Econometrica Supplementary Material

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

(*Econometrica*, Vol. 90, No. 3, May 2022, 1261–1282)

TASSO ADAMOPOULOS
Department of Economics, York University

LOREN BRANDT
Department of Economics, University of Toronto

JESSICA LEIGHT
International Food Policy Research Institute

DIEGO RESTUCCIA
Department of Economics, University of Toronto and NBER

APPENDIX A: MEASURES OF OUTPUT AND INPUTS FROM PANEL DATA

WE DESCRIBE in more detail the measures of outputs and inputs that we use in our analysis.

A.1. *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.

A.2. *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

Tasso Adamopoulos: aadamo@yorku.ca
Loren Brandt: loren.brandt@utoronto.ca
Jessica Leight: j.leight@cgiar.org
Diego Restuccia: diego.restuccia@utoronto.ca

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 10 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.

APPENDIX 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

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

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 (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,

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

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

APPENDIX C: EQUILIBRIUM AND IDENTIFICATION OF DISTORTIONS

In the basic framework, we 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 distributed as a lump sum transfer to 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

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

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:

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

$$\frac{\text{MRPK}_i}{(1-\alpha)\gamma} = \frac{y_i}{k_i} = \frac{r\left(1+\tau_i^k\right)}{(1-\alpha)\gamma\left(1-\tau_i^y\right)} \propto \frac{\left(1+\tau_i^k\right)}{\left(1-\tau_i^y\right)}, \tag{C.2}$$

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.1) and (C.2) 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.1)–(C.2) 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 a summary measure of distortions faced by farm i,

$$\text{TFPR}_i = \frac{y_i}{\ell_i^\alpha k_i^{1-\alpha}} = \widetilde{\text{TFPR}}\frac{\left(1+\tau_i^\ell\right)^\alpha\left(1+\tau_i^k\right)^{1-\alpha}}{\left(1-\tau_i^y\right)}, \tag{C.3}$$

where $\widetilde{\text{TFPR}} \equiv (\frac{q}{\alpha\gamma})^\alpha(\frac{r}{(1-\alpha)\gamma})^{1-\alpha}$ 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,

$$Y = \text{TFP}\cdot M^{1-\gamma}\left[L^\alpha K^{1-\alpha}\right]^\gamma, \tag{C.4}$$

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

$$\text{TFP} = \left[\frac{A_a \sum_{i=1}^{M} s_i \left(\frac{\overline{\text{TFPR}}}{\text{TFPR}_i} \right)^{\frac{\gamma}{1-\gamma}}}{M} \right]^{1-\gamma}, \tag{C.5}$$

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

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

Equation (C.5) makes clear that with no dispersion in TFPR_i across farm households, the equilibrium allocations, aggregate output, and TFP coincide with the corresponding efficient statistics.

APPENDIX D: MODEL OF MISALLOCATION AND SELECTION ACROSS SECTORS

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

D.1. *Environment*

We consider a representative closed village economy and for simplicity we drop village subscripts to focus on individual and sector differences. At each date, there are two goods produced, agricultural (a) and nonagricultural (n). The nonagricultural 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.

D.1.1. *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 nonagricultural 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 is allocated to the agricultural good. Individual i faces the following budget constraint:

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

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

Working in the agricultural sector involves operating a farm and is subject to idiosyncratic distortions, captured by φ_i, which we define more fully below. Income from working in the nonagricultural sector is subject to a tax η, common to all individuals. η operates as a barrier to labor mobility from agriculture to nonagriculture 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 nonagricultural labor productivity, but otherwise plays no significant role in our analysis.

D.1.2. *Individual Abilities and Distortions*

Individuals are heterogeneous with respect to their abilities in agriculture and nonagriculture, 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 nonagricultural 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 nonagricultural goods; and (b) an occupational choice, whether to work in the nonagricultural sector or the agricultural sector. We denote the income an individual i would earn in agriculture as I_{ai} and that in nonagriculture as I_{ni}. The individual chooses the sector with the highest income. We denote by H_n and H_a, the sets of $(s_{ai}, \varphi_i, s_{ni})$ values for which agents choose each sector $H_n = \{(s_{ai}, \varphi_i, s_{ni}) : I_{ai} < I_{ni}\}$, and $H_a = \{(s_{ai}, \varphi_i, s_{ni}) : I_{ai} \geq I_{ni}\}$.

D.1.3. *Consumption Allocation*

To determine the allocation of income between agricultural and nonagricultural 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:

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

D.1.4. *Production in Nonagriculture*

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

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

The representative firm in the nonagricultural 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 nonagriculture implies $w_n = A_n$.

D.1.5. *Production in Agriculture*

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

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

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 of the paper. Tax revenues are redistributed equally to the N workers independently of occupation, and equal to T per individual.

The profit maximization problem for farm i is given by

$$\max_{\ell_i, k_i} \left\{\pi_i = p_a(1 - \tau_i) y_{ai} - r k_i - q \ell_i\right\}, \tag{D.2}$$

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

$$\ell_i = A_a (\gamma p_a)^{\frac{1}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}} (1 - \tau_i)^{\frac{1}{1-\gamma}} s_{ai}, \tag{D.3}$$

$$k_i = A_a (\gamma p_a)^{\frac{1}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha\gamma}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}} (1 - \tau_i)^{\frac{1}{1-\gamma}} s_{ai}, \tag{D.4}$$

$$y_{ai} = A_a (\gamma p_a)^{\frac{\gamma}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}} (1 - \tau_i)^{\frac{\gamma}{1-\gamma}} s_{ai}, \tag{D.5}$$

$$\pi_i = A_a (1-\gamma) p_a^{\frac{1}{1-\gamma}} \gamma^{\frac{\gamma}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}} (1 - \tau_i)^{\frac{1}{1-\gamma}} s_{ai}. \tag{D.6}$$

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

$$I_{ai} = w_a \varphi_i s_{ai}, \tag{D.7}$$

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

$$w_a \equiv p_a^{\frac{1}{1-\gamma}} A_a \gamma^{\frac{\gamma}{1-\gamma}} \left(\frac{1-\alpha}{r}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}}. \tag{D.8}$$

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 rewrite land input demand, capital input demand, output supply, and profits for farmer i in terms of their agricultural ability and farm-specific distortions,

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

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

D.1.6. *Occupational Choice*

Individuals choose to operate a farm in the agricultural sector or be workers in the nonagricultural sector. If individual i chooses to operate a farm, their income is given by (D.7), while if they choose to work in nonagriculture their income is

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

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

D.1.7. *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 nonagricultural 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 nonagricultural production allocation $\{Y_n, N_n\}$ maximizes the profits of the nonagricultural 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 nonagricultural goods clear; and (f) the government budget constraint from the tax-transfer scheme is satisfied.

D.2. *Analysis*

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

$$N_a = \Phi(b), \tag{D.9}$$

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

$$b \equiv \frac{b_a - b_n}{\sigma}, \qquad b_a \equiv \log(w_a) + \mu_\varphi + \mu_a, \qquad b_n \equiv \log(w_n) + \log(1-\eta) + \mu_n, \tag{D.10}$$

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

Define deviations of log draws from means,

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

Define the deviation of log effective agricultural ability from mean,

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

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

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

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

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

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

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

The log-incomes of individual i from agriculture and nonagriculture, 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 rewrite agricultural and nonagricultural incomes as the sums of constants and log mean deviations,

$$\log(I_{ai}) = b_a + u_{\varphi i} + u_{ai} = b_a + \widehat{u}_{ai}, \tag{D.11}$$

$$\log(I_{ni}) = b_n + u_{ni}, \tag{D.12}$$

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

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

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

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, that is, $N_a = n_a$. Similarly, we can show that the probability an individual chooses to work in nonagriculture (and therefore the fraction of individuals that choose nonagriculture) is $N_n = 1 - \Phi(b)$.

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

$$E\{\log(\widehat{s}_{ai})|i \in H_a\} = \widehat{\mu}_a + \frac{\sigma_{an} - \widehat{\sigma}_a^2}{\sigma}\lambda^l(b), \tag{D.13}$$

while the average log-ability in nonagriculture among those choosing nonagriculture is

$$E\{\log(s_{ni})|i \in H_n\} = \mu_n + \frac{\sigma_n^2 - \sigma_{an}}{\sigma}\lambda^u(b), \tag{D.14}$$

where $\widehat{\mu}_a = \mu_a + \mu_\varphi$. $\lambda^l(b) \equiv E[\xi|\xi \le b] < 0$ and $\lambda^u(b) \equiv E[\xi|\xi > b] > 0$ represent lower tail truncation and upper tail truncation of a standard normal random variable ξ. The coefficients in (D.13) and (D.14) can be rewritten as $\frac{\widehat{\sigma}_a\sigma_n}{\sigma}[\rho_{an} - \frac{\widehat{\sigma}_a}{\sigma_n}]$ and $\frac{\widehat{\sigma}_a\sigma_n}{\sigma}[\frac{\sigma_n}{\widehat{\sigma}_a} - \rho_{an}]$. 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 nonagricultural 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 (nonagriculture) is lower (higher) than the population average.

APPENDIX E: CALIBRATION OF SECTORAL MODEL WITH SELECTION

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

E.1. *Inferring Population Moments*

Assuming a multivariate log-normal distribution for the joint population distribution of abilities and distortions, we first back out the moments of that distribution (variances and covariances) so that we match observed moments on incomes across sectors and farm-specific distortions. We use the moments implied by our permanent fixed-effects estimates of these variables. Normalizing the means of these distributions to zero, there are five population moments that need to be calibrated: three variances, 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,

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

2. Variance of log income in nonagriculture conditional on choosing nonagriculture,

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

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

$$\begin{aligned}&\mathrm{COV}\{\log(I_{ai}), \log(I_{ni})|i \in H_a\}\\ &\quad\equiv \widehat{c}_{an} = \sigma_{an} - \left(\frac{\sigma_n^2 - \sigma_{an}}{\sigma}\right)\left(\frac{\sigma_{an} - \widehat{\sigma}_a^2}{\sigma}\right)\lambda^l(b)\left[\lambda^l(b) - b\right].\end{aligned} \tag{E.3}$$

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

$$\mathrm{VAR}\{\log(\varphi_i)|i \in H_a\} \equiv \widehat{v}_\varphi = \sigma_\varphi^2\left\{1 - \left(\frac{\sigma_\varphi^2 + \sigma_{a\varphi}}{\sigma\sigma_\varphi}\right)^2 \lambda^l(b)\left[\lambda^l(b) - b\right]\right\}. \tag{E.4}$$

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

$$\begin{aligned}&\mathrm{COV}\{\log(I_{ai}), \log(\varphi_i)|i \in H_a\}\\ &\quad\equiv \widehat{c}_{I_a,\varphi} = \left(\sigma_{a\varphi} + \sigma_\varphi^2\right) + \left(\frac{\sigma_{an} - \widehat{\sigma}_a^2}{\sigma}\right)\left(\frac{\sigma_\varphi^2 + \sigma_{a\varphi}}{\sigma}\right)\lambda^l(b)\left[\lambda^l(b) - b\right].\end{aligned} \tag{E.5}$$

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 nonagricultural income for each household. Specifically, we use panel methods to estimate permanent measures of land input and nonagricultural income and then along with the permanent estimates of TFP and TFPR we back out all the other variables of interest. To obtain fixed-effect estimates for other farm-level variables including land input and nonagricultural income, we apply a method similar to that in equations (4) and (5), and use the model equations to solve for the rest. We first decompose land input and nonagricultural income,

$$\log \ell_{vit} = \mu_t^\ell + \mu_i^\ell + e_{ivt}^\ell, \tag{E.6}$$

$$\log I_{n,vit} = \mu_t^{I_n} + \mu_i^{I_n} + e_{ivt}^{I_n}, \tag{E.7}$$

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

$$\mu_i^\ell = \mu_v^\ell + \zeta_i^\ell, \tag{E.8}$$

$$\mu_i^{I_n} = \mu_v^{I_n} + \zeta_i^{I_n}. \tag{E.9}$$

The interpretation for the regressors is the same as for equations (4) and (5) in the paper. Using this procedure, we estimate permanent farm-specific components of land input $\widehat{\zeta}_i^\ell$ and non-agricultural income $\widehat{\zeta}_i^{I_n}$. Denote the permanent measures of land input ℓ_i and nonagricultural 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 $\varphi_i = (1/\text{TFPR}_i)^{1/(1-\gamma)}$, 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 estimated permanent components of distortions, agricultural income, and nonagricultural income.

With the permanent farm measures, we then compute empirical moments on the standard deviations of log agricultural income; log nonagricultural income; log distortions for farm operators; covariance of log agricultural income and log distortions; and the covariance of log agricultural income and log nonagricultural 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 nonagricultural activities, with many households switching from agriculture to nonagriculture under a variety of definitions of switchers in our panel data. The moment we use as our baseline is the contemporaneous covariance of log sectoral incomes across households, which implies a correlation of 0.033 in our microdata.

This moment is robust to two alternative classifications of households in agriculture and nonagriculture over the sample period. Our first classification defines a household in non-agriculture if their reported cultivated land and farm output are zero; the second classifies a household in nonagriculture if they self-report based on time allocation as mainly in non-agriculture or are full-time in nonagriculture. Identifying switchers on the basis of these alternative definitions, we reestimate the correlation in log incomes in agriculture and nonagriculture. Compared to an estimate of 0.03 in the contemporaneous correlation, we find a correlation between log income in agriculture and nonagriculture of 0.036 (0.020) for switchers using the first (second) classification.

In our system of equations, the five population moments of variances and covariances are identified by the five conditional moments of variances and covariances given by equations (E.1)–(E.5). Table E.I contains the empirical conditional variances and covariances along with the share of employment in agriculture that we target.

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

1. Using equation (D.9), we invert the standard normal to recover the parameter b that generates a share of employment in agriculture of 46 percent. This gives a $b = -0.10$.

TABLE E.I

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 nonagricultural income	0.46
$\widehat{v}_\varphi$	STD of farm distortions	1.05
$\widehat{c}_{an}$	COV between agricultural and nonagricultural incomes	0.005
$\widehat{c}_{a\varphi}$	COV of agricultural income and farm distortions	−0.14

Note: All variables refer to logs.

2. Note that equations (E.1), (E.2), and (E.3) give the variance of the log of agricultural income conditional on choosing agriculture, $\hat{v}_a$; the variance of the log of nonagricultural income conditional on choosing nonagriculture, $\hat{v}_n$; 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 nonagriculture 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 (E.4), (E.5), and the definition of $\widehat{\sigma}_a$. These equations give the variance of the log of distortions $\hat{v}_\varphi$, the covariance of log agricultural income and log distortions conditional on working in agriculture $\hat{c}_{a,\varphi}$, and the definition of the variance of agricultural ability in relation to the variance of effective agricultural ability solved in previous step 2.

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

In Table E.II, we report the resulting population moments using the procedure outlined above. Note that instead of reporting the covariances of agricultural and nonagricultural 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 nonagricultural occupations. This implies that individuals sort into the sector in which they possess a comparative advantage, and that individuals working in each sector are on average more skilled in that sector than the general population.

TABLE E.II

CALIBRATED POPULATION MOMENTS.

Parameter	Description	Value
σ_a	STD of agricultural ability	1.30
σ_n	STD of nonagricultural 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–nonagricultural ability	−0.15

Note: All variables refer to logs.

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.

E.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.[1]

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 nonagriculture, 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 of the paper; the endowment of land L is set to match an average farm size of 0.45 hectares observed in our microdata, 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 "long-run" share of employment in agriculture to 1 percent.

TABLE E.III

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 nonagricultural to agricultural productivity gap	3.95
Z_a/N_a	Average ability in agriculture	3.43
Z_n/N_n	Average ability in nonagriculture	1.72
$(Z_n/N_n)/(Z_a/N_a)$	Ratio of nonagricultural to agricultural ability	0.50
Y/N	Real GDP per worker	1.13
Y_a^e/Y_a	Efficiency gain among existing farmers	1.15
	Conditional microlevel 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

[1]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.

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 relative price of agriculture normalized to one, a capital to output ratio in agriculture of 0.3 observed in our microdata, and a ratio of labor productivity in nonagriculture to agriculture of 3.96. Table E.III displays the aggregate and microlevel statistics for the benchmark economy, as well as the values for the calibrated parameters. Real agricultural labor productivity is total agricultural output Y_a divided by employment in agriculture N_a. TFP_a is total agricultural output divided by the bundle of inputs $N_a^{1-\gamma}(L^\alpha K^{1-\alpha})^\gamma$ from the implied aggregate production function.

APPENDIX F: ADDITIONAL QUANTITATIVE EXPERIMENTS

Our main counterfactual in Section 5 of the paper eliminates the correlation of distortions with agricultural productivity across farmers. In this section, we present the results from additional quantitative experiments.

F.1. *Eliminating All Distortions*

In Table F.I, we present the results from removing all farm-specific distortions in agriculture, that is, setting $\varphi = 1$ for all i. Eliminating all distortions on farms has a substantial impact on the economy. Agricultural labor productivity increases 3.4-fold and the share of employment in agriculture falls 32 percentage points, from 46 percent to 14 percent. The magnitude of these effects is similar to those when we eliminate only the correlation

TABLE F.I

EFFECTS OF ELIMINATING DISTORTIONS.

Statistic	Benchmark Economy BE	No Correlated Distortions	No Distortions $\varphi_i = 1$
	Aggregate statistics		
Real agricultural productivity (Y_a/N_a)	1.00	2.96	3.42
Share of employment in agriculture (N_a) (%)	0.46	0.16	0.14
TFP in agriculture (TFP_a)	1.00	1.67	1.80
TFP in agriculture, constant BE farms	1.00	1.10	1.15
Real nonagricultural productivity (Y_n/N_n)	1.00	0.78	0.77
Average ability in agriculture (Z_a/N_a)	1.00	2.34	2.65
Average ability in nonagriculture (Z_n/N_n)	1.00	0.78	0.77
Real GDP per worker (Y/N)	1.00	1.18	1.19
	Conditional microlevel statistics		
STD of log–farm TFP	0.56	0.39	0.35
STD of log–farm TFPR	0.48	0.14	0
CORR of log–(farm TFP, farm TFPR)	0.97	0.44	–
CORR of log–(agr. ability, nonagr. ability)	0.15	0.49	0.50
CORR of log–(agr. income, nonagr. income)	0.03	0.40	0.50

Note: The counterfactual "No Correlated Distortions" refers to the economy when eliminating correlated distortions. The counterfactual "No Distortions" eliminates all farm-level distortions, that is, 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 microlevel statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

of distortions with agricultural ability analyzed in Section 5 of the paper. This implies that the key driver of the results is the systematic nature of the distortions, associated with the land institution we emphasize, that constrains the more productive farmers.

F.2. *Amplification Effect of Distortions and* ρ_{an}

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

TABLE F.II

ELIMINATING CORRELATED DISTORTIONS WITH ALTERNATIVE ρ_{an}.

Statistic	$\rho_{an} = -0.15$		$\rho_{an} = 0$		$\rho_{an} = 0.15$	
	BE	NC	BE	NC	BE	NC
Aggregate statistics:						
Real agricultural productivity (Y_a/N_a)	1.00	2.96	1.00	3.44	1.00	4.23
Share of employment in agriculture (N_a) (%)	0.46	0.16	0.46	0.14	0.46	0.11
TFP in agriculture (TFP_a)	1.00	1.67	1.00	1.80	1.00	1.98
Real nonagricultural productivity (Y_n/N_n)	1.00	0.78	1.00	0.78	1.00	0.77
Average ability in agriculture (Z_a/N_a)	1.00	2.34	1.00	2.72	1.00	3.35
Average ability in nonagriculture (Z_n/N_n)	1.00	0.78	1.00	0.78	1.00	0.77
Real GDP per worker	1.00	1.18	1.00	1.21	1.00	1.22
Conditional microlevel statistics:						
STD of log–farm TFP	0.56	0.39	0.56	0.35	0.56	0.31
STD of log–farm TFPR	0.48	0.14	0.48	0.13	0.48	0.12
CORR of log–(farm TFP, farm TFPR)	0.97	0.44	0.97	0.24	0.97	−0.08
CORR of log–(agr. ability, nonagr. ability)	0.15	0.49	0.28	0.54	0.38	0.57
CORR of log–(agr. income, nonagr. income)	0.03	0.40	0.45	0.58	0.76	0.72

Note: "BE" refers to the calibrated benchmark economy with each population correlation of abilities ρ_{an}. "NC" refers to the experiment of eliminating the correlation of distortions within the economy with the corresponding ρ_{an}. All aggregate variables, except for the share of employment in agriculture, are reported relative to the same statistic in the benchmark economy in each correlation case. All microlevel statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

which increases by 80 and 98 percent compared to 67 percent in the baseline, and the consequent larger decline in agricultural employment.

F.3. *Comparison to an Exogenous Increase in Productivity*

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

In the first two columns of Table F.III, we reproduce the results for the benchmark economy and the economy without correlated farm-level distortions (our main counterfactual). In columns 3–4, we show in turn the effects of increasing exogenously TFP in agriculture and then in both agriculture and nonagriculture by 10 percent (keeping all farm-level distortions in place). An exogenous increase in TFP reduces the share of employment in agriculture from 46 percent in the benchmark economy to 34 percent, however, agricultural TFP increases only by 15 percent, compared to 67 percent when eliminating correlated idiosyncratic distortions. Agricultural TFP increases a bit more than the exogenous increase in TFP because there is a small effect on selection into agriculture, increasing average quality of workers in agriculture by 11 percent, compared to 134 percent when eliminating correlated distortions. The effects are similar when nonagricultural TFP also increases exogenously by 10 percent (column four).

The reduction in misallocation associated with the elimination of correlated farm-level distortions has a much larger effect on agricultural labor productivity than an equivalent-

TABLE F.III

COMPARISON OF REMOVING DISTORTIONS VS. EXOGENOUS TFP INCREASES.

Statistic	BE	No Corr Dist	$\uparrow (A_a^{1-\gamma}) \times 1.10$	$\uparrow (A_a^{1-\gamma}, A_n) \times 1.10$
		Aggregate statistics		
Real agricultural productivity (Y_a/N_a)	1.00	2.96	1.35	1.35
Share of employment in agriculture (N_a) (%)	0.46	0.16	0.34	0.34
TFP in agriculture (TFP_a)	1.00	1.67	1.15	1.15
Real nonagricultural productivity (Y_n/N_n)	1.00	0.78	0.92	1.01
Average ability in agriculture (Z_a/N_a)	1.00	2.34	1.11	1.11
Average ability in nonagriculture (Z_n/N_n)	1.00	0.78	0.92	0.92
Real GDP per worker (Y/N)	1.00	1.18	1.09	1.18
		Conditional microlevel statistics		
STD of log– farm TFP	0.56	0.39	0.56	0.56
STD of log– farm TFPR	0.48	0.14	0.48	0.48
CORR of log–(farm TFP, farm TFPR)	0.97	0.44	0.97	0.97
CORR of log– (agr. ability, nonagr. ability)	0.15	0.49	0.20	0.20
CORR of log–(agr. income, nonagr. income)	0.03	0.40	0.12	0.12

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

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

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 nonagriculture. 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.

F.4. *Model Specification and Calibrated* ρ_{an}

We show that model specification matters for the calibrated population correlation of abilities ρ_{an}, but targeting the same empirical correlation of incomes across sectors imposes discipline on the amplification effect from eliminating correlated farm-level distortions. Since in our baseline calibration we abstract from other idiosyncratic factors determining occupational choice, a near zero income correlation even with distortions in agriculture requires a low value of ability correlation. Does our amplification result hinge on this particular inference and model structure?

We explore alternative settings for the occupational choice problem in the model to include idiosyncratic preference or mobility barriers, denoted by θ_i. The occupational decision depends now on the barrier draw, that is, an individual selects agriculture if $I_{ai} > \theta_i I_{ni}$. For simplicity, we assume that θ_i is *i.i.d.* and drawn from a log-normal distribution with mean 0 and variance σ_θ^2. In Table F.IV, we recalibrate the benchmark economy for different values of σ_θ, targeting the exact same correlation of incomes across sectors of 0.03. In each case, we recalibrate all moments and parameters to match the same other data targets. Column "BE" in each case presents the moments for the recalibrated economy under each σ_θ.

We find that increasing the dispersion of idiosyncratic barriers increases the implied ρ_{an} required to produce the same empirical correlation of incomes across sectors. The column "NC" in each case shows the results from the main quantitative experiment of eliminating correlated farm distortions. The results are fairly similar across the different values of σ_θ. The fact that the results are robust against different σ_θ and implied ρ_{an} indicates that while the calibration does not identify the "true" value of ρ_{an}, the targeted value of the cross-sector income correlation imposes discipline on the magnitude of the amplification mechanism from eliminating correlated distortions in the model.

TABLE F.IV

NO CORRELATED DISTORTIONS WITH IDIOSYNCRATIC BARRIERS.

Statistic	$\sigma_\theta = 0$		$\sigma_\theta = 0.5$		$\sigma_\theta = 0.9$	
	BE	NC	BE	NC	BE	NC
Calibrated ability correlation	$\rho_{an} = -0.15$		$\rho_{an} = -0.08$		$\rho_{an} = -0.03$	
Aggregate statistics:						
Real agricultural productivity (Y_a/N_a)	1.00	2.96	1.00	3.17	1.00	3.10
Share of employment in agriculture (N_a) (%)	0.46	0.16	0.46	0.15	0.46	0.15
TFP in agriculture (TFP_a)	1.00	1.67	1.00	1.73	1.00	1.72
Real nonagricultural productivity (Y_n/N_n)	1.00	0.78	1.00	0.83	1.00	0.89
Average ability in agriculture (Z_a/N_a)	1.00	2.34	1.00	2.51	1.00	2.47
Average ability in nonagriculture (Z_n/N_n)	1.00	0.78	1.00	0.83	1.00	0.89
Real GDP per worker	1.00	1.18	1.00	1.26	1.00	1.33
Conditional microlevel statistics:						
STD of log–farm TFP	0.56	0.39	0.56	0.38	0.56	0.42
STD of log–farm TFPR	0.48	0.14	0.48	0.13	0.48	0.13
CORR of log–(farm TFP, farm TFPR)	0.97	0.44	0.97	0.30	0.97	0.17
CORR of log–(agr. ability, nonagr. ability)	0.15	0.49	0.10	0.40	0.05	0.26
CORR of log–(agr. income, nonagr. income)	0.03	0.40	0.03	0.35	0.03	0.24

Note: "Calibrated ability correlation" refers to the calibrated ρ_{an} targeting the same empirical correlation of incomes across sectors of 0.03, under different σ_θ. "BE" refers to the calibrated benchmark economy with each dispersion in mobility barriers σ_θ. "NC" refers to the experiment of eliminating the correlation of distortions within the economy with the corresponding pair of σ_θ and ρ_{an}. All aggregate variables, except for the share of employment in agriculture, are reported relative to the same statistic in the benchmark economy in each correlation case. All microlevel statistics are reported in levels, and are conditional on choosing agriculture in the corresponding simulated economy.

REFERENCES

LAGAKOS, DAVID, AND MICHAEL E. WAUGH (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *The American Economic Review*, 103 (2), 948–980. [17]

Co-editor Fabrizio Zilibotti handled this manuscript.

Manuscript received 18 August, 2018; final version accepted 1 October, 2021; available online 22 October, 2021.