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 January 2020

We are grateful to Chaoran Chen for providing excellent research assistance. We also thank for useful comments the editor, three anonymous referees, Carlos Urrutia, Daniel Xu, and seminar participants at Toronto, McGill, Ottawa, REDg Barcelona, NBER, IMF, Cornell, Michigan, UBC, Xiamen U., ASU, Tsinghua-UCSD, HKUST, West- ern, 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. Restuccia gratefully acknowledges the support from the Canada Research Chairs program. Brandt thanks the support from the Noranda Chair for International Trade and Economics. Adamopoulos, Brandt, and Restuccia gratefully acknowledge the support from the Social Sciences and Humanities Research Council of Canada. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 January 2020
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 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 in China 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). We show that selection can substantially amplify the static misallocation effect of distortionary policies by affecting occupational choices that worsen the distribution of productive units in agriculture.

Tasso Adamopoulos
York University
Canada
aadamo@yorku.ca

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

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

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.[1] 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.[2] 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).

The paper contributes to the literature on two fronts. 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 land quality, other village factors,

[1] See Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Restuccia and Rogerson (2013, 2017) review the expanding literature on misallocation and productivity.

[2] Recent studies linking resource misallocation to land market institutions include: land reforms in Adamopoulos and Restuccia (2019), the extent of marketed land across farm households in Restuccia and Santaeulalia-Llopis (2015), and the role of land titling in Chen (2017) and Gottlieb and Grobovsek (2015). De Janvry et al. (2015) study a land certification reform in Mexico delinking land rights from land use which allowed for a more efficient allocation of individuals across space.

and time effects. We show that these adjustments considerably limit the possibility of measurement error in our fixed effect 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 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. This reduces average agricultural productivity and widens even further the real productivity gap between the agricultural and non-agricultural sectors. 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 aggregate 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. This average size compares with 16 and 17 hectares in Belgium and the Netherlands, countries with similar amounts of arable land per capita as China, and to 178 hectares in the United States. 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.[3] The data allow us to construct precise real

[3]The data are collected by the Research Center for the Rural Economy under the Ministry of Agriculture as part of a nation-wide survey.

measures of value added and productivity at the farm-level, and to observe the incomes of the same households across sectors. In the context of a widespread shift into non-agricultural activities, these data offer a unique opportunity to examine the selection effect of distortionary policies.

Our goal is to measure the extent of misallocation across farmers implied by the land market institutions in China, and to quantify its consequences for occupational choices, agricultural productivity, and real GDP per capita. We do so by combining our detailed panel-level data from China and a two-sector model with selection and idiosyncratic distortions in agriculture.[4] The panel dimension of the data is important for our purposes because it allows us to: (a) obtain better measures of farm productivity and idiosyncratic distortions than what typical cross-sectional analyses of misallocation can; and (b) identify selection across sectors, by tracing the sectoral shifts of households and their incomes.

We proceed in two steps. Our first step is to use a diagnostic tool from modern macroeconomics, a heterogeneous firm-industry framework with minimal structure, to measure the deviations between marginal products and the overall extent of static inefficiency. In this set-up, the land market institutions 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 the richness and time-dimension of our longitudinal household data and use panel data methods to estimate "permanent" fixed effect farm-level productivity and distortions, adjusting for village-level land quality, other village factors and time effects. 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. First, our estimates control for land quality differences across villages by bringing in land quality and geographic data for China from the

[4] 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 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.

Global Agro-Ecological Zones project. Second, our estimates control for common factors varying over time and across villages. Third, our estimates control for transitory household idiosyncratic shocks, and extract permanent components of productivity and distortions. To formally show the value of our approach in dealing with measurement error we apply the methodology of Bils et al. (2018), which assesses the extent of additive measurement error in measures of distortions. We find that our estimates of permanent 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 sizable. Using our fixed effect estimates of productivity and distortions, the reallocation of capital and land 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 2/3 of these gains accounted for by labor mobility across villages. This suggests that the land policy in China may be a substantial contributor to the barriers of labor mobility across space (De Janvry et al., 2015; Ngai et al., 2019) and its depressing impact on productivity. Moreover, we do not find substantial changes in the extent of misallocation over time, consistent with an absence of substantial changes in China's land market institutions over this period.

Our second step is to 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 allow us to pin down the population correlation of abilities across sectors. We then conduct a series of counterfactual experiments to assess the quantitative importance of misallocation and its overall impact on aggregate agricultural productivity, accounting for distortions in sectoral occupational choices. We emphasize three sets of counterfactuals.

First, we assess the effect of misallocation on village-level productivity by eliminating all distortions within villages. This counterfactual generates a large 3.4-fold increase in agricultural labor productivity; a significant increase in agricultural TFP of 1.8-fold; and a substantial reallocation of labor across sectors, with the share of employment in agriculture falling from 46 percent to 14 percent. The total effect on agricultural productivity is substantially larger than the static effect of eliminating misallocation across existing farmers. The difference is due to the significant amplification effect that distortions have on the selection of farmers in the model, which produces an additional increase in agricultural TFP of 1.6-fold. That is, selection more than triples the impact of reduced misallocation on agricultural productivity.

Second, to isolate the contribution of correlated distortions within villages (i.e., the property of distortions that they increase with farm productivity), we eliminate only these correlated distortions by setting their correlation with agricultural ability to zero. This results in an increase of agricultural productivity of 2.4-fold, which is more than 70 percent of the increase in productivity from eliminating all distortions within villages. This implies that it is the systematic component of distortions, i.e., the fact that they affect more heavily the more productive farmers, which is responsible for most of the amplification effect on productivity through distorted occupational choices and selection.

Third, we compare the productivity gains from eliminating farm-level distortions to an equivalent exogenous increase in village-wide productivity. When increasing village-wide productivity exogenously we find that there is only a small amplification effect on agricultural TFP through general equilibrium prices, while the share of employment in agriculture falls to 30 percent compared to 14 percent when eliminating distortions. Distortions in the agricultural sector generate much larger selection effects than comparable changes in village-wide TFP because distortions have a direct impact on occupational choices instead of just the general equilibrium effects generated via common across-household shifts in production parameters.

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.[5] Second, we study the broader impact of misallocation, in particular, the effect of distortionary policies on misallocation and 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 in our work 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. Moreover, the panel dimension of the data allows us to identify a key parameter in models of selection capturing the correlation of abilities across sectors that has important implications for aggregate outcomes. We underline that in our calibration, economy-wide changes in TFP have no substantial amplification effects on agricultural TFP; hence, our selection results are distinct from Lagakos and Waugh (2013) in that they are mostly driven by the impact of idiosyncratic distortions on occupational choices.

The paper proceeds as follows. In the next section, we describe the specifics of the land market institutions in China which are intertwined with additional mobility restrictions across space through the "hukou" registration system. 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, discuss our measures of farm productivity and distortions from the panel data, and present the main results on misallocation in agriculture in China. Section 5 embeds the framework of agriculture into a heterogeneous-ability two-sector model with non-agriculture. We calibrate the model to aggregate, sectoral, and micro moments from the Chinese data in Section 6. Section 7 reports the main results from our quantitative experiments. We

[5]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 in Song et al. (2011). Tombe and Zhu (2019) and Ngai et al. (2019) analyze the important impact of migration restrictions in the "hukou" system for reallocation and welfare in China.

conclude in Section 8.

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.[6]

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 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.

[6] 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).

3 Data

We use household survey data collected by the Research Center for the Rural Economy under the Ministry of Agriculture of China.[7] 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 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.

[7] For a detailed description and analysis of the data see Benjamin et al. (2005).

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. Appendix A describes in more detail the measures of output and inputs that we use. Focusing on the household as the unit of observation rather than the plot helps reduce measurement error that arises in attributing inputs to individual plots and other plot-level measurement issues. Moreover, the panel structure of the data allows us to address important measurement concerns that we discuss in the next section.

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 an important ingredient into our two-sector analysis with selection in Section 5. We also report the static efficiency gains for the agricultural sector in China from eliminating farm-specific distortions, providing a benchmark number relative to which we can compare the efficiency gains that would result from eliminating distortions in our two-sector model with selection.

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,

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

where y, ℓ, 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 α captures the relative importance of land in production.[8]

Our starting point is the static efficient allocation of factors of production in the village economy at any point in time obtained from the solution to a simple planner's problem that takes the distribution of productivities as given. In Appendix B 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. 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 actual observed allocations in the Chinese economy. While this representation is not required for assessing the aggregate consequences of misallocation—since we can directly compare efficient allocations and output with the actual data—it will be useful in the estimation of the two-sector economy in

[8]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, generating 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 will be 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.

Section 5 and the subsequent counterfactual exercises. In Appendix C we derive the equilibrium of the distorted economy and describe our identification of the farm-input-specific distortions from the equilibrium conditions and data.

We construct the following summary measure of farm-specific distortions faced by farm i in village v at time t,

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

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. As shown in Appendix C, $TFPR_{ivt}$ 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 (the two input wedges, or the output wedge and one input wedge), but this choice will not influence the magnitude of the overall farm-specific distortion.

We emphasize that $TFPR$ is different from "physical productivity" or real TFP, which in our model is,

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

for farm i in village v at time t. In a world without distortions farms with higher physical productivity TFP_{ivt} command more land ℓ_{ivt} and capital k_{ivt}, and marginal products of each factor (and $TFPR$) equalize across farms. However, with idiosyncratic distortions this need not be the case.

Using the fact that total output is $Y_{vt} = \sum_{i=1}^{M_v} y_{ivt}$, in Appendix C we 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 observations in the 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 "physical productivity" or TFP, which we construct residually from the farm-level production function using equation (3) and data on operated land, capital, labor, and value added as described in Section 3. In our framework, household 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 from equation (3) requires values for the parameters γ and α. 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 estimates for China. 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, which pins down our value for $\gamma = 1 - 0.46$. Estimates for the land income share based on aggregate data agree on a value of 0.36 (Cao and Birchenall, 2013; Chow, 2002). Given γ, the share of land implies $\alpha = 2/3$. The implied factor share for capital 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, 2002).

Based on these values of (α, γ), equation (3) is used to compute TFP for each household, in each village, for each year. These TFP measures however are possibly subject to measurement error, unobserved land quality differences, 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 across farms. To address these issues we net out the effect of land quality, time factors, and other village characteristics to estimate "permanent" or farmer fixed effect levels of TFP. In particular, we decompose the logarithm of farm-level TFP as follows,

$$\log TFP_{vit} = \beta_0^{TFP} + \beta_1^{TFP} \log LQ_v + \zeta_t^{TFP} + \zeta_{iv}^{TFP} + e_{ivt}^{TFP}, \tag{4}$$

where β_0^{TFP} is a common intercept, and LQ_v is our measure of observed village-level land quality; ζ_t^{TFP} is a year fixed effect component that captures time varying shocks to productivity that are common to all farmers; ζ_{iv}^{TFP} is a farm-village-specific component that does not vary over time, and embeds factors specific to a farm within a village since equation (4) cannot separately identify village and farm fixed effects; and e_{ivt}^{TFP} captures idiosyncratic shocks specific to the farmer in a given year. Note however that we are interested in extracting ζ_i^{TFP} which is a farm-specific intercept or fixed farmer component that captures a farm's "permanent" productivity that is constant across years, and purged of village level factors. Our econometric procedure for recovering ζ_i^{TFP} is as follows: (a) we use fixed effect panel data methods to estimate equation (4) to extract the household fixed effects which are inclusive of the village-specific effects that do not change over time; and (b) we run the above household specific terms on village dummies (without a constant), and extract the residuals as our estimate of the pure farm idiosyncratic fixed-effect (permanent) component $\widehat{\zeta}_i^{TFP}$.

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, for all of the villages in our sample. 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 10×10 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 (2018), 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 with a standard deviation of the log of 0.096.[9]

Our baseline measure of farm TFP, which we denote as TFP_i is the exponential of the estimated household fixed effect $\widehat{\zeta}_i^{TFP}$. There are a few points to note. First, farm TFP is devoid of changes over time and differences across villages. Second, importantly the household fixed effect estimate controls for potential measurement error which is subsumed in the residual. To appreciate these points, note that dispersion, measured by the standard deviation of the log, in the cross-sectional measures of farm TFP that control for land quality varies between 0.85 and 0.94 over the ten year period and has a mean of 0.90. By contrast, the dispersion for our baseline measure of farm TFP is 0.35. The 90/10 percentile ratio in farm TFP is 2.2-fold whereas the 75/25 percentile ratio is 1.5-fold.[10]

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. While we have controlled for land quality differences across villages, it is possible there remains land quality differences across households within a village and that these explain the variation in TFP across households we observe. We argue 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 officials and farmers that households

[9] We construct the potential yield for each village in our micro dataset 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 (2018) for details.

[10] While the dispersion in farm-level TFP in our data is substantial, it is much lower than the cross-sectional dispersion of plant-level TFP documented in Hsieh and Klenow (2009) for Indian, Chinese and U.S. manufacturing. For instance, in Chinese manufacturing, the 90/10 percentile ratio of plant-level TFP is 12.7-fold whereas the 75/25 percentile ratio is 3.8-fold.

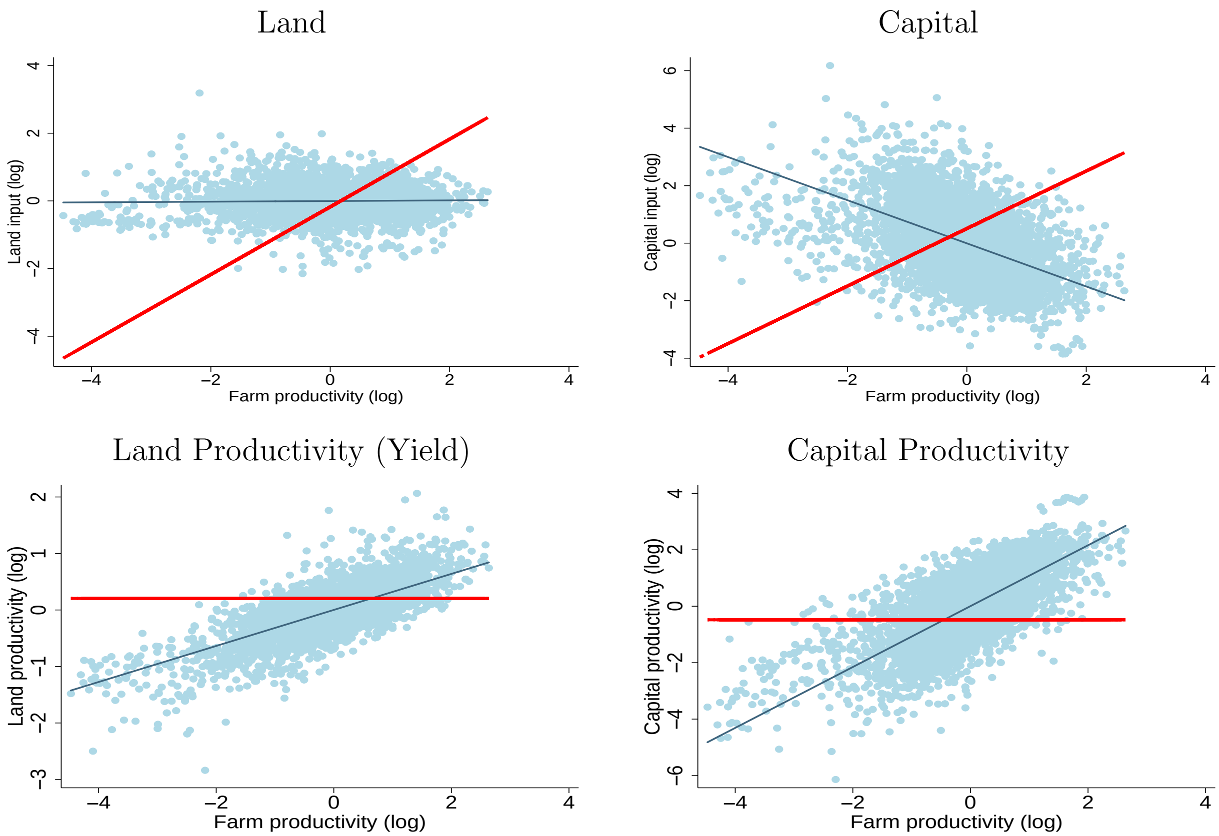
are allocated "bundles" of land, and hold similar portfolios of land in terms of land quality, and distance among plots.

4.3 Distortions and Agricultural Productivity

If land and capital were allocated across farms in a decentralized fashion through unhindered factor markets, the resulting allocations should resemble the efficient allocations characterized in Appendix B, with relatively more productive farmers operating at a larger scale with more land and capital. In other words, the relationship between 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 fact, 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 inputs. Land use and capital use are not systematically positively correlated with farm productivity in China. In addition, the average productivity of land and capital inputs are systematically positively correlated with farm productivity across farms in China. 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 we described in Section 2 including the fairly uniform administrative allocation of land among members of the village. The lack of ownership over the allocated plots (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).

Figure 1: Factor Allocations by Farm Productivity



Notes: The data refers to the household fixed effects from a panel regression. 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 and productivity. The input allocations across farmers in China indicate that there is substantial misallocation. In the context of the decentralized framework we presented in Section 4.1, this misallocation is manifested through farm-specific distortions or "wedges," measured as deviations of the observed input allocations from the efficient ones. We summarize the distortions faced by a farmer in both the land and capital markets by the measure of revenue productivity TFPR in equation (2). The further $TFPR_{ivt}$ for farmer i is from the village-time average, the higher the overall distortion that farmer i faces. (See also equation C.6 in the Appendix.) We follow the same procedure as with physical productivity to estimate farm-specific distortions using the panel data on $TFPR_{ivt}$ with controls for land quality and household, village, and time fixed effects. In particular we use panel methods to estimate,

$$\log TFPR_{vit} = \beta_0^{TFPR} + \beta_1^{TFPR} \log LQ_v + \zeta_t^{TFPR} + \zeta_{iv}^{TFPR} + e_{ivt}^{TFPR}. \tag{5}$$

The interpretation for the regressors is the same as for equation (4). We follow the same procedure as for equation (4) to estimate the permanent farm-specific components of distortions $\widehat{\zeta}_i^{TFPR}$. We refer to farm-$TFPR_i$ as the exponential of the estimated household fixed effect $\widehat{\zeta}_i^{TFPR}$.

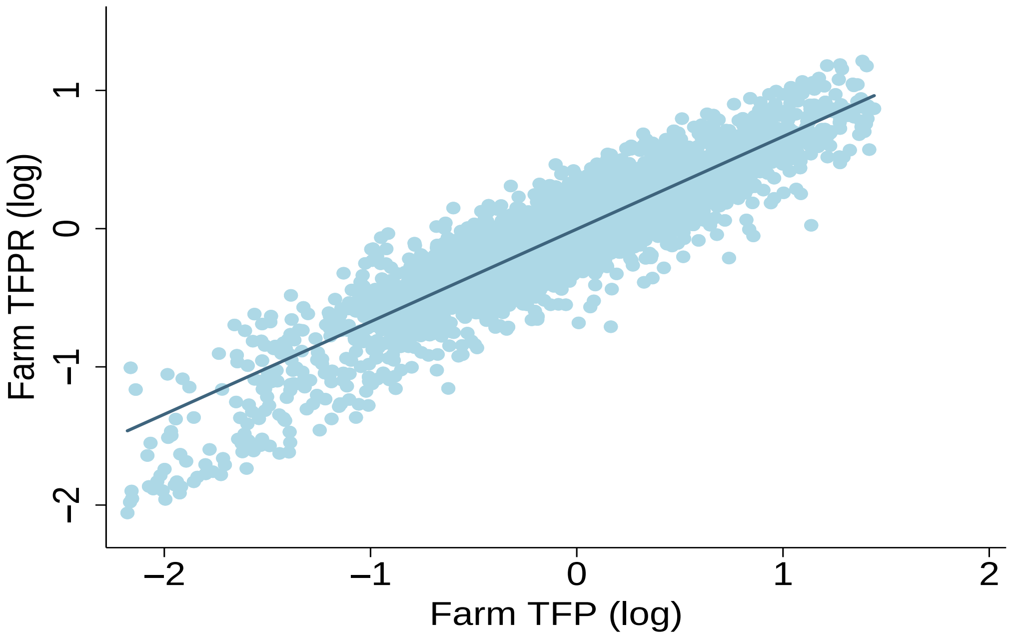
In Figure 2 we plot the farm-specific distortions, as captured by TFPR, against farm TFP (in logs) for our baseline measures. There is a strong positive correlation between farm distortions (TFPR) and farm productivity (TFP), with the correlation equal to 0.91.[11] The more productive farmers face higher 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 ones who would have wanted to expand the most in unfettered markets, acquiring more land and capital. In the context of our decentralized framework, this is reflected in higher farm-specific distortions on the more productive farmers.

Static efficiency gains. In order to measure the efficiency gains associated with the misallocation of resources in agriculture we conduct a counterfactual exercise. We ask how much larger would aggregate output (and as a result TFP) be in agriculture if all 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, $Y^e/Y - 1$.

Table 1 reports the efficiency gains from factor reallocation using our baseline measure of farm

[11] The dispersion in TFPR is also substantial. The standard deviation of log $TFPR_i$ is 0.48, while the 90/10 and 75/25 percentile ratios are 3.1 and 1.8 respectively.

Figure 2: Farm-specific Distortions and Productivity



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

TFP, estimated as the fixed effect component from a panel regression. 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 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). Allowing for input reallocation across villages, i.e., without removing the village specific effects, 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

Table 1: 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	–
+ across villages	53.2	24.9	–	83.0

Notes: Reports the 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.

considering also the possibility of resource reallocation across villages, with about 2/3 accounted for by household mobility across space. This is not surprising in view of the results in Ngai et al. (2019) showing that China's *hukou* system—the effective limited ability to trade land for cultivation and social transfers associated with education and health tied to local registration—constitute substantial barriers to population movements across sectors and space. In particular, Ngai et al. (2019) argue that the land policy component of the *hukou* system is a key barrier to industrialization.

We also emphasize that if we consider instead the more conventional cross-sectional measures of farm TFP and distortions typically used in the literature, the reallocation gains more than triple to 83 percent relative to the gains associated with our estimated household fixed effect measures, highlighting the importance of the panel data in dealing with measurement issues, which we discuss further in the next subsection.

Note that in general it is difficult to identify the sources of dispersion in marginal products when measuring input wedges indirectly, as evidenced by inspecting equations (C.4) and (C.5) in Appendix C. 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. We perform this decomposition using equation (C.8) in Appendix C. 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 1). We focus on the TFPR variation as an output wedge associated with the land institution in our quantitative analysis in Section 5.

The reallocation gains are calculated relative to an efficient benchmark, where the average (and marginal) product of factors are hypothetically equalized across farms. 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 around 80 percent between farm size and several alternative measures of farm TFP. We note that this high correlation contrasts sharply with the land allocation in China documented earlier where there is essentially no correlation between land input and farm productivity. This finding is consistent with recent empirical evidence emphasizing farm TFP as a key measure of productivity. 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 productivity growth (Adamopoulos and Restuccia, 2014). We conclude from this evidence that the efficient benchmark is useful in contrasting the land allocation in China and providing reasonable and attainable gains from reallocation.

While there are no explicit prohibitions of rentals in China, frequent reallocations of land within villages likely lead households to fear loss of their use rights if they do not farm their land themselves. Indeed, rental markets are very thin in China during our sample period, constituting less than 5 percent of cultivated land. Moreover, there is no significant change in the amount of rented land over time. In addition, rentals of land typically involve family members or close relatives and hence do not necessarily direct the land to best uses. Nevertheless, we can assess the extent to which rented land alleviates misallocation. Controlling for farm TFP, we find that the reallocation gains among farms with no rented land are roughly 20 percent larger than among farms with some rented land. This finding suggests that rentals help reduce misallocation, but their scale is too limited to prevent large productivity losses due to misallocation.

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 0.90 across the years, and as low as 0.85 in 1993, as high as 0.94 in 1997, and 0.90 in 2002. This finding contrasts with the reduction in misallocation found for the manufacturing sector in China in Hsieh and Klenow (2009) over a similar period. Similarly, the reallocation gains associated with this misallocation appear roughly constant over time. These 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.

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. Our analysis is less subject to these critiques for two reasons. First, an important factor in our analysis mitigating measurement concerns is the fact that we focus on the household farm as the production unit rather than the individual plots operated by the household. In this context, we expect a lesser role of unobserved transitory shocks and measurement error in our data compared to studies using for example plant or establishment level data. Second, we exploit the panel structure of the data by estimating farm productivity and farm distortions as the household fixed effect of a panel regression with controls as discussed earlier. We illustrate the value of our approach by explicitly applying the state-of-the-art method in Bils et al. (2018), 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. (2018) 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 with higher TFPR display larger output responses to input changes? Bils et al. (2018) show that this question can be addressed by regressing plant growth in measured output on plant growth in measured inputs and the interaction of the growth in inputs and the level of measured TFPR. We implement the Bils et al. (2018) methodology using our panel data on farms for China. In particular, the econometric model we estimate is the following,

$$\Delta \log (y_{it}) = \beta_1 \cdot \log (TFPR_{it}) + \beta_2 \cdot \Delta \log (I_{it}) + \beta_3 \cdot interaction_{it} + \mu_v + \mu_t + u_{it}, \qquad (6)$$

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 μ_v and μ_t are village and time fixed effects. Bils et al. (2018) show that from the regression coefficients in equation (6) we can identify an estimate of the share of the dispersion in TFPR that is due to true variation in distortions, λ as: $\widehat{\lambda} = 1 + \widehat{\beta}_3/\widehat{\beta}_2$. A λ close to 1 reflects minimal measurement error, implying that measured differences across farms in TFPR reflect largely differences in true distortions. Also the smaller λ the lower the productivity gains from reallocation than those implied by measured TFPR.

We estimate equation (6) 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; (b) a fixed-effect TFPR measure but without controlling for village factors (other than land quality); (c) a raw unadjusted cros-sectional measure of TFPR, controlling only for village-level land quality. The estimated λ in each case is reported in the last row of Table 2. When we use a raw unadjusted measure of TFPR we estimate $\widehat{\lambda} = 0.914$, which implies that measurement error is 8.6%. When we use a fixed effect estimate of TFPR from which village factors have not been purged we find $\widehat{\lambda} = 0.95$, implying 5% measurement error. When we use our baseline measure of permanent TFPR the estimated λ becomes 1, implying that there is little scope for the type of measurement error this method can capture with our baseline measure. Consider that in the analysis of Bils et al. (2018) for manufacturing plants in the United States and India, measurement error was found to be substantial. In particular, the estimated λ was stable over time around 0.5 for India, whereas for the United States λ is 0.86 at the beginning of the sample but falls to around 0.5 at the end of the sample. We note that the method applied is suitable for assessing the extent of additive measurement error that is orthogonal to true farm productivity.

Table 2: Mismeasurement in Productivity and Distortions

	Fixed Effect Estimates		Cross-section
	Household farm	+ Village	average
Farm TFP:			
STD(log)	0.35	0.64	0.71
p90/p10	2.18	4.35	5.53
p75/p25	1.48	2.06	2.32
Farm TFPR:			
STD(log)	0.48	0.81	0.90
p90/p10	3.14	7.17	9.48
p75/p25	1.78	2.71	3.15
CORR (logTFP, logTFPR)	0.91	0.88	0.88
BKR $\hat{\lambda}$	1.00	0.95	0.91

Notes: Fixed effects estimates from panel regression. Cross-section average refers to statistics computed in each year and averaged across years. Farm TFP and TFPR are computed as specified in the text. BKR $\hat{\lambda}$ is the estimate of the fraction of TFPR variation that is not measurement error as in Bils et al. (2018). Each BKR $\hat{\lambda}$ is based on an OLS regression of (6) with village and year dummies, and clustering of standard errors at the village level. Extreme values of TFPR are winsorized at 0.5 percent tails.

We note also in Table 2 that mismeasurement has important implications for the overall dispersion in productivity and distortions, as we move from cross-sectional measures of farm TFP and distortions to fixed-effects measures. The dispersion in productivity and distortions are substantially reduced as evidenced by the standard deviation of the log, or the 90 to 10 and 75 to 25 ratios. For instance, 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, and a further 40 percent reduction when measured as only the household fixed effect. Mismeasurement has less of an impact on the systematic component of distortions as measured by the correlation between farm productivity and distortions. The correlation is strengthened somewhat, from 0.88 to 0.91, when removing transitory variation or measurement error, again consistent with our description of the land institution in China, where within a village, land is allocated fairly uniformly across households regardless of their farming ability. The systematic component of distortions is a key element in the amplification effect via selection we discuss in the next section.

5 Misallocation and Selection across Sectors

We now integrate our framework of agriculture into a two-sector general-equilibrium Roy (1951) model of selection across sectors to assess: (1) how farm-specific distortions in agriculture alter the occupational choice of individuals between agriculture and non-agriculture; and (2) how selection affects measured misallocation in the agricultural sector and the productivity gains from factor reallocation.

We augment our village model of agriculture along the following dimensions. First, we extend the model to a two-sector model by introducing a non-agricultural sector. Second, we consider preferences for individuals over consumption goods for agriculture and non-agriculture, with a subsistence constraint for the agricultural good. Third, individuals are endowed with a pair of productivities, one for each of the two sectors. Fourth, individuals make an occupational choice of whether to become farm operators in agriculture or workers in the non-agricultural sector. We show that a key determinant of the occupational choice is the farm-specific distortion individuals face if they become farm operators. For analytical tractability, we consider a continuum of individuals. The fraction of individuals that choose agriculture, and thus the number and productivity distribution of farms are endogenous. In what follows, we discuss in detail the economic environment and define the equilibrium, and then describe some key properties of the model.

5.1 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,

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

where c_a and c_n denote the consumption of the agricultural and non-agricultural good, $\overline{a}$ is a minimum subsistence requirement for the agricultural good, and ω is the preference weight on agricultural goods. The subsistence constraint implies that when income is low a disproportionate amount will be 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 φ_i, which we define more fully below. Income from working in the non-agricultural sector is subject to a tax η, common to all individuals. η 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 φ_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}, and 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,

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

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. As described previously, 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 used previously,

$$y_{ai} = \left(A_a s_{ai}\right)^{1-\gamma}\left(\ell_i^{\alpha} k_i^{1-\alpha}\right)^{\gamma}, \tag{7}$$

where y_a is real farm output, ℓ 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 τ_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. 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,

$$\max_{\ell_i, k_i} \left\{ \pi_i = p_a \left(1-\tau_i\right) y_{ai} - r k_i - q \ell_i \right\}, \tag{8}$$

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 the farm-specific distortions,

$$\ell_i = A_a \left(\gamma p_a\right)^{\frac{1}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}} \left(1-\tau_i\right)^{\frac{1}{1-\gamma}} s_{ai}, \tag{9}$$

$$k_i = A_a \left(\gamma p_a\right)^{\frac{1}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha\gamma}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(1-\tau_i\right)^{\frac{1}{1-\gamma}} s_{ai}, \tag{10}$$

$$y_{ai} = A_a \left(\gamma p_a\right)^{\frac{\gamma}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(1-\tau_i\right)^{\frac{\gamma}{1-\gamma}} s_{ai}, \tag{11}$$

$$\pi_i = A_a \left(1-\gamma\right) p_a^{\frac{1}{1-\gamma}} \gamma^{\frac{\gamma}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(1-\tau_i\right)^{\frac{1}{1-\gamma}} s_{ai}. \tag{12}$$

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

$$I_{ai} = w_a \varphi_i s_{ai}, \tag{13}$$

where $\varphi_i \equiv \left(1-\tau_i\right)^{\frac{1}{1-\gamma}}$ 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,

$$w_a \equiv p_a^{\frac{1}{1-\gamma}} A_a \gamma^{\frac{\gamma}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}}. \tag{14}$$

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 can become farm operators in the agricultural sector or workers in the non agricultural sector. If individual i chooses to become a farm operator in agriculture their income is given by (13), while if they become a non-agricultural worker their income is,

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

We note that these incomes are net of the transfer T, which is common to all individuals and hence does not affect occupational choices. Individual i will choose the sector that provides the highest possible income, given the individual's triplet $(s_{ai}, \varphi_i, s_{ni})$. Individual i will choose 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 φ_i. We can define an individual's effective ability as the product of the two, $\widehat{s}_{ai} \equiv s_{ai}\varphi_i$. An individual will then choose to operate a farm if $w_a\widehat{s}_{ai} \geq (1-\eta)\, w_n s_{ni}$. We note that, holding relative prices constant, farm-specific taxes directly distort the occupational choices of individuals. For given common sectoral returns (w_a, w_n), barrier η, and individual abilities (s_{ai}, s_{ni}), a lower φ (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.

5.2 Analysis

The model laid out above has implications for the share of employment in agriculture, the occupational choices of individuals, the pattern of selection, and sectoral and aggregate productivity. In addition to aggregate implications, the model has micro-level implications summarized by moments of sectoral incomes conditional on sectoral choices, as well as by moments of farm-level productivity and farm-level distortions for those operating in agriculture. We exploit the properties of the multivariate log-normal distribution over $(s_{ai}, \varphi_i, s_{ni})$ in order to provide analytical results.

In Appendix D 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,

$$N_a = \Phi\left(b\right), \tag{15}$$

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

$$b \equiv \frac{b_a - b_n}{\sigma}, \quad b_a \equiv \log\left(w_a\right) + \mu_\varphi + \mu_a, \quad b_n \equiv \log\left(w_n\right) + \log\left(1 - \eta\right) + \mu_n, \tag{16}$$

where σ is the variance of relative effective abilities between non-agriculture and agriculture.

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,

$$E\left\{\log\left(\widehat{s}_{ai}\right)|i \in H_a\right\} = \widehat{\mu}_a + \frac{\sigma_{an} - \widehat{\sigma}_a^2}{\sigma}\lambda^l\left(b\right), \tag{17}$$

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

$$E\left\{\log\left(s_{ni}\right)|i \in H_n\right\} = \mu_n + \frac{\sigma_n^2 - \sigma_{an}}{\sigma}\lambda^u\left(b\right), \tag{18}$$

where $\widehat{\mu}_a = \mu_a + \mu_\varphi$. $\lambda^l\left(b\right) \equiv E\left[\xi|\xi \leq b\right] < 0$ and $\lambda^u\left(b\right) \equiv E\left[\xi|\xi > b\right] > 0$ represent lower tail truncation and upper tail truncation of a standard normal random variable ξ. The coefficients in (17) and (18) can be re-written as $\frac{\widehat{\sigma}_a\sigma_n}{\sigma}\left[\rho_{an} - \frac{\widehat{\sigma}_a}{\sigma_n}\right]$ and $\frac{\widehat{\sigma}_a\sigma_n}{\sigma}\left[\frac{\sigma_n}{\widehat{\sigma}_a} - \rho_{an}\right]$. 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 $\widehat{\sigma}_a$ and non-agriculture σ_n, and their correlation ρ_{an}. For example, if effective abilities are sufficiently positively correlated across sectors and the dispersion of non-agricultural ability is larger in relative terms ($\widehat{\sigma}_a^2 < \sigma_{an}$ and $\sigma_n^2 > \sigma_{an}$), then the average effective ability of those in agriculture (non-agriculture) is lower (higher) than the population average.

6 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 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 now describe these steps in detail.

6.1 Inferring Population Moments from Observed 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. Normalizing the means of these distributions to zero, there are five population moments that need to be calibrated: three variances, σ_a^2, σ_n^2, σ_φ^2 and two covariances, $\sigma_{a\varphi}$, σ_{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,

$$VAR\left\{\log\left(I_{ai}\right)|i\in H_a\right\}\equiv\widehat{v}_a=\widehat{\sigma}_a^2\left\{1-\left(\frac{\sigma_{an}-\widehat{\sigma}_a^2}{\sigma\widehat{\sigma}_a}\right)^2\lambda^l\left(b\right)\left[\lambda^l\left(b\right)-b\right]\right\}. \tag{19}$$

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

$$VAR\left\{\log\left(I_{ni}\right)|i\in H_n\right\}\equiv\widehat{v}_n=\sigma_n^2\left\{1-\left(\frac{\sigma_n^2-\sigma_{an}}{\sigma\sigma_n}\right)^2\lambda^u\left(b\right)\left[\lambda^u\left(b\right)-b\right]\right\}. \tag{20}$$

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

$$COV\left\{\log\left(I_{ai}\right),log\left(I_{ni}\right)|i\in H_a\right\}\equiv\widehat{c}_{an}=\sigma_{an}-\left(\frac{\sigma_n^2-\sigma_{an}}{\sigma}\right)\left(\frac{\sigma_{an}-\widehat{\sigma}_a^2}{\sigma}\right)\lambda^l\left(b\right)\left[\lambda^l\left(b\right)-b\right]. \tag{21}$$

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

$$VAR\left\{\log\left(\varphi_i\right)|i\in H_a\right\}\equiv\widehat{v}_\varphi=\sigma_\varphi^2\left\{1-\left(\frac{\sigma_\varphi^2+\sigma_{a\varphi}}{\sigma\sigma_\varphi}\right)^2\lambda^l\left(b\right)\left[\lambda^l\left(b\right)-b\right]\right\}. \tag{22}$$

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

$$COV\left\{\log\left(I_{ai}\right),\log\left(\varphi_i\right)|i\in H_a\right\}\equiv\widehat{c}_{I_a,\varphi}=$$
$$\left(\sigma_{a\varphi}+\sigma_\varphi^2\right)+\left(\frac{\sigma_{an}-\widehat{\sigma}_a^2}{\sigma}\right)\left(\frac{\sigma_\varphi^2+\sigma_{a\varphi}}{\sigma}\right)\lambda^l\left(b\right)\left[\lambda^l\left(b\right)-b\right]. \tag{23}$$

In these expressions, the relationship between the variance of ability and the variance of effective ability in agriculture is given by,

$$\widehat{\sigma}_a^2=\sigma_a^2+\sigma_\varphi^2+2\sigma_{a\varphi}, \tag{24}$$

and the measures of sectoral incomes and distortions were discussed previously.

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. This procedure is outlined in Appendix E. 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 0.034 in our micro-data.

We emphasize that this moment is very robust to alternative classifications of workers as being in "non-agriculture" and switching from agriculture to non-agriculture over the sample period. We consider alternative definitions of switchers by classifying households in non-agriculture as those cultivating zero land and producing zero agricultural output and those who self report as mainly in non-agriculture or full-time in non-agriculture based on time allocations. Then non-switchers are those households that are classified in agriculture or non-agriculture during the entire panel, whereas switchers are the rest. For instance, the correlation between log income in agriculture and non-agriculture for switchers according to the no-output-land definition is 0.036 and 0.020 according to the two self-reported time allocation definitions, compared to 0.034 in the contemporaneous correlation.

Table 3: Targeted Empirical Conditional Moments

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

Notes: All variables refer to logs.

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 (19)-(23). Table 3 contains the empirical conditional variances and covariances along with the share of employment in agriculture that we target. The details of how we use the targeted empirical moments with the above system of equations to infer the population moments are provided in Appendix F. Our procedure ensures that the occupational choices of individuals are consistent with the observed share of employment in agriculture of 46 percent in China.

In Table 4 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 ρ_{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 will sort into the sector in which they possess a comparative advantage, and that individuals working in each sector will be on average more skilled in that sector than the general population. The correlation of ability in agriculture s_a and distortions φ is strongly negative (-0.95), consistent with our description of the institutional environment in China.

Table 4: Calibrated Population Moments

Parameter	Description	Value
σ_a	STD of agricultural ability	1.30
σ_n	STD of non-agricultural ability	0.65
σ_φ	STD of distortions	1.06
$\rho_{a\varphi}$	CORR of agricultural ability and distortions	-0.95
ρ_{an}	CORR of agricultural–non-agricultural ability	-0.15

Notes: All variables refer to logs.

6.2 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, φ, s_n), drawn from a multivariate log-normal distribution, using the inferred population moments from the previous step.[12]

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; (α,γ) 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; 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 ω, 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 determined by the parameter ω 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

[12] 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.

"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 η; 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 5 displays the aggregate and micro-level statistics for the benchmark economy, as well as the values for the calibrated parameters. The model reproduces well the macro and micro statistics for China, in particular, the log-normal assumption for the distributions of distortions and abilities that afford us substantial tractability, provides a good fit of the empirical distributions. Note that whereas the static gain from efficient reallocation in the data is 24.4 percent, it is only 15 percent in the model. The difference arises mainly because we focus on the reallocation gains associated with the land institution which we model as an output wedge as discussed in Section 4.3.

7 Quantitative Experiments

We conduct a set of counterfactual experiments in order to assess the quantitative importance of farm-specific distortions for the allocation of resources within agriculture, the sector-occupation choices of individuals, as well as the amplification effect that selection imparts on sectoral productivity and real GDP per worker. We consider in turn: (a) the effects of eliminating farm-specific distortions; (b) these effects compared to those from an exogenous increase in TFP; and (c) the pattern of selection with sectoral reallocation.

Table 5: 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	3.96
Z_a/N_a	Average ability in agriculture	3.42
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	Static gain of misallocation	1.15
	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
η	labor mobility barrier	0.74

7.1 Eliminating Distortions

Our main experiment involves studying the effects from removing farm-specific distortions in agriculture. In Section 4 we showed that not only is there a large dispersion in implicit distortions across farms but that they are also strongly positively correlated with farm productivity. In the data, more productive farmers face larger distortions as they are unable to operate larger farms (i.e. obtain more land and capital for their operation). Land market institutions that restrict the allocation of land within villages are at the heart of these patterns. In order to assess the aggregate and micro-level effects of distortions we consider two cases: first, we eliminate all distortions on farmers in the economy, i.e., we set $\varphi = 1$ for all i; and second, we eliminate only the correlation of distortions with farmer ability, keeping uncorrelated distortions the same as in the benchmark. In this case, we set $\rho_{a\varphi} = 0$ and since we showed that 60 percent of the static reallocation gains is due to the correlation with farmers ability, we also reduce σ_φ so that the remaining static reallocation gains is 40 percent of the overall static gains in the Benchmark Economy, implying $\sigma_\varphi = 0.66 \times \sigma_\varphi^{BE}$. The results from these counterfactuals highlight the importance of distortions for aggregate, sectoral, and micro outcomes as well as the contribution of the systematic component of distortions for these patterns. The results of these two counterfactuals along with the benchmark economy are presented in Table 6.

Table 6: Counterfactuals: 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	2.39	3.42
Share of Employment in Agriculture (N_a) (%)	0.46	0.20	0.14
TFP in Agriculture (TFP_a)	1.00	1.50	1.80
Real Non-Agricultural Productivity (Y_n/N_n)	1.00	0.79	0.77
Average Ability in Agriculture (Z_a/N_a)	1.00	2.03	2.65
Average Ability in Non-Agriculture (Z_n/N_n)	1.00	0.79	0.77
Real GDP per Worker (Y/N)	1.00	1.14	1.19
Micro-level Statistics			
STD of log–farm TFP	0.56	0.41	0.34
STD of log–farm TFPR	0.48	0.30	0
CORR of log–(farm TFP, farm TFPR)	0.97	0.42	–
CORR of log–(agr. ability, non-agr. ability)	0.16	0.28	0.51
CORR of log–(agr. income, non-agr. income)	0.04	0.44	0.51

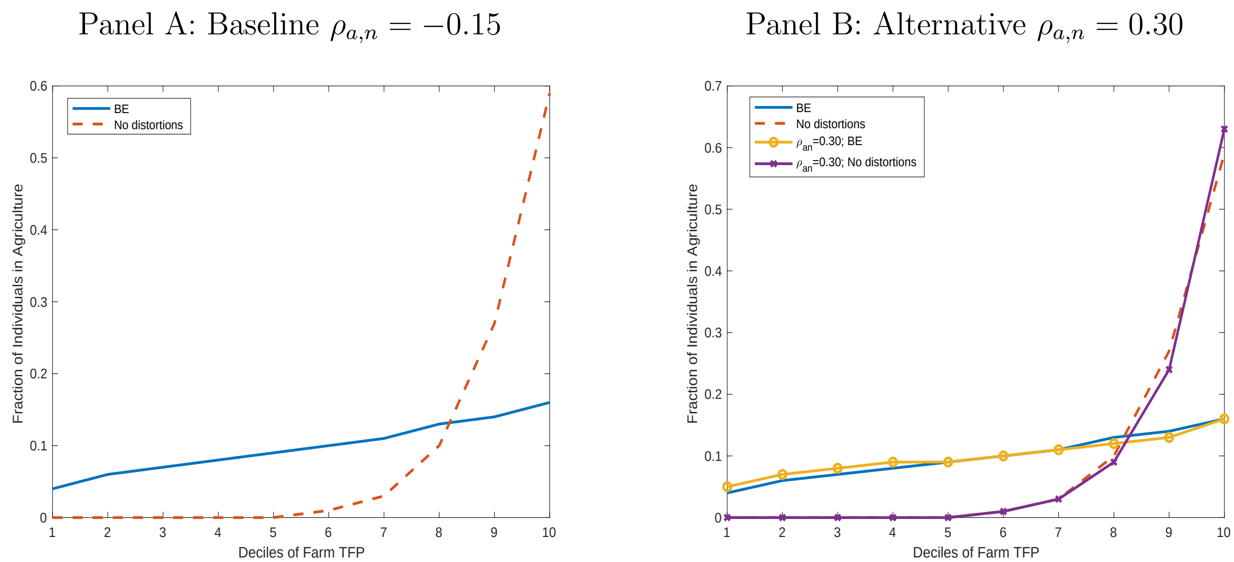
Notes: The counterfactual "No Correlated Distortions" refers to the economy when only eliminating the correlation of farm-level distortions with farmer ability. In this case, we set $\sigma_{a\varphi} = 0$ and since the correlation accounts for 60 percent the static reallocation gains, we set $\sigma_\varphi = 0.66 \times \sigma_\varphi^{BE}$ to match the static reallocation gains of 1.06. The counterfactual "No Distortions" eliminates all farm-level distortions, correlated and uncorrelated. In this case, 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.

Eliminating all distortions. As can be seen from the third column in Table 6, eliminating all distortions on farmers 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. As a reference for comparison, consider that the "static" gains of the efficient reallocation from eliminating distortions in the model involve an increase in agricultural TFP of 15 percent while the total increase in agricultural TFP is a factor of 1.80-fold. To understand the relationship between the increases in agricultural labor productivity and TFP, we use aggregate agricultural output in equation (C.7) to express agricultural labor productivity as,

$$\underbrace{\frac{Y_a}{N_a}}_{3.42} = \underbrace{A}_{1.15} \cdot \underbrace{\bar{Z}_a^{\,1-\gamma}}_{1.57} \cdot \underbrace{\left[\frac{L^{\alpha}K^{1-\alpha}}{N_a}\right]^{\gamma}}_{1.90}, \tag{25}$$

where A includes both $A_a^{1-\gamma}$ and the effect of idiosyncratic distortions on TFP (see equation C.8). Since A_a is constant, changes in A reflect the static efficiency gains from eliminating misallocation. The term $\bar{Z}_a^{\,1-\gamma}$ captures the effects of selection. Note that these first two terms represent the components of agricultural TFP that are affected by misallocation. The last term in (25) represents the impact of the share of employment in agriculture on labor productivity holding fixed aggregate endowments of land and capital. Eliminating farm-specific distortions increases agricultural labor productivity directly by eliminating misallocation across farms, but it also induces higher ability farmers to sort back into agriculture since they no longer face restrictions on consolidating land. These two effects constitute the increase in agricultural TFP. At the same time, the associated increases in productivity via eliminating 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. To appreciate the effect of eliminating distortions on selection, Figure 3, Panel A, reports the fraction of employment in agriculture for each decile of the farming ability distribution. Whereas in the benchmark economy with distortions employment in agriculture is uniformly distributed across all agricultural ability types, removing distortions

Figure 3: Selection Effect in Agriculture



Notes: Reports the fraction of employment in agriculture from each decile of the farming ability distribution in the benchmark economy and the economy with no distortions. The ratio of farm TFP between the top and bottom deciles is a factor of 8.2-fold.

improves selection into agriculture of mostly high ability farmers.

Eliminating distortions is the driver of all these changes, but to gauge the magnitude of each, we use equation (25) to decompose the change in agricultural labor productivity into: (1) the misallocation effect on productivity which results in an increase of 1.15-fold or about 11 percent of the overall increase (log(1.15)/log(3.42)); (2) the effect of a 2.7-fold increase in average ability in agriculture as the more able farmers find it optimal to stay in agriculture, which translates into a 1.57 fold increase in agricultural labor productivity or 36 percent of the overall increase; and (3) the effect of the reduction in the share of employment in agriculture N_a, which implies a further increase in agricultural labor productivity of 1.90-fold or 53 percent of the overall increase. This is how a 1.15 fold increase in agricultural TFP due to reduced misallocation translates into a 3.42-fold increase in agricultural labor productivity. In terms of TFP in agriculture, the overall increase is a factor of 1.80-fold, out of which more than two-thirds is the effect of selection (log(1.57)/log(1.80)). In other words, the amplification effect of selection on agricultural TFP more than triples the static gains from reduced misallocation.

Average ability in non-agriculture, and as a result labor productivity in non-agriculture, falls to 77 percent of their benchmark economy values, due to the mass influx of workers, not all of which are as productive in non-agriculture. Despite the significant impact on agricultural productivity, real GDP per worker increases about 19 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.[13] The dispersion of TFP in agriculture is now lower, albeit with a higher mean.

Eliminating correlated distortions. Consider next the case of eliminating only the correlation of distortions. Note that in this experiment there is still misallocation related to the dispersion of distortions. As discussed above, in the data there are two types of misallocation: misallocation of factors across farmers with different productivity, as well as misallocation of factors among farmers with the same productivity. Misallocation across farmers with different productivity accounts for 60 percent of the static reallocation gains. In this experiment, we only eliminate the distortions across farmers with different productivity so we set $\rho_{a\varphi} = 0$ and $\sigma_\varphi = 0.66 \times \sigma_\varphi^{BE}$. The results of this quantitative experiment are summarized in the second column of Table 6. The results are in the same direction and almost of the same magnitude as when we eliminate all distortions. For instance, agricultural labor productivity increases by almost 2.4-fold versus 3.4-fold when eliminating all distortions, implying that the strong correlation between distortions and farmer ability accounts for 71 percent ($\log(2.39)/\log(3.42)$) of the change in agricultural labor productivity from eliminating all distortions. A similar contribution of 71 percent applies to agricultural TFP. Whereas correlated distortions account for 60 percent of the static gains from reduced factor misallocation, they account for a larger share of the productivity gains with selection, suggesting that the correlation property of distortions is quite important in the amplification mechanism of selection.

[13]In reality, however, non-agricultural labor productivity increases by 84 percent over the period we examine. Taking this into account, the effect on real GDP per capita from eliminating distortions is more pronounced, with aggregate real GDP per capita increasing 2.2-fold, compared with 1.19-fold with just reduced misallocation in agriculture.

Discussion. We noted earlier that the land policy in China may be responsible for the sizeable misallocation of inputs observed across villages, which can potentially double the efficiency gains estimated from reallocation across households within villages. In this context, the productivity implications described from reforming the village land institutions in the model represent a lower bound of the potential impact on productivity, as factor mobility across Chinese villages can substantially enlarge those gains. Substantial labor mobility across space has indeed been found in reform episodes delinking land use from land rights (De Janvry et al., 2015).

The aggregate quantitative impact from eliminating farm-level distortions depends on two key empirical moments: the dispersion of farm-level distortions, which largely determines the extent of static misallocation in the model, and the correlation of distortions with farm productivity, which as just shown largely determines the amplification effect through selection. We have calibrated these empirical moments to estimates of household fixed effects in panel regressions in China. As mentioned earlier, typical cross-sectional analyses of misallocation are often criticized for misattributing measurement error in the quality of inputs and outputs and idiosyncratic transitory shocks to misallocation. As expected both the panel dimension and the narrower geographical characterization of the data reduce the extent of misallocation as evidenced in the lower TFPR dispersion. But as shown earlier, the amplification effect of selection hinges largely on the systematic pattern of distortions with respect to farm TFP and this component remains strong, if not even stronger, in the panel and within villages (see again Table 2). This result is to be expected under the interpretation of the implicit distortions in the agricultural sector as stemming from the land institution in China, which roughly provides equal amounts of land to farmers of very different productivities.

We also discuss our results with an alternative value for the population correlation of abilities across sectors ρ_{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 ρ_{an} is -0.15. We consider an alternative value of 0.3 and re-calibrate the benchmark economy to assess the impact of removing distortions. Table 7 reports the results. Removing distortions has a substantial effect on aggregate agricultural productivity, 3.4-fold in the baseline calibration and 3.7-fold in the alternative calibration. This is largely due to a stronger effect of selection on agricultural TFP which increases by 1.9-fold compared to 1.8-fold in the baseline and the consequent larger decline in agricultural labor. The stronger selection effect in agriculture can be observed in Figure 3, Panel B, where the positive correlation between distortions and abilities generates a slightly stronger bias towards low productive farmers selecting into agriculture.

Table 7: Eliminating Distortions with Alternative Population ρ_{an}

Statistic	Benchmark Economy	No distortions: Baseline $\rho_{an} = -0.15$	No distortions: Alternative $\rho_{an} = 0.30$
Aggregate Statistics			
Real Agricultural Productivity (Y_a/N_a)	1.00	3.42	3.74
Share of Employment in Agriculture (N_a) (%)	0.46	0.14	0.11
TFP in Agriculture (TFP_a)	1.00	1.80	1.87
Real Non-Agricultural Productivity (Y_n/N_n)	1.00	0.77	0.80
Average Ability in Agriculture (Z_a/N_a)	1.00	2.65	3.04
Average Ability in Non-Agriculture (Z_n/N_n)	1.00	0.77	0.80
Real GDP per Worker	1.00	1.19	1.15
Micro-level Statistics			
STD of log–farm TFP	0.56	0.34	0.36
STD of log–farm TFPR	0.48	0	0
CORR of log–(farm TFP, farm TFPR)	0.97	–	–
CORR of log–(agr. ability, non-agr. ability)	0.16	0.51	0.75
CORR of log–(agr. income, non-agr. income)	0.04	0.51	0.75

Notes: "Baseline" is the main calibration of data moments in China that results in a population correlation of abilities across sectors $\rho_{an} = -0.15$. For $\rho_{an} = 0.30$, the benchmark economy is calibrated to the same targets in the second stage and distortions are removed. 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.

To summarize, our results suggest that farm-specific distortions have an important effect on the occupational choices of farmers, particularly high productivity farmers, which in turn substantially affect agricultural productivity, and the allocation of labor across sectors.

7.2 Comparison to an Exogenous Increase in TFP

In the context of our model, improvements in resource allocation in agriculture produce an increase in aggregate 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 15 percent exogenous increase in TFP, corresponding to the static gains from eliminating misallocation in our model.

Table 8: Comparison of Removing Distortions vs. Exogenous TFP Increases

Statistic	BE	No Distortions	$\uparrow (A_a^{1-\gamma}) \times 1.15$	$\uparrow (A_a^{1-\gamma}, A_n) \times 1.15$
Aggregate Statistics				
Real Agricultural Productivity (Y_a/N_a)	1.00	3.42	1.54	1.54
Share of Employment in Agriculture (N_a) (%)	0.46	0.14	0.30	0.30
TFP in Agriculture (TFP_a)	1.00	1.80	1.23	1.23
Real Non-Agricultural Productivity (Y_n/N_n)	1.00	0.77	0.89	1.02
Average Ability in Agriculture (Z_a/N_a)	1.00	2.65	1.16	1.16
Average Ability in Non-Agriculture (Z_n/N_n)	1.00	0.77	0.89	0.89
Real GDP per Worker (Y/N)	1.00	1.19	1.12	1.27
Micro-level Statistics				
STD of log– farm TFP	0.56	0.34	0.56	0.56
STD of log– farm TFPR	0.48	0	0.48	0.48
CORR of log–(farm TFP, farm TFPR)	0.97	–	0.97	0.97
CORR of log– (agr. ability, non-agr. ability)	0.16	0.51	0.22	0.22
CORR of log–(agr. income, non-agr. income)	0.04	0.51	0.16	0.16

Notes: The first column "BE" refers to the benchmark economy. The second column "No Distortions" refers to the counterfactual of eliminating all farm-level distortions. The third column refers to the case of exogenously increasing TFP in agriculture 1.15-fold relative to the benchmark, and the fourth column refers to the case of increasing TFP in both agriculture and non-agriculture 1.15-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.

In the first two columns of Table 8 we reproduce the results for the benchmark economy and the economy without 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 15 percent (keeping 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 30 percent, however, agricultural TFP increases only by 23 percent, compared to 80 percent when eliminating distortions. Agricultural TFP increases slightly more than the exogenous increase in TFP because there is only a small effect on selection into agriculture, increasing average quality of workers in agriculture by 16 percent, compared to 165 precent when eliminating distortions. The effects are similar when non-agricultural TFP also increases exogenously by 15 percent (column four).

The reduction in misallocation associated with the elimination of farm-level distortions has a much larger aggregate 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 15 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 static misallocation gains of 15 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.

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.

7.3 Sectoral Reallocation and Selection Patterns

Our empirical findings indicate that farm-level distortions have not changed much over the period we examine, consistent with the unchanged land-market institutions in China. Yet in the data we observe that households are moving out of agriculture and into non-agriculture at a rate of about 1 percent per year. This occurs because the economy grows even with farm-level distortions. Non-agricultural labor productivity increases 7 percent per year over the period we examine, with an overall increase of 84 percent over 1993-2002. In Table 9 we show what happens in the benchmark economy in the presence of the observed farm-level distortions when TFP increases in both agriculture and non-agriculture. The second column displays the results under year-to-year (annual) exogenous changes in TFP, whereby TFP in agriculture increases by 2.2 percent (to capture the one percentage point drop in the share of employment in agriculture), and TFP in non-agriculture increases by 7 percent, relative to the benchmark. The third column captures the overall changes over the sample period 1993-2002, an implied exogenous increase in agriculture TFP of 24.2 percent and non-agricultural TFP of 84 percent, relative to the benchmark. As productivity increases in both sectors, the share of employment in agriculture drops by an amount consistent with the drop we observe in our survey data.

One question that arises is whether there is a particular type of household that switches from agriculture to non-agriculture as productivity rises in the presence of distortions. We examine

Table 9: Effects of TFP Increases on Selection

Statistic	BE	$\uparrow A_a^{1-\gamma}$, A_n Annual	$\uparrow A_a^{1-\gamma}$, A_n Overall
		Aggregate Statistics	
Real Agricultural Productivity (Y_a/N_a)	1.00	1.03	1.37
Share of Employment in Agriculture (N_a) (%)	0.46	0.45	0.34
TFP in Agriculture (TFP_a)	1.00	1.02	1.16
Real Non-Agricultural Productivity (Y_n/N_n)	1.00	1.06	1.68
Average Ability in Agriculture (Z_a/N_a)	1.00	1.01	1.11
Average Ability in Non-Agriculture (Z_n/N_n)	1.00	0.99	0.91
Real GDP per Worker (Y/N)	1.00	1.07	1.87
		Micro-level Statistics	
STD of log–farm TFP	0.56	0.56	0.55
STD of log–farm TFPR	0.48	0.48	0.48
CORR of log–(farm TFP, farm TFPR)	0.97	0.97	0.97
CORR of log–(agr. ability, non-agr. ability)	0.16	0.16	0.20
CORR of log–(agr. income, non-agr. income)	0.04	0.05	0.13

Notes: The first column "BE" refers to the benchmark economy. The second column refers to the case of increasing TFP in agriculture by 2.2 percent and TFP in non-agriculture by 7 percent (annual), relative to the benchmark. In the third column, TFP in agriculture increases 24.2 percent and non-agricultural TFP by 84 percent (overall) 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 question in the context of our model by considering the micro-level implications of exogenous increases in TFP, and in particular the implications for the correlation of abilities and the correlation of incomes across sectors as the share of individuals in agriculture drops. The results of Table 9 indicate that for year-to-year changes (second column) the correlation of agricultural and non-agricultural abilities or incomes across sectors exhibits hardly any change relative to the benchmark. Over the entire period (third column), these correlations change more, particularly the correlation of incomes. These changes however are well within the range of changes observed in the data. These results suggest that the bulk of changes in sectoral reallocation in China during the sample period are driven by exogenous changes in productivity and not so much by changes in the agricultural distortions associated with the land market institutions and the "hukou" system.

8 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 us with information on income in agriculture and wages in non-agriculture for households that switch occupations, which we use to restrict the correlation of abilities across sectors in the population. 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 aggregate 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 will not only allow farmers to increase their operational scales through land consolidations, but would 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. (2018). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. Technical report, National Bureau of Economic Research.

Adamopoulos, T. and Restuccia, D. (2019). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics*.

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. (2018). Misallocation or mismeasurement? *manuscript, Stanford University*.

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.

Chen, C. (2017). Untitled land, occupational choice, and agricultural productivity. *American Economic Journal: Macroeconomics*, 9(4):91–121.

Chow, G. (2002). Capital formation and economic growth in china. *Quarterly Journal of Economics*, 108.

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 Grobovsek, J. (2015). Communal Land and Agricultural Productivity,. Unpublished manuscript, University of Edinburgh.

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 Santaeulalia-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.

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.

Appendix — Not for publication

A Measures of Output and Inputs

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

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

and the resource constraints,

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

Using the first-order conditions of this problem along with the rural village resource constraints in equation (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 ℓ 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,

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

where Y^e is agricultural output under the efficient allocation, A^e is agricultural TFP $A^e = \left(A_a \bar{S}\right)^{1-\gamma}$, where $\bar{S} = \left(\sum_{i=1}^{M} s_i\right)/M$ is average farm productivity.

C Equilibrium and Identification of Distortions

Denote by τ_i^{ℓ} and τ_i^{k} the land and capital input taxes, and by τ_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,

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

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:

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

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

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,

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

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,

$$Y = TFP \cdot M^{1-\gamma}\left[L^{\alpha}K^{1-\alpha}\right]^{\gamma}, \tag{C.7}$$

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

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

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

$$\overline{TFPR} = \frac{\widetilde{TFPR}}{\left[\sum_{i=1}^{M} \frac{y_i}{Y} \frac{\left(1-\tau_i^y\right)}{\left(1+\tau_i^\ell\right)} \right]^{\alpha} \left[\sum_{i=1}^{M} \frac{y_i}{Y} \frac{\left(1-\tau_i^y\right)}{\left(1+\tau_i^k\right)} \right]^{1-\alpha}}. \tag{C.9}$$

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

D Model Implications Under Log-Normality

Define deviations of log draws from means,

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

Define the deviation of log effective agricultural ability from mean,

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

Note that u_{ni} is normally distributed with mean $E\left(u_{ni}\right) = 0$ and variance $VAR\left(u_{ni}\right) = E\left(u_{ni}^2\right) = \sigma_n^2$. In turn, $\widehat{u}_{ai}$ is also normally distributed with mean $E\left(\widehat{u}_{ai}\right) = E\left(u_{\varphi i}\right) + E\left(u_{ai}\right) = 0$ and variance,

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

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

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

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

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

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,

$$\log(I_{ai}) = b_a + u_{\varphi i} + u_{ai} = b_a + \widehat{u}_{ai}, \tag{D.10}$$

$$\log(I_{ni}) = b_n + u_{ni}, \tag{D.11}$$

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

Sectoral employment The probability an individual chooses to become a farm operator in agriculture,

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

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 $\Phi(.)$ 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 become a worker in non-agriculture (and therefore the fraction of individuals that choose non-agriculture) is $N_n = 1 - \Phi(b)$.

E Fixed Effect Estimates of Farm-level Measures

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 decompose land input, and non-agricultural income as follows,

$$\log \ell_{vit} = \beta_0^\ell + \beta_1^\ell \log LQ_v + \zeta_t^\ell + \zeta_{iv}^\ell + e_{ivt}^\ell \tag{E.12}$$

$$\log I_{n,vit} = \beta_0^{I_n} + \zeta_t^{I_n} + \zeta_{iv}^{I_n} + e_{ivt}^{I_n} \tag{E.13}$$

The interpretation for the regressors is the same as for equation (4). Note that equation (E.13) does not have a land quality term, as it does not directly impact non-agricultural income. We follow the same procedure as for equation (4) to 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 ℓ_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 ℓ_i. We then compute the moments of interest over the above estimated permanent components of distortions, agricultural income, and non-agricultural income.

F Inferring Population Moments from Observed Moments

The following are the specific steps in the procedure we follow to recover the population moments of the distributions of abilities across sectors and distortions.

1. Using equation (15) 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 (19), (20), and (21) 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 $\widehat{\sigma}_a$ and non-agriculture σ_n, and the covariance of abilities σ_{an} alone. We solve this 3×3 system for the three population moments $\widehat{\sigma}_a$, σ_n, σ_{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 σ_a, the dispersion of distortions σ_{φ}, and the covariance of abilities in agriculture and distortions $\sigma_{a\varphi}$ using the 3×3 system in equations (22), (23), and (24). 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.