Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China†

Tasso Adamopoulos
York University

Loren Brandt
University of Toronto

Jessica Leight
IFPRI

Diego Restuccia
University of Toronto and NBER

September 2021

Abstract

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

JEL classification: O11, O14, O4, E02, Q1.
Keywords: agriculture, misallocation, selection, productivity, China.

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

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

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

with attributing inputs to plots within farms.

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

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

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

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

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

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

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

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.

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

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

The paper proceeds as follows. In the next section, we describe the specifics of the land market institutions in China. Section 3 describes the panel data from China and the variables we use in our

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

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

2 Land Market Institutions in China

The Household Responsibility System (HRS), established in rural China in the late 1970s and early 1980s, dismantled the system of collective management set up under Mao and extended use rights over farmland to rural households. These reforms triggered a spurt in productivity growth in agriculture in the early 1980s that subsequently dissipated. This level effect is often attributed to the improved effort incentives for households as they became residual claimants in farming (McMillan et al., 1989; Lin, 1992). Ownership of agricultural land however remained vested with the collective, and specifically, 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, the 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.[5]

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

[5]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).

and Brandt, 2002). These frictions in the land market could generate allocative inefficiency or misallocation by distorting the allocation of land across farmers. Also, the inability to use land as collateral for borrowing purposes could result in the misallocation of other inputs such as capital.

3 Data

We use household survey data collected by the Research Center for the Rural Economy under the Ministry of Agriculture of China.[6] 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

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

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

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

4 Measuring TFP and Misallocation in Agriculture

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

4.1 Basic Framework

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

As in Lucas Jr (1978), the production technology available to farmer i in village v at time t with

productivity s_{ivt} exhibits decreasing returns to scale in variable inputs and is given by the Cobb-Douglas function,

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

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.[7]

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

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

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

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

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

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

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

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

for farm i in village v at time t. 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; with idiosyncratic distortions this need not be the case.

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

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

4.2 Measuring Farm Productivity

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

make a consistent mapping of the data to model variables, we remove the variation in labor input by normalizing value added, land, and capital by total labor days.

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

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

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

where μ_t^{TFP} is a year fixed-effect component that captures time-varying shocks to productivity that are common to all farmers; μ_i^{TFP} is a farm-specific component that does not vary over time; and e_{ivt}^{TFP} captures idiosyncratic shocks specific to the farmer in a given year. We use fixed effect panel data methods to estimate equation (4) to extract the household fixed effects. Note that μ_i^{TFP} is

inclusive of village-specific effects that do not change over time, but differ across individuals in different villages. The village effects cannot be separately identified from (4) because they are collinear with household fixed effects. In a second step, we remove the village-specific effects by regressing the household fixed effect on village dummies (μ_v) and extracting the residual,

$$\mu_i^{TFP} = \mu_v^{TFP} + \zeta_i^{TFP}, \tag{5}$$

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

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

Land Quality. An alternative interpretation of the differences in farm TFP is that they represent unobserved variation in land quality across households, a form of measurement error. Given that our micro data do not provide information on land quality at the farm-level, we use detailed land quality data from the Global Agro Ecological Zones (GAEZ) project of the Food and Agricultural

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

Organization (FAO) to compute measures of land quality at the village level. Variation in this measure provides bounds on potential differences in land quality between households within villages. GAEZ provides data on a rich set of land quality characteristics relevant for agricultural production at the 5 arc-minute resolution (roughly cell size of 10×10 kilometers) for the whole world. These data include soil attributes (e.g., fertility, depth), climate attributes (e.g., moisture, temperature), and terrain attributes (e.g., elevation, slope). Following Adamopoulos and Restuccia (2021), we use the detailed geographical data from GAEZ to construct a summary measure of land quality for each village. We find that the dispersion in this measure of land productivity is relatively small across our sample villages in China with a standard deviation of the log of 0.096.[10] By comparison, in the Philippines the dispersion of a similar measure of land quality across villages is 0.21 (Adamopoulos and Restuccia, 2020).

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

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

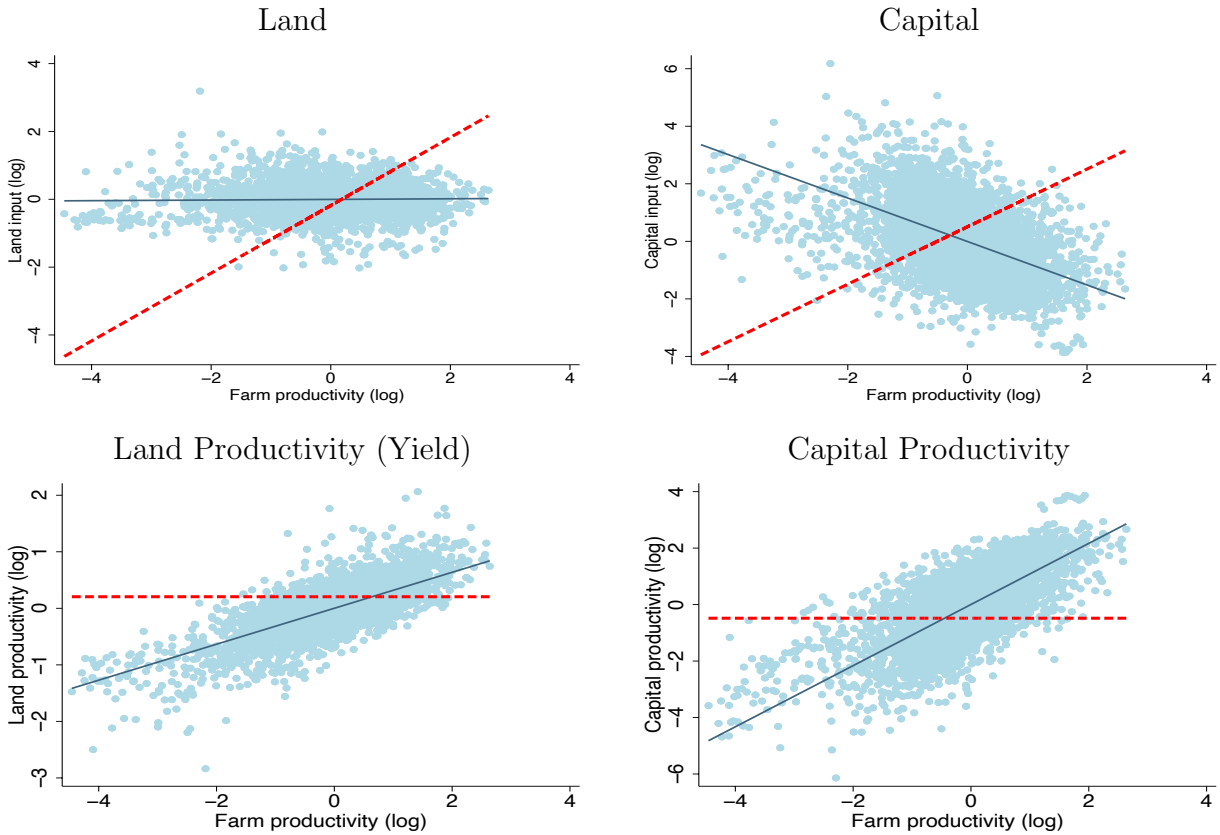
4.3 Distortions and Productivity

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

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

The input allocations across farmers in China indicate that there is substantial misallocation. In the context of our decentralized framework, this misallocation is manifested through farm-specific distortions or "wedges," measured as deviations of input allocations between actual and efficient. We summarize the distortions faced by a farm in both the land and capital markets by the measure of revenue productivity TFPR in equation (2). We follow the same procedure as with physical

Figure 1: Factor Allocations by Farm Productivity



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

productivity to estimate permanent farm-specific distortions using the panel data on $TFPR_{ivt}$. In particular, using panel methods we first estimate,[11]

$$\log TFPR_{ivt} = \mu_t^{TFPR} + \mu_i^{TFPR} + e_{ivt}^{TFPR}. \tag{6}$$

The interpretation of the regressors is the same as for equation (4). Then we remove village-specific

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

factors from the estimated household fixed effects by regressing them on village dummies (μ_v) to extract the residual,

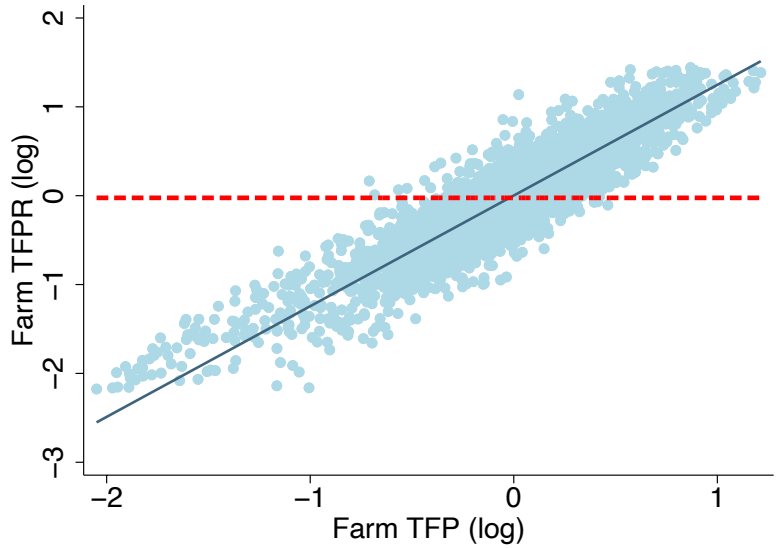
$$\mu_i^{TFPR} = \mu_v^{TFPR} + \zeta_i^{TFPR}. \tag{7}$$

This procedure provides an estimate of 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. The more productive farmers face higher farm-specific distortions. This relationship reflects the nature of the land market institutions in China. The administrative egalitarian allocation of land, along with the thin rental land markets, provide little scope for farmers to adjust the operational size of the farm they are "endowed" with. The farmers hurt the most from such an institutional setting are the more productive who would have optimally expanded the most in unfettered markets, operating more land and capital. In the context of our decentralized framework, this is reflected in higher farm-specific distortions on the more productive farmers.

4.4 Mismeasurement

Typical analyses of misallocation with cross-sectional data are often criticized for potentially misinterpreting idiosyncratic transitory shocks or measurement error as permanent TFP differences. Our analysis is less subject to these critiques because we exploit the panel structure of the data to obtain fixed-effect estimates of farm productivity and farm distortions. We illustrate the value of our approach by applying the method of Bils et al. (2017) for inferring measurement error in TFPR when panel micro-data are available, and comparing statistics from our fixed effects estimates and cross-sectional data.

Figure 2: Farm-specific Distortions and Productivity



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

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

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

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

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

We estimate equation (8) by OLS, clustering standard errors at the village level. We estimate it for three different measures of TFPR: (a) our baseline farm fixed effects TFPR measure that adjusts for village factors; (b) a farm fixed-effect TFPR measure, but without controlling for village factors (other than land quality); and (c) a raw unadjusted cross-sectional measure of TFPR that controls only for village-level land quality. Changes in output and inputs are the same across the different specifications; what varies is the measure of TFPR used. The point estimates for λ along with standard errors and the corresponding 95% confidence intervals are reported in the last row of Table 1. With the raw unadjusted measure of TFPR, we estimate $\widehat{\lambda} = 0.900$, which implies that measurement error is 10%. For the fixed effect estimate of TFPR from which village factors have not been purged, $\widehat{\lambda} = 0.955$, implying just under 5% measurement error. For our baseline measure of permanent TFPR the estimated λ is 1.00, implying none of the type of measurement error this method can capture with our baseline measure. Our findings suggest that the productivity gains from reallocation, implied by our baseline farm fixed effect estimates of TFP and TFPR, are fairly

reflective of the true potential gains from such a reallocation.

Table 1: Mismeasurement in Productivity and Distortions

	Fixed Effect Estimates		Cross-section
	Household Farm	+ Village	average
Farm TFP:			
STD(log)	0.35	0.64	0.72
p90/p10	2.19	4.35	5.59
p75/p25	1.48	2.06	2.32
Farm TFPR:			
STD(log)	0.48	0.81	0.92
p90/p10	3.14	7.17	9.70
p75/p25	1.78	2.71	3.23
CORR (logTFP, logTFPR)	0.91	0.88	0.88
BKR $\hat{\lambda}$	1.00	0.96	0.90
Standard error	(.026)	(.039)	(.024)
95% confidence interval	[0.95, 1.05]	[0.88, 1.03]	[0.85, 0.95]

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

A caveat of the method is that it can capture only certain types measurement error. In particular, it is suitable for assessing the extent of additive measurement error that is orthogonal to true farm productivity. If, for example, measurement error is multiplicative it cannot be identified through this method. As a result, the lack of measurement error in our baseline farm fixed effects estimates may represent a lower bound on the extent of measurement error in the data.

Estimates reported in Table 1 also reveal a marked reduction in the dispersion of productivity and distortions as we move from the cross-section to the fixed effects estimates. Consistent with a reduction in mismeasurement, the standard deviation of log-distortions is reduced by 10 percent

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

4.5 Other Evidence

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

Comparison to developed countries. In Figure 1 we contrasted actual allocations of land and capital in China against an efficient benchmark where more productive farms command more inputs. The question arises whether this is a reasonable benchmark. Unfortunately, we do not have access to micro level farm data for a developed country with secure property rights in land to confirm the tight link between land use and farm productivity implied by the benchmark efficient allocation. Nevertheless, the evidence suggests a much stronger relationship between farm size and productivity in developed economies. We present two pieces of evidence. First, using the U.S. Census of Agriculture data in Adamopoulos and Restuccia (2014) that provide information on land, capital, value added, and number of farms across 12 farm land-size categories, we calculate the correlation between farm size (operated land input) and several measures of productivity. We find a high positive correlation among these variables: a correlation around 90 percent between farm

size and labor productivity, and more than 80 percent between farm size and several alternative measures of farm TFP. This high correlation contrasts sharply with the land allocation in China documented earlier that results in no correlation between land input and farm productivity. The sharp difference between the United States and China is consistent with recent empirical evidence for developed and developing countries. Rada and Fuglie (2019) summarize the evidence from micro studies suggesting a positive relationship between farm size and productivity in developed economies (such as Australia and the United States) and no systematic relationship in poor and developing countries. Second, the evidence among many developed economies including the longer historical data for the United States and Canada is that average farm sizes have grown substantially along with high rates of agricultural productivity growth (Adamopoulos and Restuccia, 2014).

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

We examine the role of land rental by focusing on those provinces with land rental above the 5 percent mean. Applying the same fixed-effects methodology to this sample, the correlation of our permanent measures (in logs) of land input and farm productivity across households increases from 0.02 in the full sample to 0.13 for the provinces with more significant land rental activity.[12] This evidence indicates that rentals contribute to a land allocation that is more connected to farm productivity, however the effect is fairly small. In related work, Chari et al. (2021) examine the

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

effects of a land tenancy reform that was announced in China in 2003, and was implemented after the period in our data set. Using a difference-in-differences approach, they find that provinces in China exposed to the reform saw increased land rental activity directed towards more productive farmers, leading to significant output and productivity gains.[13]

4.6 Efficiency Gains

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

Table 2 reports the efficiency gains from factor reallocation using our baseline measure of farm TFP, estimated as the fixed effect component from a panel regression, purged of village effects. Extreme values of agricultural output, distortions, and non-agricultural income are winsorized at 0.5 percent tails. 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). In addition, if we allow for input reallocation across villages, i.e., we use household productivity measures without removing the village-specific effects but controlling for village land quality, the reallocation gain increases to 53.2 percent (second row, first column). When only capital can be reallocated across villages, in addition to capital and land between farms

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

Table 2: Efficiency Gains from Reallocation

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

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

within villages, the efficiency gains are 33.2 percent. These results imply that the gains from reallocation more than double when allowing for resource reallocation across villages, with about two-thirds accounted for by household mobility across space. This finding is consistent with the results of Ngai et al. (2019) who show that the limited ability to trade land for cultivation and lack of access to social services in the cities for those with rural hukou constitute substantial barriers to population movements across sectors and space.

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

In general, it is difficult to identify the sources of dispersion in marginal products when measuring input wedges indirectly. The identification issue is particularly acute when the underlying institution creating misallocation affects multiple inputs. In this case, the impact of the institution is better captured by an output wedge that affects dispersion in all marginal products. The land institution we emphasize for China also has effects on capital allocations. As a result, to decompose the impact

of the land institution in the efficiency gains we documented, we separate between an output wedge and a capital to land wedge. Our interpretation is that the output wedge is driven by the land institution whereas the capital to land wedge relates to other factors affecting the capital to land ratio across farmers, including potentially the use of different technologies with varying capital intensities. This is equivalent to interpreting the reported TFPR variation as output wedges, which leaves the capital-to-land ratio constant, and the residual variation as a capital-to-land wedge. Under this decomposition and for the estimated household fixed effect measures, the efficiency gain associated with the land institution is 13.6 percent, which accounts for almost 60 percent of the efficiency gains (24.4) reported earlier, while the remaining relates to other factors affecting the capital-to-land ratio across farmers (see Table 2). We focus on the TFPR variation as an output wedge associated with the land institution in our quantitative sectoral analysis.

We also find no substantial changes in the magnitude of the misallocation in the rural sector, and thus the gains from reallocation over time in China during our sample period. The standard deviation of log TFPR in the cross-section data is roughly constant over time, as are the reallocation gains associated with this misallocation. This finding contrasts with the reduction in misallocation found for the manufacturing sector in China, documented in Hsieh and Klenow (2009) over a similar period. Our findings are consistent with the view of costly farm-specific distortions that are tied to land market institutions in China that have not changed much over the period we study.

5 Misallocation and Selection across Sectors

We integrate our framework of agriculture into a two-sector general-equilibrium model of selection (Roy, 1951; Lagakos and Waugh, 2013) to assess: (1) how farm-specific distortions in agriculture alter the occupational choice between agriculture and non-agriculture; and (2) how selection affects measured misallocation and the productivity gains from reallocation. Individuals are endowed with productivities for each sector and make an occupational choice. Production in agriculture is

undertaken by heterogeneous farms that differ not only in agricultural ability, but also with respect to farm-specific distortions. A key innovation is that farm distortions, such as those tied to the land market institutions in China, directly distort occupational choices and amplify the productivity effects of misallocation.

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

An individual i choosing to work for the non-agricultural sector earns $I_{ni} = w_n s_{ni}$, where w_n is the non-agricultural wage per unit of skill. An individual that is a farm-operator in agriculture chooses inputs and output to maximize profits subject to the idiosyncratic distortion. Farm income is the value of output which includes not only the return to farm labor but also the land and capital incomes. We express an individual's income from agriculture as,

$$I_{ai} = w_a \varphi_i s_{ai}, \tag{9}$$

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

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

$$w_a \varphi_{ai} s_{ai} \geq w_n s_{ni} \tag{10}$$

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

The sectoral model has implications for the share of employment in agriculture, the pattern of selection, and sectoral and aggregate productivity. In addition, the model has micro-level implications summarized by moments of sectoral incomes conditional on sectoral choices, as well as by moments on farm-level productivity and farm-level distortions for those operating in agriculture. We provide more details on the model, calibration, and additional results in Supplemental Material Appendix D, E, and F.

Calibration. Our strategy is to calibrate distortions, abilities, and sectoral selection in a Benchmark Economy (BE) to the panel household-level data from China. We assume a trivariate log-normal distribution for the joint distribution of agricultural ability, non-agricultural ability and farm-specific distortions. We then proceed in two steps. First, we infer the population moments on abilities and distortions from observed moments on sectoral incomes and estimated wedges. Second, given the population moments, we calibrate the remaining parameters from the general equilibrium

equations of the model to match relevant data targets. To back out the population moments, we: (i) construct model moments on sectoral incomes and farm distortions, conditional on sectoral choices as functions of the population moments; (ii) compute the counterparts to the conditional moments in our panel-data from China, using our estimated fixed effect permanent components of distortions, agricultural income, and non-agricultural income for each household; and (iii) solve a system of equations for the population moments.

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

The calibration of the remaining parameters ensures that at the aggregate level the share of employment in agriculture in China is 46 percent. The model reproduces well the macro and micro statistics for China, with the log-normal assumption for the distributions of distortions and abilities, which affords substantial tractability, providing a good fit of the empirical distributions. In particular, whereas the productivity gain from eliminating distortions across existing farmers with different levels of productivity (correlated distortions) in the data is 13.9 percent (Table 2), it is 10 percent in the model with parametric distributions of distortions and abilities.

The effect of farm-specific distortions. We conduct a counterfactual experiment to assess the effects of the land market institution in China. We do so by eliminating distortions that are correlated with agricultural ability. We remove the systematic component of distortions simply by regressing distortions $\log(\varphi_i)$ on log agricultural ability for all individuals in the benchmark economy and retaining the residual. We then compute the equilibrium with only residual distortions. Note

that in this experiment there is still misallocation associated with the dispersion of distortions unrelated to farm productivity. Table 3 reports the results. Agricultural labor productivity increases by almost 3-fold. This increase arises from an increase in agricultural TFP of 67 percent and the associated reduction in the labor share in agriculture from 46 percent to 16 percent. Note that the effect on agricultural TFP of eliminating correlated distortions, holding constant farmers in the benchmark economy, is 10 percent in our calibrated model. As a result correlated distortions contribute substantially to agricultural productivity via distorted selection in occupational choices. We note that the overall labor productivity gains from removing the land market friction are in absolute terms well within values in the literature on agricultural labor productivity gaps (Restuccia et al., 2008).

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

Eliminating correlated farm-specific distortions increases agricultural labor productivity directly by reducing misallocation across farms, but also induces higher ability farmers to work in agriculture since they face no systematic restrictions on consolidating operational land. These two effects constitute the increase in agricultural TFP. At the same time, the associated increases in productivity via reduced misallocation and improved selection reduce the share of employment in agriculture, which further raises agricultural labor productivity by increasing the amount of land and capital available per farmer. In terms of selection, whereas in the benchmark economy with distortions employment in agriculture is roughly uniformly distributed across all agricultural ability types, removing distortions improves selection into agriculture of mostly high ability farmers. In terms of TFP in agriculture, out of the overall 1.67-fold increase, more than four-fifths $(1 - \log(1.10)/\log(1.67))$

Table 3: The Effects of Correlated Distortions

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

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

represent the effect of selection (an additional 50 percent). In other words, the amplification effect of selection on agricultural TFP quadruples the gains from reduced misallocation when keeping the set of farmers fixed (10 percent).

Average ability in non-agriculture, and hence real labor productivity in non-agriculture, falls to 78 percent of their benchmark economy values, due to more workers in non-agriculture, not all of whom are as productive in that sector. The nominal non-agricultural to agricultural productivity ratio falls from 3.95 to 3.13, when the share of labor in agriculture drops (from 46 to 16 percent). This finding is quantitatively consistent with the evidence for a wide range of countries that lower employment shares in agriculture are associated with lower nominal productivity ratios (see for example Figure 2 and Table 7 in Storesletten et al., 2019). Despite the significant impact on agricultural productivity,

real GDP per worker increases about 18 percent. The reason for the dampened effect on aggregate output is that there is a large shift of labor to the sector that experiences a drop in productivity relative to the benchmark. The dispersion of TFP in agriculture is lower, albeit with a higher mean.

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

6 Conclusions

Using a simple quantitative framework and micro panel data, we present evidence that capital and land are severely misallocated across farmers within villages in China. Exploiting the panel nature of the micro data, we estimate for each household permanent fixed effect farm-level productivity and distortions devoid of village differences that considerably limit the extent of measurement error in these measures. 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 mis-

allocation shows no systematic tendency to improve, consistent with the persistent nature of the institutional restrictions in the Chinese economy.

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

Our analysis implies that implementing a system of secure property rights to facilitate a decentralized allocation of land would generate large productivity gains. To the extent that village officials do not observe farmer ability or do not make land allocation decisions based on ability, any administrative (re)allocation of land would be unable to channel land to farmers that value it the most or can make the most out of it. Developing a market allocation mechanism by extending fully transferable use rights over land to farmers would not only allow farmers to increase their operational scales through land consolidations, but also induce the best farmers to stay in agriculture, while releasing labor to non-agriculture. The productivity and farm size increases due to a better allocation of factors of production among farmers and an improved selection of farmers in agriculture can arguably also induce changes in farm operations by incentivizing farmers to use modern inputs and better technologies. We leave this important extension of our framework for future research.

References

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–97.

Adamopoulos, T. and Restuccia, D. (2020). Land reform and productivity: A quantitative analysis

with micro data. *American Economic Journal: Macroeconomics*, 12(3):1–39.

Adamopoulos, T. and Restuccia, D. (2021). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. *Review of Economic Studies*.

Benjamin, D. and Brandt, L. (2002). Property rights, labour markets, and efficiency in a transition economy: the case of rural china. *Canadian Journal of Economics*, pages 689–716.

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

Bils, M., Klenow, P. J., and Ruane, C. (2017). Misallocation or mismeasurement? *Stanford Center for International Development*, (Working Paper No. 599).

Brandt, L., Huang, J., Li, G., and Rozelle, S. (2002). Land rights in rural china: Facts, fictions and issues. *The China Journal*, (47):67–97.

Brandt, L., Tombe, T., and Zhu, X. (2013). Factor market distortions across time, space and sectors in china. *Review of Economic Dynamics*, 16(1):39–58.

Cao, K. H. and Birchenall, J. A. (2013). Agricultural productivity, structural change, and economic growth in post-reform china. *Journal of Development Economics*, 104:165–180.

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

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

Chen, C., Restuccia, D., and Santaeulàlia-Llopis, R. (2021). The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics*.

Chow, G. C. (1993). Capital formation and economic growth in china. *The Quarterly Journal of Economics*, 108(3):809–842.

De Janvry, A., Emerick, K., Gonzalez-Navarro, M., and Sadoulet, E. (2015). Delinking land rights from land use: Certification and migration in mexico. *American Economic Review*, 105(10):3125–49.

Fan, S. and Zhang, X. (2002). Production and productivity growth in chinese agriculture: New national and regional measures. *Economic Development and Cultural Change*, 50(4):819–838.

Gollin, D., Parente, S., and Rogerson, R. (2002). The Role of Agriculture in Development. *American Economic Review*, 92(2):160–164.

Gottlieb, C. and Grobovšek, J. (2019). Communal land and agricultural productivity. *Journal of Development Economics*, 138:135–152.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Jin, S., Huang, J., and Rozelle, S. (2010). Agricultural productivity in china. *The shifting patterns of agricultural production and productivity worldwide*, pages 229–278.

Kung, J. K.-s. and Liu, S. (1997). Farmers' preferences regarding ownership and land tenure in post-mao china: unexpected evidence from eight counties. *The China Journal*, (38):33–63.

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

Lin, J. Y. (1992). Rural reforms and agricultural growth in china. *The American Economic Review*, pages 34–51.

Lucas Jr, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523.

McMillan, J., Whalley, J., and Zhu, L. (1989). The impact of china's economic reforms on agricultural productivity growth. *The Journal of Political Economy*, pages 781–807.

Ngai, L. R., Pissarides, C. A., and Wang, J. (2019). China's 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.

Storesletten, K., Zhao, B., and Zilibotti, F. (2019). Business cycle during structural change: Arthur lewis' theory from a neoclassical perspective. Technical report, National Bureau of Economic Research.

Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, 109(5):1843–72.

Yang, D. T. (1997). China's land arrangements and rural labor mobility. *China Economic Review*, 8(2):101–115.

Zhu, X. (2012). Understanding china's growth: Past, present, and future (digest summary). *Journal of Economic Perspectives*, 26(4):103–124.