \} = \mu_n + \frac{\sigma_n^2 - \sigma_{an}}{\sigma} \lambda^u (b), \quad (\text{D.22})$$

where  $\hat{\mu}_a = \mu_a + \mu_\varphi$ .  $\lambda^l (b) \equiv E [\xi | \xi \leq b] < 0$  and  $\lambda^u (b) \equiv E [\xi | \xi > b] > 0$  represent lower tail truncation and upper tail truncation of a standard normal random variable  $\xi$ . The coefficients in (D.21) and (D.22) can be re-written as  $\frac{\hat{\sigma}_a \sigma_n}{\sigma} \left[ \rho_{an} - \frac{\hat{\sigma}_a}{\sigma_n} \right]$  and  $\frac{\hat{\sigma}_a \sigma_n}{\sigma} \left[ \frac{\sigma_n}{\hat{\sigma}_a} - \rho_{an} \right]$ . As a result, the average quality of those that choose to work in a given sector relative to the average quality in the population depends on the dispersions of effective abilities in agriculture  $\hat{\sigma}_a$  and non-agriculture  $\sigma_n$ , and their correlation  $\rho_{an}$ . For example, if effective abilities are sufficiently positively correlated across sectors and the dispersion of non-agricultural ability is larger in relative terms ( $\hat{\sigma}_a^2 < \sigma_{an}$  and  $\sigma_n^2 > \sigma_{an}$ ), then the average effective ability of those in agriculture (non-agriculture) is lower (higher) than the population average.

## E Calibration of Sectoral Model with Selection

We calibrate distortions, abilities, and sectoral selection in a Benchmark Economy (BE) to the panel household-level data from China. We proceed in two steps. First, we infer population parameters on abilities and distortions from observed moments on sectoral incomes and estimated wedges. Second, given the calibrated population moments in the first step, we calibrate the remaining parameters from the general equilibrium equations of the sectoral model to match relevant data targets. We describe these steps in detail.

**Inferring population moments.** Assuming a multivariate log-normal distribution for the joint population distribution of abilities and distortions, we first back out the moments of that distribution (variances and covariances) so that we match observed moments on incomes across sectors and farm-specific distortions. We use the moments implied by our permanent fixed effects estimates of these variables. Normalizing the means of these distributions to zero, there are five population moments that need to be calibrated: three variances,  $\sigma_a^2$ ,  $\sigma_n^2$ ,  $\sigma_\varphi^2$  and two covariances,  $\sigma_{a\varphi}$ ,  $\sigma_{an}$ . These moments govern the occupational choices of individuals in the economy. To back out the population moments we: (i) construct in the model moments on sectoral incomes and farm distortions, conditional on sectoral choices as functions of the population moments; (ii) compute the counterparts to the conditional moments in our panel-data from China; and (iii) solve a system of equations for the population moments.

Exploiting log-normality, our system of equations on conditional moments consists of:

1. Variance of log income in agriculture conditional on choosing agriculture,

$$VAR \{\log(I_{ai}) | i \in H_a\} \equiv \hat{v}_a = \hat{\sigma}_a^2 \left\{ 1 - \left( \frac{\sigma_{an} - \hat{\sigma}_a^2}{\sigma \hat{\sigma}_a} \right)^2 \lambda^l(b) [\lambda^l(b) - b] \right\}. \quad (\text{E.23})$$

2. Variance of log income in non-agriculture conditional on choosing non-agriculture,

$$VAR \{\log(I_{ni}) | i \in H_n\} \equiv \hat{v}_n = \sigma_n^2 \left\{ 1 - \left( \frac{\sigma_n^2 - \sigma_{an}}{\sigma \sigma_n} \right)^2 \lambda^u(b) [\lambda^u(b) - b] \right\}. \quad (\text{E.24})$$

3. Covariance of log incomes in agriculture, non-agriculture conditional on choosing agriculture,

$$COV \{\log(I_{ai}), \log(I_{ni}) | i \in H_a\} \equiv \hat{c}_{an} = \sigma_{an} - \left( \frac{\sigma_n^2 - \sigma_{an}}{\sigma} \right) \left( \frac{\sigma_{an} - \hat{\sigma}_a^2}{\sigma} \right) \lambda^l(b) [\lambda^l(b) - b]. \quad (\text{E.25})$$

4. Variance of log-distortions in agriculture conditional on choosing agriculture,

$$VAR \{\log(\varphi_i) | i \in H_a\} \equiv \hat{v}_\varphi = \sigma_\varphi^2 \left\{ 1 - \left( \frac{\sigma_\varphi^2 + \sigma_{a\varphi}}{\sigma \sigma_\varphi} \right)^2 \lambda^l(b) [\lambda^l(b) - b] \right\}. \quad (\text{E.26})$$

5. Covariance of log agricultural income and log distortions in agriculture conditional on choosing agriculture,

$$COV \{\log(I_{ai}), \log(\varphi_i) | i \in H_a\} \equiv \hat{c}_{I_a, \varphi} = (\sigma_{a\varphi} + \sigma_\varphi^2) + \left( \frac{\sigma_{an} - \hat{\sigma}_a^2}{\sigma} \right) \left( \frac{\sigma_\varphi^2 + \sigma_{a\varphi}}{\sigma} \right) \lambda^l(b) [\lambda^l(b) - b]. \quad (\text{E.27})$$

A key aspect of our empirical approach is that we compute conditional moments in our panel data over the estimated fixed effect permanent components of distortions, agricultural income, and non-agricultural income for each household. Specifically, we use panel methods to estimate permanent measures of land input and non-agricultural income and then along with the permanent estimates of TFP and TFPR we back out all the other variables of interest. To obtain fixed effect estimates for other farm-level variables we apply a method similar to that in equations (4) and (5), on land input and non-agricultural income and then use the model equations to solve for the rest. In particular, we first decompose land input, and non-agricultural income as follows,

$$\log \ell_{vit} = \mu_t^\ell + \mu_i^\ell + e_{ivt}^\ell \quad (\text{E.28})$$



$$\log I_{n,vit} = \mu_t^{I_n} + \mu_i^{I_n} + e_{ivit}^{I_n} \quad (\text{E.29})$$

and then we partial out village-level factors by extracting the residuals from,

$$\mu_i^\ell = \mu_v^\ell + \zeta_i^\ell, \quad (\text{E.30})$$

$$\mu_i^{I_n} = \mu_v^{I_n} + \zeta_i^{I_n}, \quad (\text{E.31})$$

The interpretation for the regressors is the same as for equations (4) and (5) in the paper. Using this procedure we estimate permanent farm-specific components of land input  $\widehat{\zeta}_i^\ell$ , and non-agricultural income  $\widehat{\zeta}_i^{I_n}$ . Denote the permanent measures of land input  $\ell_i$  and non-agricultural income  $I_{n,i}$  as the exponential of the estimates  $\widehat{\zeta}_i^\ell$ , and  $\widehat{\zeta}_i^{I_n}$  respectively. We also have that our measure of distortions in the model is,  $\varphi_i = (1/TFPR_i)^{1/(1-\gamma)}$ , and permanent agricultural income  $I_{a,i}$  in the model is agricultural output, which can itself be recovered from  $TFP_i$ ,  $TFPR_i$ , and  $\ell_i$ . We then compute the moments of interest over the above estimated permanent components of distortions, agricultural income, and non-agricultural income.

With the permanent farm measures, we then compute empirical moments on the standard deviations of log agricultural income; log non-agricultural income; log distortions for farm operators; covariance of log agricultural income and log distortions; and the covariance of log agricultural income and log non-agricultural income. This last moment requires some discussion as a typical limitation of empirical models of selection is that income is observed only for chosen occupations. An advantage of our setting is that for the vast majority of households (around 96 percent), income is observed in both agricultural and non-agricultural activities; and many households switch from agriculture to non-agriculture under a variety of definitions of switchers in our panel data. The moment we use as our baseline is the contemporaneous covariance of log sectoral incomes across households, which implies a correlation of 0.033 in our micro-data.

We emphasize that this moment is robust to two alternative classifications of households in agriculture and non-agriculture over the sample period. Our first classification defines a household in non-agriculture if their reported cultivated land and farm output are zero; the second classifies a household in non-agriculture if they self-report based on time allocation as mainly in non-agriculture or are full-time in non-agriculture. Identifying switchers on the basis of these alternative definitions, we re-estimate the correlation in log incomes in agriculture and non-agriculture. Compared to an estimate of 0.03 in the contemporaneous correlation, we find a correlation between log income in agriculture and non-agriculture of 0.036 (0.020) for switchers using the first (second) classification. In our system of equations, the 5 population moments of variances and covariances are identified by the 5 conditional moments of variances and covariances given by equations (E.23)-(E.27). Table E.1

contains the empirical conditional variances and covariances along with the share of employment in agriculture that we target.

Table E.1: Targeted Empirical Conditional Moments

Statistic	Description	Value
$N_a$	Share of labor in agriculture	0.46
$\hat{v}_a$	STD of agricultural income	0.34
$\hat{v}_n$	STD of non-agricultural income	0.46
$\hat{v}_\varphi$	STD of farm distortions	1.05
$\hat{c}_{an}$	COV between agricultural and non-agricultural incomes	0.005
$\hat{c}_{a\varphi}$	COV of agricultural income and farm distortions	-0.14

Notes: All variables refer to logs.

We follow these specific steps to recover the population moments of the distributions of abilities across sectors and distortions:

1. Using equation (D.17) we invert the standard normal to recover the parameter  $b$  that generates a share of employment in agriculture of 46 percent. This gives a  $b = -0.10$ .
2. Note that equations (E.23), (E.24), and (E.25) give the variance of the log of agricultural income conditional on choosing agriculture,  $\hat{v}_a$ ; the variance of the log of non-agricultural income conditional on choosing non-agriculture,  $\hat{v}_n$ ; and the covariance of the two conditional on having chosen agriculture  $\hat{c}_{an}$ , in terms of the dispersion in effective abilities in agriculture  $\hat{\sigma}_a$  and non-agriculture  $\sigma_n$ , and the covariance of abilities  $\sigma_{an}$  alone. We solve this  $3 \times 3$  system for the three population moments  $\hat{\sigma}_a$ ,  $\sigma_n$ ,  $\sigma_{an}$  to match the observed conditional moments on incomes from the panel data on China.
3. We then solve for the dispersion of abilities in agriculture  $\sigma_a$ , the dispersion of distortions  $\sigma_\varphi$ , and the covariance of abilities in agriculture and distortions  $\sigma_{a\varphi}$  using the  $3 \times 3$  system in equations (E.26), (E.27), and the definition of  $\hat{\sigma}_a$ . These equations give the variance of the log of distortions  $\hat{v}_\varphi$ , the covariance of log agricultural income and log distortions conditional on working in agriculture  $\hat{c}_{a,\varphi}$ , and the definition of the variance of agricultural ability in relation to the variance of effective agricultural ability solved in previous step 2.

This procedure ensures that the occupational choices of individuals are consistent with the observed share of employment in agriculture of 46 percent in China.

Table E.2: Calibrated Population Moments

Parameter	Description	Value
$\sigma_a$	STD of agricultural ability	1.30
$\sigma_n$	STD of non-agricultural ability	0.65
$\sigma_\varphi$	STD of distortions	1.06
$\rho_{a\varphi}$	CORR of agricultural ability and distortions	-0.95
$\rho_{an}$	CORR of agricultural–non-agricultural ability	-0.15

Notes: All variables refer to logs.

In Table E.2 we report the resulting population moments following the procedure outlined above. Note that instead of reporting the covariances of agricultural and non-agricultural abilities and of agricultural ability and distortions, we report their respective correlations,  $\rho_{a\varphi}$  and  $\rho_{an}$ , which have a more intuitive interpretation. The correlation of abilities across sectors is negative ( $-0.15$ ), which means that individuals that are skilled farmers tend to be less skilled in non-agricultural occupations. This implies that individuals sort into the sector in which they possess a comparative advantage, and that individuals working in each sector are on average more skilled in that sector than the general population. The correlation of ability in agriculture  $s_a$  and distortions  $\varphi$  is strongly negative ( $-0.95$ ), consistent with our description of the institutional environment in China.

**Calibrating remaining parameters.** In order to calibrate the remaining parameters and to simulate the model we generate correlated data of 1,000,000 triplets  $(s_a, \varphi, s_n)$ , drawn from a multivariate log-normal distribution, using the inferred population moments from the previous step.<sup>1</sup>

We calibrate the remaining parameters using the generated correlated data, which embed the distributional properties of the population moments, so that the model equations constitute an equilibrium. The parameters to calibrate in this step are:  $A_n$  productivity in non-agriculture which is normalized to 1;  $(\alpha, \gamma)$  the elasticity parameters in the technology to produce the agricultural good, which are set to  $\alpha = 0.66$  and  $\gamma = 0.54$ , following our analysis of measuring farm TFP and distortions in agriculture in Section 4 of the paper; the endowment of land  $L$  is set to match an average farm size of 0.45 hectares observed in our micro data, which given our target for the share of employment in agriculture implies  $L = 0.207$ ; and  $\omega$ , the weight of the agricultural good in preferences, is set to 0.01, which implies a long-run share of employment in agriculture of 1 percent. In our model with period-by-period growth, the subsistence constraint of agricultural goods becomes asymptotically negligible (i.e.,  $\bar{a} = 0$ ) and the share of employment would be solely de-

<sup>1</sup>We find that drawing a large sample of 1,000,000 data points produces the same results as drawing 10,000 samples of 10,000 data points each, and taking the average.

terminated by the parameter  $\omega$  in preferences, regardless of the presence of barriers or distortions. Today’s developed countries observe a share of employment in agriculture below 1.5 percent, hence we conservatively set this “long-run” share of employment in agriculture to 1 percent.

The remaining four parameters: subsistence of agriculture goods  $\bar{a}$ , productivity in agriculture  $A_a$ , capital endowment in agriculture  $K_a$ , and labor mobility barrier  $\eta$ ; are selected by solving the equilibrium of the model to match four targeted moments: the share of employment in agriculture of 46 percent, a normalized to one relative price of agriculture, a capital to output ratio in agriculture of 0.3 observed in our micro data, and a ratio of labor productivity in non-agriculture to agriculture of 3.96. Table E.3 displays the aggregate and micro-level statistics for the benchmark economy, as well as the values for the calibrated parameters. Real agricultural labor productivity is total agricultural output  $Y_a$  divided by employment in agriculture  $N_a$ .  $TFP_a$  is total agricultural output divided by the bundle of inputs  $N_a^{1-\gamma} (L^\alpha K^{1-\alpha})^\gamma$  from the implied aggregate production function.

Table E.3: Calibrated Benchmark Economy (BE)

Statistic	Description	Value in BE
$Y_a/N_a$	Real agricultural labor productivity	0.44
$N_a$	Share of employment in agriculture	0.46
$TFP_a$	TFP in agriculture	0.84
$(Y_n/N_n) / (Y_a/N_a)$	Real non-agricultural to agricultural productivity gap	3.95
$Z_a/N_a$	Average ability in agriculture	3.43
$Z_n/N_n$	Average ability in non-agriculture	1.72
$(Z_n/N_n) / (Z_a/N_a)$	Ratio of non-agricultural to agricultural ability	0.50
$Y/N$	Real GDP per worker	1.13
$Y_a^e/Y_a$	Efficiency gain among existing farmers	1.15
Conditional Micro-level Statistics		
	STD of log– farm TFP	0.56
	STD of log– farm TFPR	0.48
	CORR of log– farm TFP and log– farm TFPR	0.97
Calibrated Parameters		
$\bar{a}$	subsistence constraint	0.20
$A_a$	productivity in agriculture	0.27
$K_a$	capital stock in agriculture	0.06
$\eta$	labor mobility barrier	0.74

## F Additional Quantitative Experiments

Our main counterfactual in Section 5 of the paper eliminates the correlation of distortions with agricultural productivity across farmers. This section presents the results from additional quantitative experiments.

**Eliminating all distortions.** In Table F.4 we present the results from removing all farm-specific distortions in agriculture, i.e., setting  $\varphi = 1$  for all  $i$ . Eliminating all distortions on farms has a substantial impact on the economy. Agricultural labor productivity increases 3.4-fold and the share of employment in agriculture falls 32 percentage points, from 46 percent to 14 percent. Note, however that the magnitude of these effects is similar to those when we eliminate only the correlation of distortions with agricultural ability analyzed in Section 5 of the paper. This implies that the key driver of the results is the systematic nature of the distortions, associated with the land institution we emphasize, that particularly constrain the more productive farmers.

Table F.4: Effects of Eliminating Distortions

Statistic	Benchmark Economy BE	No Correlated Distortions	No Distortions $\varphi_i = 1$
Aggregate Statistics			
Real Agricultural Productivity ( $Y_a/N_a$ )	1.00	2.96	3.42
Share of Employment in Agriculture ( $N_a$ ) (%)	0.46	0.16	0.14
TFP in Agriculture ( $TFP_a$ )	1.00	1.67	1.80
TFP in Agriculture, constant BE farms	1.00	1.10	1.15
Real Non-Agricultural Productivity ( $Y_n/N_n$ )	1.00	0.78	0.77
Average Ability in Agriculture ( $Z_a/N_a$ )	1.00	2.34	2.65
Average Ability in Non-Agriculture ( $Z_n/N_n$ )	1.00	0.78	0.77
Real GDP per Worker ( $Y/N$ )	1.00	1.18	1.19
Conditional Micro-level Statistics			
STD of log–farm TFP	0.56	0.39	0.35
STD of log–farm TFPR	0.48	0.14	0
CORR of log–(farm TFP, farm TFPR)	0.97	0.44	–
CORR of log–(agr. ability, non-agr. ability)	0.15	0.49	0.50
CORR of log–(agr. income, non-agr. income)	0.03	0.40	0.50

Notes: The counterfactual “No Correlated Distortions” refers to the economy when eliminating correlated distortions. The counterfactual “No Distortions” eliminates all farm-level distortions, i.e., we set  $\varphi_i = 1$  for all  $i$ . All aggregate variables, except for the share of employment in agriculture, are reported relative to the same statistic in the Benchmark Economy (BE). All micro-level statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

**Amplification effect of distortions and  $\rho_{an}$ .** We discuss our main results with alternative values for the population correlation of abilities across sectors  $\rho_{an}$  considered in the literature and show that our results are conservative in terms of potential amplification effects. In the benchmark economy, the calibrated population correlation  $\rho_{an}$  is  $-0.15$ . The results from this calibration are reproduced in columns 2-3 in Table F.5. We consider two alternatives, of  $\rho_{an} = 0$  (columns 4-5) and  $\rho_{an} = 0.15$  (columns 6-7). In each case we re-calibrate all parameters of the benchmark economy, including the rest of the population moments, to match the same empirical moments in the data (except for the conditional correlation of incomes across sectors). This is column “BE” under each  $\rho_{an}$ . Comparing the “BE” column in each case the share of employment in agriculture and the micro-level statistics are identical by virtue of the calibration. The only exception is the correlation of incomes across sectors, which increases from 0.03 in the baseline (targeted) calibration to 0.40 under no correlation of abilities ( $\rho_{an} = 0$ ), and 0.76 under a positive correlation of abilities ( $\rho_{an} = 0.15$ ). The column “NC” under each  $\rho_{an}$  refers to the counterfactual experiment of eliminating correlated farm-level distortions in each economy. Table F.5 shows that our baseline results are robust to higher values of  $\rho_{an}$ , with the effects on agricultural productivity being stronger. Removing distortions has a substantial effect on aggregate agricultural productivity, 2.96-fold in the baseline calibration, 3.44 and 4.23-fold with the calibrations for  $\rho_{an}$  of 0 and 0.15, respectively. This is largely due to a stronger effect of selection on agricultural TFP, which increases by 80 and 98 percent compared to 67 percent in the baseline and the consequent larger decline in agricultural employment.

**Comparison to an exogenous increase in productivity.** Improvements in resource allocation in agriculture produce an increase in agricultural productivity and labor reallocation away from agriculture. Qualitatively such effects can also be generated through an exogenous increase in agricultural TFP or economy-wide TFP. To put our results from reduced misallocation in context, we compare them to the results from a 10 percent exogenous increase in TFP, corresponding to the efficiency gains from eliminating misallocation across existing farms with different productivity (correlated distortions) in our model.

In the first two columns of Table F.6 we reproduce the results for the benchmark economy and the economy without correlated farm-level distortions (our main counterfactual). In columns three and four we show in turn the effects of increasing exogenously TFP in agriculture and then in both agriculture and non-agriculture, by 10 percent (keeping all farm-level distortions in place). An exogenous increase in TFP reduces the share of employment in agriculture from 46 percent in the benchmark economy to 34 percent, however, agricultural TFP increases only by 15 percent, compared to 67 percent when eliminating correlated idiosyncratic distortions. Agricultural TFP increases a bit more than the exogenous increase in TFP because there is a small effect on selection

Table F.5: Eliminating Correlated Distortions with Alternative  $\rho_{an}$ 

Statistic	$\rho_{an} = -0.15$		$\rho_{an} = 0$		$\rho_{an} = 0.15$	
	BE	NC	BE	NC	BE	NC
Aggregate Statistics:						
Real Agricultural Productivity ( $Y_a/N_a$ )	1.00	2.96	1.00	3.44	1.00	4.23
Share of Employment in Agriculture ( $N_a$ ) (%)	0.46	0.16	0.46	0.14	0.46	0.11
TFP in Agriculture ( $TFP_a$ )	1.00	1.67	1.00	1.80	1.00	1.98
Real Non-Agricultural Productivity ( $Y_n/N_n$ )	1.00	0.78	1.00	0.78	1.00	0.77
Average Ability in Agriculture ( $Z_a/N_a$ )	1.00	2.34	1.00	2.72	1.00	3.35
Average Ability in Non-Agriculture ( $Z_n/N_n$ )	1.00	0.78	1.00	0.78	1.00	0.77
Real GDP per Worker	1.00	1.18	1.00	1.21	1.00	1.22
Conditional Micro-level Statistics:						
STD of log–farm TFP	0.56	0.39	0.56	0.35	0.56	0.31
STD of log–farm TFPR	0.48	0.14	0.48	0.13	0.48	0.12
CORR of log–(farm TFP, farm TFPR)	0.97	0.44	0.97	0.24	0.97	-0.08
CORR of log–(agr. ability, non-agr. ability)	0.15	0.49	0.28	0.54	0.38	0.57
CORR of log–(agr. income, non-agr. income)	0.03	0.40	0.45	0.76	0.76	0.72

Notes: “BE” refers to the calibrated benchmark economy with each population correlation of abilities  $\rho_{an}$ . “NC” refers to the experiment of eliminating the correlation of distortions within the economy with the corresponding  $\rho_{an}$ . All aggregate variables, except for the share of employment in agriculture, are reported relative to the same statistic in the benchmark economy in each correlation case. All micro-level statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

into agriculture, increasing average quality of workers in agriculture by 11 percent, compared to 134 percent when eliminating correlated distortions. The effects are similar when non-agricultural TFP also increases exogenously by 10 percent (column four).

The reduction in misallocation associated with the elimination of correlated farm-level distortions has a much larger effect on agricultural labor productivity than an equivalent-in-magnitude exogenous increase in TFP. When TFP increases exogenously, there is only a small effect in selection as explained above, which operates through general equilibrium effects (via changes in relative prices). To see this, note that if relative prices remained unchanged, a 10 percent increase in both  $A_a^{1-\gamma}$  and  $A_n$  would have no effect on occupational choices as they would leave the relative return to agriculture and non-agriculture unaltered for every individual. In the case of reduced misallocation, selection works to generate a large amplification effect on agricultural labor productivity, over and above the misallocation gains across existing farmers of 10 percent. The reason for this is that farm-level distortions directly impact the occupational choices of individuals, particularly for those with high agricultural ability. Removing farm-level distortions alters the pattern of occupational choices of individuals even holding constant aggregate prices.

Table F.6: Comparison of Removing Distortions vs. Exogenous TFP Increases

Statistic	BE	No Corr Dist	$\uparrow (A_a^{1-\gamma})$ $\times 1.10$	$\uparrow (A_a^{1-\gamma}, A_n)$ $\times 1.10$
Aggregate Statistics				
Real Agricultural Productivity ( $Y_a/N_a$ )	1.00	2.96	1.35	1.35
Share of Employment in Agriculture ( $N_a$ ) (%)	0.46	0.16	0.34	0.34
TFP in Agriculture ( $TFP_a$ )	1.00	1.67	1.15	1.15
Real Non-Agricultural Productivity ( $Y_n/N_n$ )	1.00	0.78	0.92	1.01
Average Ability in Agriculture ( $Z_a/N_a$ )	1.00	2.34	1.11	1.11
Average Ability in Non-Agriculture ( $Z_n/N_n$ )	1.00	0.78	0.92	0.92
Real GDP per Worker ( $Y/N$ )	1.00	1.18	1.09	1.18
Conditional Micro-level Statistics				
STD of log- farm TFP	0.56	0.39	0.56	0.56
STD of log- farm TFPR	0.48	0.14	0.48	0.48
CORR of log-(farm TFP, farm TFPR)	0.97	0.44	0.97	0.97
CORR of log- (agr. ability, non-agr. ability)	0.15	0.49	0.20	0.19
CORR of log- (agr. income, non-agr. income)	0.03	0.40	0.12	0.12

Notes: The first column “BE” refers to the benchmark economy. The second column “No Corr Dist” refers to the counterfactual of eliminating all the correlation of farm-level distortions with farm-level productivity. The third column refers to the case of exogenously increasing TFP in agriculture 1.10-fold relative to the benchmark, and the fourth column refers to the case of increasing TFP in both agriculture and non-agriculture 1.10-fold relative to the benchmark. All aggregate variables, except for the share of employment in agriculture, are reported relative to the same statistic in the benchmark economy. All micro-level statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

This result is important as a challenge in the literature is to find measurable drivers of sectoral reallocation and increased productivity in agriculture relative to non-agriculture. In an important contribution, [Lagakos and Waugh \(2013\)](#) highlighted selection as a substantial amplification mechanism of productivity differences, an insight we build on in our paper. But as emphasized in our results, reasonable economy-wide productivity differences are unlikely to generate differences in sectoral reallocation and selection large enough to explain the large real sectoral productivity gaps across rich and poor countries. We provide a measure of idiosyncratic distortions in agriculture as a specific and distinct driver of sectoral reallocation that has a strong effect on occupational choices and selection, generating both a direct effect on agricultural productivity and an amplification effect that is orders of magnitude larger than the effect from aggregate distortions or economy-wide productivity differences.

**Model specification and calibrated  $\rho_{an}$ .** We show that model specification matters for the calibrated population correlation of abilities  $\rho_{an}$ , but targeting the same empirical correlation of



Table F.7: No Correlated Distortions with Idiosyncratic Barriers

Statistic	$\sigma_\theta = 0$		$\sigma_\theta = 0.5$		$\sigma_\theta = 0.9$	
	BE	NC	BE	NC	BE	NC
Calibrated Ability Correlation	$\rho_{an} = -0.15$		$\rho_{an} = -0.09$		$\rho_{an} = -0.04$	
Aggregate Statistics						
Real Agricultural Productivity ( $Y_a/N_a$ )	1.00	2.96	1.00	3.06	1.00	2.94
Share of Employment in Agriculture ( $N_a$ ) (%)	0.46	0.16	0.46	0.16	0.46	0.16
TFP in Agriculture ( $TFP_a$ )	1.00	1.67	1.00	1.71	1.00	1.68
Real Non-Agricultural Productivity ( $Y_n/N_n$ )	1.00	0.78	1.00	0.81	1.00	0.85
Average Ability in Agriculture ( $Z_a/N_a$ )	1.00	2.34	1.00	2.43	1.00	2.36
Average Ability in Non-Agriculture ( $Z_n/N_n$ )	1.00	0.78	1.00	0.81	1.00	0.85
Real GDP per Worker	1.00	1.18	1.00	1.23	1.00	1.28
Conditional Micro-level Statistics						
STD of log-farm TFP	0.56	0.39	0.56	0.40	0.57	0.43
STD of log-farm TFPR	0.48	0.14	0.48	0.14	0.48	0.14
CORR of log-(farm TFP, farm TFPR)	0.97	0.44	0.96	0.31	0.96	0.18
CORR of log-(agr. ability, non-agr. ability)	0.15	0.49	0.10	0.41	0.05	0.30
CORR of log-(agr. income, non-agr. income)	0.03	0.40	0.03	0.37	0.03	0.28

Notes: “Calibrated Skill Corr.” refers to the calibrated  $\rho_{an}$  targeting the same empirical correlation of incomes across sectors of 0.03, under different  $\sigma_\theta$ . “BE” refers to the calibrated benchmark economy with each dispersion in mobility barriers  $\sigma_\theta$ . “NC” refers to the experiment of eliminating the correlation of distortions within the economy with the corresponding pair of  $\sigma_\theta$  and  $\rho_{an}$ . All aggregate variables, except for the share of employment in agriculture, are reported relative to the same statistic in the benchmark economy in each correlation case. All micro-level statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

incomes across sectors imposes discipline on the amplification effect from eliminating correlated farm-level distortions. In our baseline calibration, as we have abstracted from other idiosyncratic factors determining occupational choice, a near zero income correlation even with distortions in agriculture, requires a low value of ability correlation. Does our amplification result hinge on this particular inference and model structure? We explore alternative settings for the occupational choice problem in the model to include idiosyncratic preference or mobility barriers, denoted by  $\theta_i$ . The occupational decision depends now on the barrier draw, i.e., an individual selects agriculture if  $I_{ai} > \theta_i I_{ni}$ . For simplicity, we assume that  $\theta_i$  is *i.i.d.* and drawn from a log-normal distribution with mean 0 and variance  $\sigma_\theta^2$ . In Table F.7 for different values of  $\sigma_\theta$  we recalibrate the benchmark economy targeting the exact same correlation of incomes across sectors of 0.03. In each case, we recalibrate all moments and parameters to match the same other data targets. Column “BE” in each case presents the moments for the re-calibrated economy under each  $\sigma_\theta$ . We find that increasing the dispersion of idiosyncratic barriers increases the implied  $\rho_{an}$  required to produce the same empirical

correlation of incomes across sectors. The column “NC” in each case shows the results from the main quantitative experiment of eliminating correlated farm distortions. The results are fairly similar across the different values of  $\sigma_\theta$ . The fact that the results are robust against different  $\sigma_\theta$  and implied  $\rho_{an}$ , indicate that while the calibration does not identify the “true” value of  $\rho_{an}$ , the targeted value of the cross-sector income correlation imposes discipline on the magnitude of the amplification mechanism from eliminating correlated distortions in the model.

## References

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *The American Economic Review*, 103(2):948–980.