University of Toronto
Department of Economics

Working Paper 748

Distortions, Producer Dynamics, and Aggregate Productivity: A General Equilibrium Analysis

By Stephen Ayerst, Loren Brandt and Diego Restuccia

February 16, 2023

Distortions, Producer Dynamics, and Aggregate Productivity: A General Equilibrium Analysis*

Stephen Ayerst† Loren Brandt‡ Diego Restuccia§

February 2023

Abstract

The expansion in farm size is an important contributor to agricultural productivity in developed countries, but the reallocation process is hindered in less developed economies. How do distortions to factor reallocation affect farm dynamics and agricultural productivity? We develop a model of heterogeneous farms making cropping choices and investing in productivity improvements. We calibrate the model using detailed farm-level panel data from Vietnam, exploiting regional differences in agricultural institutions and outcomes. We focus on south Vietnam and quantify the effect of higher measured distortions in the North on farm choices and agricultural productivity. We find that the higher distortions in north Vietnam reduce agricultural productivity by 46%, accounting for around 70% of the observed 2.5-fold difference between regions. Moreover, two-thirds of the productivity loss is driven by farms' choice of lower productivity crops and reductions in productivity-enhancing investment, which more than doubles the productivity loss from factor misallocation.

Keywords: Farm dynamics, productivity, size, distortions, misallocation, Vietnam.
JEL classification: O11, O14, O4.

*We thank Tasso Adamopoulos for help with the grid land-quality data for Vietnam. For useful comments we thank Costas Azariadis, Dan Cao, Julieta Caunedo, Klaus Desmet, Mark Huggett, David Lagakos, Toshi Mukoyama, Richard Rogerson, Kirk White, Zoltan Wolf, and seminar participants at Georgetown, UBC, ASU Herrendorf memorial conference, and US Census. All remaining errors are our own. Restuccia gratefully acknowledges the support from the Canada Research Chairs program and the Bank of Canada Fellowship program. The views expressed herein are those of the authors and should not be attributed to the Bank of Canada or its Governing Council, nor the IMF, its Executive Board, or its management.

†International Monetary Fund, stephen.b.ayerst@gmail.com.

‡University of Toronto, loren.brandt@utoronto.ca.

§University of Toronto and NBER, diego.restuccia@utoronto.ca.

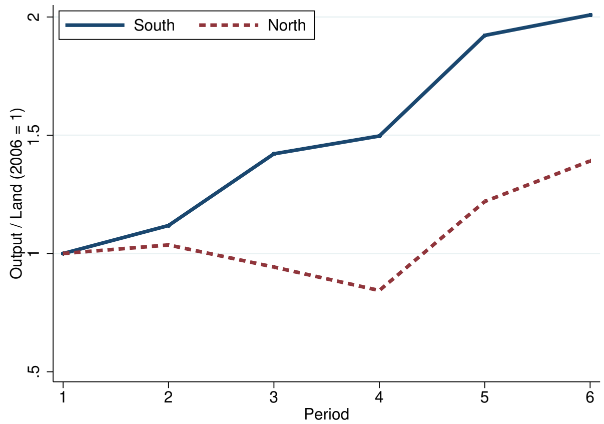
1 Introduction

The reallocation of resources across businesses is a salient feature of the growth process in developed economies: successful businesses expand, while unsuccessful businesses contract and even exit (Baily et al., 1992; Davis et al., 1998; Foster et al., 2001). Business dynamism is also observed in the agricultural sector in these countries where land consolidation and farm exit have been substantial drivers of productivity growth (Key, 2019). In sharp contrast, the process of reallocation is hindered in less developed countries by a variety of regulations, policies and institutions (Adamopoulos and Restuccia, 2014; Restuccia and Rogerson, 2017). In this paper, we examine how distortions to factor reallocation affect agricultural productivity and growth in the context of a model of farm heterogeneity and dynamics. Exploiting farm-level panel data from Vietnam and regional differences in agricultural institutions, we find substantial differences in agricultural productivity, farm dynamics, and crop choice across regions, the bulk of which can be accounted for by measured differences in distortions.

Vietnam agriculture offers a valuable context to study these issues. Since the late 1980s, the country has undergone major reforms—including decentralizing production to the household level and liberalizing output and input markets—that led to substantial improvements in productivity and growth. Reforms have been highly uneven across regions however allowing us to focus on regional differences in the cost of misallocation. We exploit detailed household panel data on output and inputs at the farm level that allow us to measure productivity and distortions at the farm and regional level.

In our data, agricultural productivity in the South is more than twice as large as in the North, with a substantial gap also observed in productivity growth. Figure 1 documents the evolution of land productivity across a 10-year period in the North and the South in our data. Land productivity grew at an annualized rate of 7.1 percent in the South compared to only 3.3 percent in the North. This gap in the evolution of productivity resembles differences in the life-cycle profiles in employment and productivity in the United States and India emphasized in Hsieh and Klenow (2014).

Figure 1: Farm Dynamism in North and South Vietnam



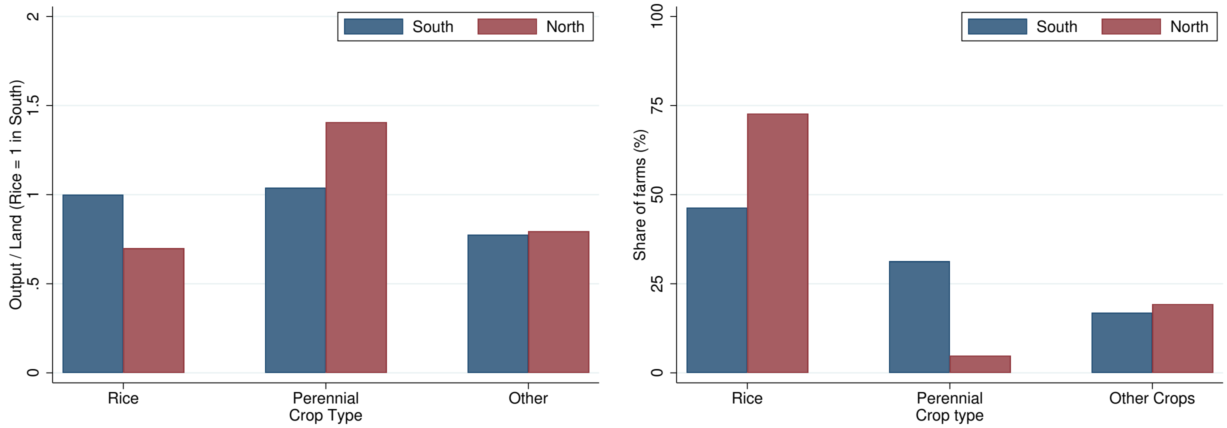
Notes: Land productivity is calculated as the total value of crop production by farms in our data divided by total cultivated area for each region and year, normalized to one in both regions for the first period of data. Data details are provided in Section 4.

We also observe striking differences between the North and the South in the crop composition of farms, which may be related to institutional differences (e.g., land-use restrictions between regions). Figure 2 shows that most farms in the North grow rice even though it is the least productive crop in the region. Indeed, land-use restrictions are more prevalent in the North in our data, where 43 percent of farms report being required to grow rice compared to 23 percent in the South. Figure 2 also illustrates that land productivity is more dispersed across crops in the North than in the South, suggesting the potential for larger misallocation in the North.

We develop a dynamic model of farms to understand the differences in productivity between the south and north Vietnam. The basic structure of the model follows Lucas (1978) in which heterogeneous farm managers hire land and labor in order to produce output using a decreasing returns to scale production technology. This leads to a non-degenerate distribution of farms in equilibrium where farm size and employment depend on the distribution of productivity.

Our main departure from the existing literature is to endogenize the productivity process of farms. Farm productivity depends on four components, the first of which is a permanent

Figure 2: Crop Comparison in North and South Vietnam



(a) Land productivity

(b) Crop shares

Notes: Land productivity is calculated as the total value of crop production by farms divided by total cultivated land, normalized to one for rice in the South. Data are from 2006. Data details and construction of crop types are described in Section 4.

farmer-specific productivity component. There is also a random productivity component that varies between periods. The third is an ability component that depends on farmer's endogenous efforts to improve their productivity. Farmers make investments to improve their managerial ability in the next period along a quality ladder. This component matches empirical evidence showing that less distorted farmers invest more in farm improvements and also experience faster productivity growth (Appendix B.2). Finally, there is a crop-specific productivity component that depends on the farmer's endogenous choice of crop. Following Adamopoulos and Restuccia (2020), this component captures factors that affect the relative profitability of farms growing different crops. Farmers select crops upon entry based on the expected value of each crop and an idiosyncratic preference across crops, which allows us to match the substantial overlap in farm productivity across different crops in the data (Appendix B.1).

We follow Restuccia and Rogerson (2008) by modeling institutional distortions as idiosyncratic output wedges at the farm level. The farm-level distortions capture institutional factors (e.g., land sale or rental restrictions, insecure property rights) that affect farmers' input choices relative to their first best. We allow institutional distortions to be correlated with the farmer's productivity, crop choice, and a random stochastic component that varies over time. In addition, we allow for government land-use regulations, such as restrictions on crop choice, that force farmers to grow a crop independent of its relative profitability or the household's idiosyncratic preferences. This constraint on household choices is motivated by our data in which farms report government restrictions on specific plots of land (see also, Le, 2020).

In equilibrium, aggregate productivity depends on both the extent of static misallocation caused by farm-level distortions as well as the endogenous distribution of farm-level productivities, which depends on farmers' investment in ability and crop selection. To understand the quantitative importance of these factors, we calibrate the model to data on south Vietnam from the Vietnam Access to Resources Household Survey (VARHS), a detailed household-level panel dataset covering 2006 to 2016. In the data, we group farms into three types: Rice, Perennial, and Other Crop farms on the basis of the value of their production in each of these crops. We discipline the model's parameters to match moments related to the productivity distribution across farms, changes in the distribution over time and differences across crop types.

Our main experiment involves adjusting distortion parameters in the benchmark economy, which is calibrated to south Vietnam, to match measured distortions in north Vietnam. Relative to the South, distortions in the North are more highly correlated with farm-level productivity, implying flatter allocation of factors with respect to farm productivity; perennial crops are more distorted compared to rice and other crops; and a higher share of plots face government-imposed crop restrictions. Imposing the North distortions on the South leads to: 1) productivity falling by 46% relative to the benchmark economy, accounting for just over two-thirds of the observed TFP gap between the North and the South; 2) a reduction in the average growth rate of farmers' output of 1.8 percentage points, or half of the observed gap between the North and the South; and 3) a drop in the share of perennial farmers from

33% to 7%, similar to the observed 5% in the North. In our quantitative analysis, a key institutional feature driving the productivity differences is the larger elasticity of distortions with respect to farm productivity, reflecting the greater inability of farmers in the North to expand their farm size.

To understand the channels through which larger measured distortions in the North account for the North-South productivity gap, we separate the impact of distortions between changes in factor misallocation, changes in the endogenous farm productivity distribution, and changes in crop choices. We find that all channels are important, with factor misallocation accounting for one-third of the productivity loss, and the remaining two-thirds arising from the reallocation of economic activity to lower-productivity crops (crop choice) and, more importantly, the shift in the within-crop farm productivity distribution due to lower investment. Comparatively, the channels of dynamic misallocation (crop choice and farm productivity) account for two-thirds of the productivity loss in the model, more than doubling the impact of static factor misallocation.

Our work connects with three strands in the literature. We relate to the broad literature emphasizing resource misallocation across production units for understanding aggregate productivity differences (Restuccia and Rogerson, 2008; Guner et al., 2008; Hsieh and Klenow, 2009) and in particular productivity in agriculture (Adamopoulos and Restuccia, 2014). We differ from this literature by comparing the effects of misallocation across regions rather than the level of misallocation, i.e., the overall cost with respect to an undistorted economy, which is potentially affected by mismeasurement. We also focus on the impact of distortions on crop choice and farm investment, emphasizing the dynamic effects of misallocation. In this regard, our work joins a recent literature studying the role of producer dynamics on aggregate outcomes (Hsieh and Klenow, 2014; Bento and Restuccia, 2017; Guner et al., 2018; Akcigit et al., 2021; Da-Rocha et al., 2017; König et al., 2022), and a literature in microeconomics studying the channels of firm-level upgrading in developing countries (Verhoogen, 2021). We differ from this literature in studying the dynamics of farm productivity, of which crop choice is subject to specific regulations as in the case of Vietnam. We also contribute to a growing literature investigating the specific aspects of economic growth and convergence in Vietnam (Tarp, 2017; Le, 2020; Ayerst et al., 2020).

The paper is organized as follows. Section 2 summarizes the institutional context. Section 3 describes the model. Section 4 summarizes the data, construction of key moments, and differences between the north and south Vietnam. Section 5 calibrates a benchmark economy with distortions to panel farm-level data from the South, discussing the model's quantitative properties and goodness of fit. Section 6 presents the quantitative analysis where the main experiment involves applying measured distortions in the North to the benchmark economy, and discusses the extent to which this counterfactual economy resembles key features of the North. Section 7 concludes.

2 Institutional Context

Our analysis begins in 2006, nearly two decades after the start of economic reform in Vietnam. Central to these efforts was the return to family farming. In the late 1980s, production rights to land reverted to households, and over time expanded to include rights to transfer, exchange, lease, inherit and mortgage. Titling of land began in 1994 with the passing of the 1993 Land Law and by 1997 Land Use Certificates had been issued to approximately one-half of all cultivated land (Benjamin and Brandt, 2004). By 2004, coverage extended to three-quarters of all cultivated land (Brandt et al., 2006) but subsequently stalled (Markussen, 2017).

Property rights' reforms were accompanied by liberalization of product markets, especially for rice, and input markets such as those for fertilizer (Benjamin and Brandt, 2004). Restrictions on the volume of rice exports were relaxed, as were internal product market barriers. Similarly, restrictions on fertilizer imports were removed. Prices came to be largely market-determined. Geographic mobility barriers were also relaxed.

Often neglected in discussions of Vietnam agriculture are important regional differences in institutions between the North and South, reflecting their unique histories. In the North, agriculture was collectivized in the 1950s and households organized into communes. Most of the land currently held by households in the North was obtained directly from the commune, which played a key role in the decentralization of land rights to households in the late 1980s. By contrast, agriculture in the South was never successfully collectivized after reunification in 1975, and farming continued to be carried-out at the household level. Property rights in land also remained de-facto private. As a result, land sales and rental exerted a much larger influence on patterns of landownership and use in the South compared to the North (Brandt et al., 2006).

Regional differences appear in other forms and are also likely a legacy of institutions and the role of the state in the North prior to reform. Restrictions persist on crop choice, largely related to rice production and national food security, and are more prominent in the North (Markussen et al., 2011; Markussen, 2017). Risk of land expropriation remains, with these risks negatively related to informal ties to local officials and cadres (Markussen and Tarp, 2014). Land titling has expanded but in sub-regions in the north remains well below national levels. Households in the North also are much more likely to report issues with respect to access to water for irrigation, and problems of flooding.

Farm households in Vietnam carry oiut productivity-augmenting investments in land and water and acquire information on new technologies and markets through their involvement in extension services. In Appendix B.2, we examine the effect of these investments on farm productivity growth, and their correlation with measures of market distortions discussed in more detail below. In both the North and the South, better and less-distorted farmers invest significantly more time and resources in improving farm productivity. However, the benefits of these investments are much weaker in the North, with only participation in extension services positively correlated with farm productivity growth. By contrast, in the South investments in cash and in-kind (labor) in land and water, as well as acquisition of new knowledge through extension services are positively correlated with farmer productivity growth.

3 Model

We develop a model of heterogeneous farms that make cropping decisions and invest in productivity improvements. Farmers face idiosyncratic distortions, as in Restuccia and Rogerson (2008), which affects the choice of inputs relative to the first best allocation; in addition, the government imposes crop restrictions on a set of farmers. Distortions, crop restrictions, and crop-specific differences affect the allocation of resources across farms and crops, as well as the productivity distribution through farmer incentives to invest in productivity improvements.

3.1 Economic Environment

Time is discrete and indexed by $t \in \{0, 1, 2, ..., \infty\}$. The economy is populated by a mass N of households, indexed by f, half that work as farm managers and half that supply labor as farm workers. We abstract from sectoral occupational choice (structural transformation) and selection, as the impact of distortions on these channels has been well-studied and are known to amplify the productivity cost of distortions (e.g., Adamopoulos and Restuccia, 2014, 2020; Adamopoulos et al., 2022). The economy is also endowed with a mass L of land that is used in agricultural production.

Production technology. We model crops $i \in \mathcal{I}$ as a technological choice to the farmer. In this regard, we take all agricultural production to be a single final good and assume that the choice of crop directly affects farm productivity. This simplification provides tractability while preserving the core economics of the farmer's crop choice.

A farmer f that grows crop $i \in \mathcal{I}$ produces output according to the following decreasing returns to scale technology in variable inputs,

$$y^i_{f,t} = (s^i_{f,t})^{1-\gamma}(\ell^\alpha_{f,t} n^{1-\alpha}_{f,t})^\gamma,$$

where $s^i_{f,t}$ is the crop-specific productivity of farmer f in period t; $\ell_{f,t}$ is the land input; and $n_{f,t}$ is the labor input. The productivity of farmer f growing crop i is equal to

$$\ln s^i_{f,t} = \ln z_f + \ln \kappa^i + \ln a_{f,t} + v_{f,t},$$

where z_f is a permanent component of farmer productivity with distribution $\Phi_z(z)$; κ^i is a crop i specific component of productivity; $a_{f,t}$ is the managerial ability of farmer f; and $v_{f,t}$ is a time varying stochastic component of farmer productivity with distribution $\Phi_v(v)$. The farmer's ability is determined through their investment decisions as we discuss below.

Investment. A farmer f can improve their managerial ability $a_{f,t}$ through investment. Farmer ability follows a quality ladder, such that a farmer that has successfully improved their ability h times has ability $a_{f,t} = a(h) = \lambda^h$. A farmer f that invests e_f units of the final output good improves their ability with probability $x_{f,t} = (e_f/\psi a_{f,t})^{1/\zeta}$. Rewriting this expression shows that expenditure on improving ability with success rate x_i is equal to

$$e(x_{f,t}, a_{f,t}) = \psi x^\zeta_{f,t} a_{f,t},$$

where $a_{f,t}$ is a scaling factor capturing that it is more costly for higher ability farmers to further improve their ability. The parameter ψ captures the level of the investment function, with the investment required to improve ability with the same probability increasing in ψ. The parameter $\zeta > 1$ captures the curvature of the investment function, which dictates how quickly increasing the success rate of improving ability increases the cost of investment. Appendix B.2 provides evidence on the types of investment that farmers engage in, such as investment in irrigation systems or extension services, and the corresponding improvements to productivity and productivity growth.

Preferences and cropping decision. The economy is populated by a mass N of households, indexed by f, half of which are farm managers (farmers) while the remaining households are workers and supply a unit of labor to farms. With probability ξ a household survives to the next period and with probability $1-\xi$ a household exits and is replaced by a new household. Household preferences are given by

$$U_f^o([C_t^o]) = \mathbb{E}_0\left[\sum_{t=0}^{\infty}(\xi\beta)^t C_t^o\right] \times b_f^o,$$

where $o \in \{\mathcal{I}, W\}$ is the occupation of the household, which can either be a worker W or a manager of a crop $i \in \mathcal{I}$ farm. The value of b^o captures an idiosyncratic preference for occupation o. We normalize the preference parameters for workers such that $b_f^W = 1$. The value of b_f^i is drawn by each farm manager from a Frechet distribution, $H(b) = \exp\{-(b/\eta^i)^{-\theta}\}$, where η^i is a crop-specific preference shifter that captures a common component of the utility cost of growing a crop i. Higher values of η^i correspond to, on average, more utility from growing crop i. The common crop-specific component η^i captures the average difficulty or lost revenues from preparing plots for specific crops. For example, growing perennial plants involves investment and seasons in which the plot does not produce output, which would be captured by lower η^i. The dispersion of preferences captures idiosyncratic factors to the farmer (e.g., slope of land, access to irrigation, soil quality) that cause farmers to prefer different crops even in the absence of market-based factors.

The idiosyncratic dispersion in the utility cost causes farmers to differ in their relative preference for growing different crops. For example, some farmers prefer to grow perennials while others prefer to grow rice, all else equal. This preference may be strong enough that some farmers choose to grow rice even if rice is less profitable than perennials. In equilibrium, this implies farmers select into crops based on both the relative market value and their relative preference for each crop, where the elasticity of selection to market forces is determined by the shape parameter θ. Modeling crop choice as a utility cost allows us to replicate the overlap in the productivity of farmers that grow different crops observed in the data (Appendix B.1). For example, there are many productive rice farmers and unproductive perennial farmers, despite perennial farmers being more productive on average.

Farmers that exit the economy are replaced by a new household that takes over management of their farm. We interpret exit as capturing both the exit of households from agriculture as well as the inter-generational transfer of the farm within families. New farmers enter the market with the same permanent productivity z as their predecessor and with ability $a_{f,t}$ drawn from a distribution that depends on the exiting farmer's ability. An exiting farmer with ability h is replaced by a farmer with ability $\tilde{h}$ drawn uniformly across values $\{0, 1, ..., h\}$ if that farmer chooses to grow the same crop i and ability $\tilde{h} = 0$ otherwise. This could be thought of as capturing a transfer of skills from parents to children or learning from other farmers in the region.

3.2 Market Structure

The final agricultural good is the numeraire. Following Restuccia and Rogerson (2008), we model institutional distortions in a reduced form as an idiosyncratic tax τ on farm revenues. Farm revenues net of the tax are equal to $(1 - \tau_f^i) y_f^i$.

Higher values of τ_f^i imply that farmers operate smaller farms than they would in the absence of the distortion. Distortions are distributed according to

$$\ln(1 - \tau_f^i) = (1 - \gamma) \left[\ln \bar{\tau} + \ln \varphi^i - \rho \ln \left(s_f^i \right) + \varepsilon_f \right], \tag{1}$$

where φ^i is a crop-specific distortion that captures institutional factors affecting crop choice (e.g., market access to sell or purchase specific inputs); ρ captures the elasticity of distortions to the underlying productivity of the farm reflecting correlated distortions (e.g., land size restrictions); and ε_f is a random idiosyncratic component of distortions with distribution $\Phi_\varepsilon(\varepsilon)$. Distortions are scaled by $(1-\gamma)$ to simplify algebra in the solution. We assume that the government's budget constraint is balanced by a lump-sum transfer T to households that is equal to the total amount collected from the idiosyncratic tax.

A fraction ω of farmers face government-imposed crop restrictions in which case the farmer must grow rice for their crop, such that $i = R$. This reflects a direct cropping restriction imposed by the Vietnamese government on individual farms that are quantitatively important for aggregate production (Le, 2020). These types of land-use restrictions are not captured by the standard misallocation wedges τ_f^i since they do not impact the farm's output or choice of inputs. Farmers receive the government-imposed crop restriction prior to making their crop choice, implying that some farmers facing the restriction would have still grown rice. The probability of facing the government-imposed crop restriction is unrelated to the distortions τ_f^i that farmers face implying that restricted rice farmers are otherwise identical to unrestricted rice farmers.

Timing. The timing of each period is: (i) farmers enter and farmer-specific productivities z_f and managerial ability $a_{f,t}$ of new farmers are realized; (ii) new farmers make crop choices; (iii) farmer ability and period-specific shocks (v, ε) are realized; (iv) farmers choose production inputs and realize profits; (v) farmers invest in future managerial ability; (vi) farmers exit.

3.3 Equilibrium

We focus on the steady state equilibrium in which the distribution of farm types, allocations, and prices are constant. We drop f and t subscript and write farmer outcomes in terms of the farmer's crop choice $i \in \mathcal{I}$, permanent farmer productivity z, current ability-level h, and current shocks (v, ε).

Production problem. Farmers choose resources (n, ℓ) to maximize total profits. The farmer's production problem is

$$\pi^i_{z,h}(v,\varepsilon) = \max_{n,\ell}(1-\tau^i_{z,h}(v,\varepsilon))s^i_{z,h}(v,\varepsilon)^{1-\gamma}(\ell^\alpha n^{1-\alpha})^\gamma - q\ell - wn.$$

Profits depend on farmer's idiosyncratic distortions $\tau^i_{z,h}(v,\varepsilon)$ and the farmer's productivity $s^i_{z,h}(v,\varepsilon)$. Solving the farmer's production problem implies that inputs are

$$\ell^i_{z,h}(v,\varepsilon) = \left[\frac{\gamma\alpha}{q}\left(\frac{1-\alpha}{\alpha}\frac{q}{w}\right)^{\gamma(1-\alpha)}\right]^{\frac{1}{1-\gamma}}(1-\tau^i_{z,h}(v,\varepsilon))^{\frac{1}{1-\gamma}}s^i_{z,h}(v,\varepsilon),$$
$$n^i_{z,h}(v,\varepsilon) = \left[\frac{\gamma(1-\alpha)}{w}\left(\frac{\alpha}{1-\alpha}\frac{w}{q}\right)^{\gamma\alpha}\right]^{\frac{1}{1-\gamma}}(1-\tau^i_{z,h}(v,\varepsilon))^{\frac{1}{1-\gamma}}s^i_{z,h}(v,\varepsilon).$$

Given the above level of inputs, output is

$$y^i_{z,h}(v,\varepsilon) = \gamma^{\frac{\gamma}{1-\gamma}}\left(\frac{\alpha}{q}\right)^{\frac{\alpha\gamma}{1-\gamma}}\left(\frac{1-\alpha}{w}\right)^{\frac{(1-\alpha)\gamma}{1-\gamma}}(1-\tau^i_{z,h}(v,\varepsilon))^{\frac{\gamma}{1-\gamma}}s^i_{z,h}(v,\varepsilon).$$

Investment problem. Farm profits are equal to $\pi^i_{z,h}(v,\varepsilon) = (1-\gamma)(1-\tau^i_{z,h}(v,\varepsilon))y^i_{z,h}(v,\varepsilon)$. The farmer's investment problem is to choose investment e, or equivalently the success rate x, to maximize the expected value of their farm. The problem is

$$V^i_{z,h}(v,\varepsilon) = \max_x\left(\pi^i_{z,h}(v,\varepsilon) - e(x,\lambda^h)\right) + (\xi\beta)\mathbb{E}_{v',\varepsilon'}\left[xV^i_{z,h+1}(v',\varepsilon') + (1-x)V^i_{z,h}(v',\varepsilon')\right].$$

The investment decision of the farmer solves

$$x^i_{z,h} = \left[\frac{(\xi\beta)\mathbb{E}_{v',\varepsilon'}\left[V^i_{z,h+1}(v',\varepsilon') - V^i_{z,h}(v',\varepsilon')\right]}{\psi\zeta\lambda^h}\right]^{\frac{1}{\zeta-1}},$$

where the farmer's investment decision does not depend on their current state (v,ε).

Crop decision. Let $\bar{V}_z^i$ denote the expected value of a new farm with crop i, permanent productivity z, and ability-level $h = 0$ before the shock (v, ε) is realized. Farmers with the government-imposed crop restriction do not choose their crop and are forced to produce rice, $i = R$. For unrestricted farmers, the crop decision is

$$\max_{i \in \mathcal{I}} \bar{V}_z^i \times b^i.$$

The resulting share of farmers that grow crop i is equal to

$$\Omega_z^i = \begin{cases} \omega + (1-\omega)\frac{(\eta^i \bar{V}_z^i)^\theta}{\sum_{i' \in \mathcal{I}} (\eta^{i'} \bar{V}_z^{i'})^\theta} & \text{for } i = R \\ (1-\omega)\frac{(\eta^i \bar{V}_z^i)^\theta}{\sum_{i' \in \mathcal{I}} (\eta^{i'} \bar{V}_z^{i'})^\theta} & \text{for } i \neq R \end{cases}. \tag{2}$$

See Appendix A for derivation of the above expression. The fraction of (unrestricted) farmers that choose a specific crop depends on both the relative expected value of growing that crop $\bar{V}_z^i$ and the relative difficulty of growing that crop, captured by the preference parameter η^i, where θ determines the elasticity of farmers to these factors.

Farm distribution. The evolution of farmer ability depends on the success rate $x_{z,h}^i$ chosen by farmers and the survival rate ξ. The evolution of the distribution of farm abilities is described by

$$\Delta \mu_{z,h}^i = \begin{cases} \mu_{E,z,0}^i - (1-\xi)\mu_{z,0}^i - \xi x_{z,h}^i \mu_{z,0}^i & \text{for } h = 0, \\ \mu_{E,z,h}^i - (1-\xi)\mu_{z,h}^i + \xi \left[\mu_{z,h-1}^i x_{z,h-1}^i - x_{z,h}^i \mu_{z,h}^i\right] & \text{for } h > 0, \end{cases}$$

where $\mu_{E,z,h}^i$ is the entry distribution of farmers with ability at node h. In the stationary equilibrium the distribution is defined by $\Delta \mu_{z,h}^i = 0$ for all values of h and the entry rate is equal to $\mu_{E,z,h}^i = (1-\xi) \sum_{\tilde{h}=h}^{\infty} \mu_{z,\tilde{h}}^i / (1 + \tilde{h})$.

Aggregate output. Production of the agricultural good is given by

$$Y = \left[\frac{\left(\int_{v,\varepsilon} e^{\gamma\varepsilon + (1-\rho\gamma)v} d\Phi_v(v) d\Phi_\varepsilon(\varepsilon) \right) \int_z \sum_i \sum_h (\varphi^i)^\gamma (z\kappa^i \lambda^h)^{1-\rho\gamma} \mu^i_{z,h} \Omega^i_z d\Phi_z(z)}{\left(\int_{v,\varepsilon} e^{\varepsilon + v(1-\rho)} d\Phi_v(v) d\Phi_\varepsilon(\varepsilon) \right)^\gamma \left(\int_z \sum_i \sum_h \varphi^i (z\kappa^i \lambda^h)^{1-\rho} \mu^i_{z,h} \Omega^i_z d\Phi_z(z) \right)^\gamma} \right] \times$$
$$N_F^{1-\gamma} (L^\alpha N_W^{1-\alpha})^\gamma, \qquad (3)$$

where $N_F = 0.5N$ is the mass of farm managers and $N_W = 0.5N$ is the mass of workers. The expression in square brackets describes the average productivity of farms and the impact of misallocation on aggregate productivity. In the undistorted economy, this expression simplifies to average productivity to the exponent $1-\gamma$, and total output is equal to output of a farm with average productivity multiplied by the total mass of farms. Aggregate productivity depends on (i) the misallocation of factors of production (n, ℓ); (ii) the share of farms growing each crop Ω^i_z; and (iii) the distribution of farmer abilities $\mu^i_{z,h}$ through farmer investment decisions. Notably, aggregate output does not depend directly on the average level of distortions $\bar{\tau}$, which is canceled out by general equilibrium effects. The final part of the expression describes the effect of decreasing returns to scale of the aggregate land and labor on output. Aggregate output has constant returns to scale in the aggregate land and labor (both workers and managers).

Equilibrium definition. The stationary competitive equilibrium is the set of values

$$\{C^W, q, w, T, n^i_{z,h}(v,\varepsilon), \ell^i_{z,h}(v,\varepsilon), V^i_{z,h}(v,\varepsilon), x^i_{z,h}, \mu^i_{z,h}, \Omega^i_z\}$$

for all $z \in \mathcal{Z}$, $h \in \{0, 1, 2, ..., \infty\}$, $i \in \mathcal{I}$ and values (v, ε) such that:

(i) Taking prices as given, $(n^i_{z,h}(v,\varepsilon), \ell^i_{z,h}(v,\varepsilon))$ maximize farm profits and $(V^i_{z,h}(v,\varepsilon), x^i_{z,h})$ maximize farm value.

(ii) The lump-sum transfer T balances the government's budget.

(*iii*) The distributions $(\mu^i_{z,h}, \Omega^i_z)$ are consistent with farm decisions and are stationary.

(*iv*) The land, labor, and output markets clear.

4 Data

We use data from the Vietnam Access to Resources Household Survey (VARHS) that covers households from 12 provinces in north and south Vietnam surveyed biennially between 2006 and 2016. We focus on a sub-sample of 2,118 households that are included in all six biennial surveys from 2006 to 2016. We provide an overview of the construction of variables and the mapping between data and model variables. The data are used in the next section to discipline the calibration of the model.

4.1 Variable Construction

We provide a brief overview of the data construction, a more detailed documentation is provided in Ayerst et al. (2020). Our main variables of interest are farm-level measures of output (value added) $y_{f,t}$, land $l_{f,t}$, and labor $n_{f,t}$. These values are also used to construct measures of farm-level productivity and distortions.

Farm-level output is measured as the sum of crop production, valued using a common price for each crop, net of intermediate input expenditure. To construct the common price, we first calculate the median price for each crop, as reported by households. We then construct the common price as the weighted sum of the median price across years, where weights are the relative total quantity of that crop's production for the year. We use crop sales and quantities, when available, to compute prices rather than the total value and quantities of farm output reported by households to avoid bias in the construction of prices.[1] Finally, we

[1] We make two adjustments to the output data for missing data and survey changes. First, the 2006 survey only asks for the total value of a crop produced by the household for some crop categories. For most crops, households are still asked to report both value and quantity. We regress crop prices on region, year, and crop fixed effects and then use the estimated fixed effects to construct predicted prices for the crops with missing quantity data. This allows us to impute a quantity for these crops in the 2006 survey. Second, the survey

set observations with negative value added to zero, which account for around 2.5% of our sample and are primarily in the North.

Farm-level land is constructed as the total cultivated area of plots owned and rented by households exclusive of any land that is used for activities unrelated to crop production (e.g., forestry, animal husbandry), left fallow for more than 48 months, or rented out.

Farm-level labor is constructed as the sum of labor hired from outside of the household and quality-adjusted labor supplied within the household. Using information on family members that hire out in agriculture, we run regressions of the wage rate on individual characteristics (e.g., age, education, sex), a year fixed effect, and a household region fixed effect. We use the average region-year wage rate to adjust household expenditure on outside labor into a quantity. We also use the estimated coefficients on individual characteristics to quality-adjust labor supplied within the household.

Following the model and Hsieh and Klenow (2009), we use the measures of output, land, and labor to construct farm-level measures of TFP and revenue productivity (TFPR). For farm f growing crop i, TFP is calculated as

$$\mathrm{TFP}^i_{f,t} = \frac{y^i_{f,t}}{(\ell^{\alpha}_{f,t} n^{1-\alpha}_{f,t})^{\gamma}} = (s^i_{f,t})^{1-\gamma},$$

and TFPR as

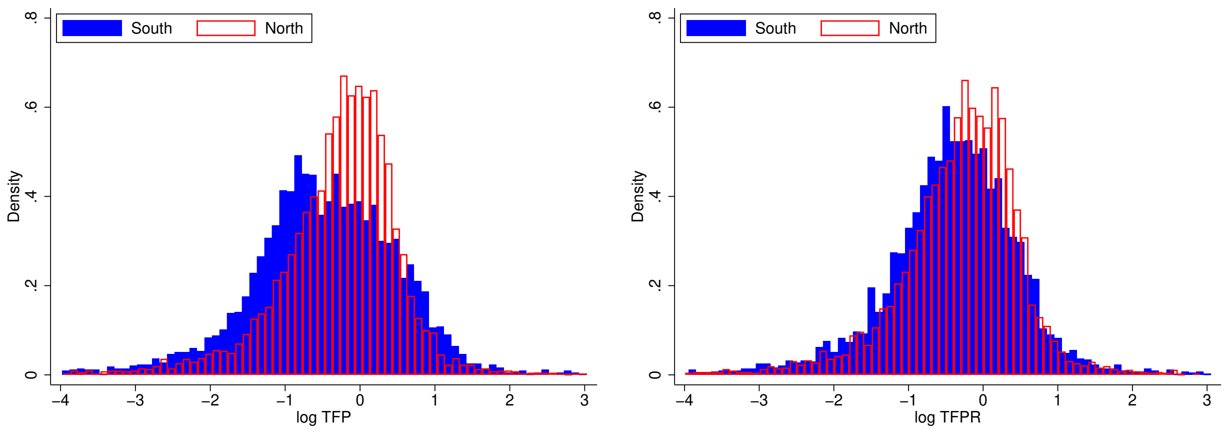
$$\mathrm{TFPR}^i_{f,t} = \frac{y^i_{f,t}}{\ell^{\alpha}_{f,t} n^{1-\alpha}_{f,t}} \propto \frac{1}{1-\tau^i_{f,t}}.$$

Note that revenue productivity (TFPR) is a model-based measure since output per unit of composite input equalizes across producers in an undistorted equilibrium and is otherwise proportional to distortions τ's. We set $\gamma = 0.70$ and $\alpha = 0.50$ to be consistent with the calibration presented in the next section. Figure 3 plots histograms of the TFP and TFPR

treats potatoes, cassava and sweet potatoes as a single crop in 2006 and as unique crops in 2008 and later surveys. For 2006, we treat this category as potatoes, noting that it only accounts for around 2.7% of total production value and all three crops fall into the 'Other Crop' farmer type (see below).

distributions in the final dataset, where both variables are normalized by the mean in each region (North / South) year. The figure highlights that both TFP and TFPR tend to be more dispersed in the South, as documented in Ayerst et al. (2020).

Figure 3: TFP and TFPR Distribution



(a) TFP Distribution

(b) TFPR Distribution

Notes: Histogram of TFP and TFPR for farm-year observations in north and south Vietnam. TFP and TFPR are normalized by the mean in each region (North / South) year.

In the next section, we use the data to discipline the calibration of the model. To minimize the influence of outliers, we winsorize the final set of variables (output, land, labor, TFP, and TFPR) at the top and bottom 2% for both the North and South in each year.[2]

4.2 Farm Types

In the model, farms select a specific crop i, which becomes its only type of output in the current and future periods. In the data, farmers often grow multiple crops. To map the data into the model, we group farms into three major crop types, which we label as rice, perennials and other (annual) crops, indexed by $i \in \{R, P, O\}$, based on the crop value of the farm.

In the survey, perennials include: fruits, coffee, tea, cocoa, cashew nuts, sugarcane, pepper,

[2] We winsorize rather than trim the data since trimming disproportionately affects the share of perennial farmers in the final dataset. Other than for the crop share, trimming the data implies similar moments compared to winsorizing.

rubber, medicinal trees and plants, and other perennial crops. Other crops include maize, potatoes, sweet potatoes, cassava, peanuts, soybeans, vegetables, and other annual crops. Farmers for whom more than half of their average yearly crop output is from rice (perennials) are rice (perennial) farmers while the remainder are "other crop" farmers. We do not impose strict cutoffs (e.g., 100% perennials) because of inter-cropping, crop-rotation, and the fact that farms may devote some of their land to other crops for home consumption. However, cropping tends to be concentrated in these categories. For example, over two-thirds of rice farmers have a rice share over 75% and just under half of rice farmers have a rice share over 90%. For perennials, these numbers are slightly higher at 70% and 50% of farmers for the same thresholds.[3]

Figure 4: Crop Composition by Farm

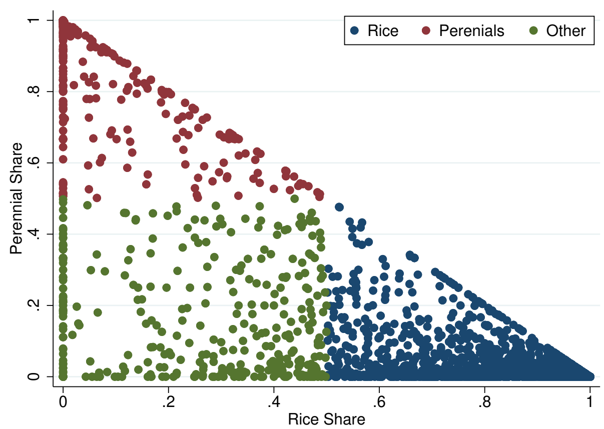


Figure 4 summarizes the empirical distribution of farms across crop types. Unsurprisingly, most farmers are classified as rice farmers as this is the most widely produced crop in Vietnam. The remaining farmers are split between perennials and other crops. The other crop farms tend to grow a large share of rice, on average around 30% of value, but still have a higher production value of other annual crops.

[3]In addition, more than 90% of the farmers we classify as rice and perennial farmers are classified the same way in all years. Differences are more common in the case of other crop farmers.

4.3 Summary Statistics

Table 1 provides summary statistics on farms by crop type for both south and north Vietnam. An observation is a farm-year and only includes farm-years for which TFP can be calculated. Farm-level TFP and TFPR are normalized to one in each region-year such that reported values capture relative productivity and distortions within regions.

Table 1: Summary Statistics

A. Vietnam, South

Crop Type	Mean						
	Output	Land	Labor	TFP	TPFR	TFP Growth	Obs
Rice	9.9	2.4	154.2	−0.4	−0.3	4.9	2,293
Perennials	10.6	3.9	294.6	−0.2	−0.3	10.9	1,368
Other	9.6	1.9	184.4	−0.7	−0.6	1.8	726
Total	10.1	2.8	203.0	–	–	6.2	4,387

B. Vietnam, North

Crop Type	Mean						
	Output	Land	Labor	TFP	TPFR	TFP Growth	Obs
Rice	9.2	1.1	148.8	−0.2	−0.2	4.1	4,840
Perennials	8.7	0.7	126.7	−0.6	−0.4	−2.1	236
Other	9.0	1.2	155.4	−0.4	−0.4	−3.4	1,063
Total	9.2	1.1	149.1	–	–	2.6	6,139

Notes: Observations are at the farm-year level. Output is reported as the log of total agricultural output using real prices common across farms to sum different crops. Land is reported in acres. Labor is reported in number of effective worker days. TFP and TFPR are reported in logs and constructed following the model. TFP and TPFR in both the South and the North are normalized by the mean in each year. TFP Growth is calculated over a two-year period as $TFP\ Growth_t = 100 \times (TFP_{f,t} - TFP_{f,t-2})/[0.5(TFP_{f,t} + TFP_{f,t-2})]$ before TFP is normalized where t is the calendar year.

There are stark differences within regions between crops. In the South, perennials tend to outperform the other crop types in terms of production, productivity, and growth. This is consistent with the fact that perennials are cash crops that incentivize investment into cropping. Among perennials, coffee is the most important. In contrast, rice and other crops are more likely to be food crops for the household's own consumption, and underperform compared to perennials. In contrast, in the North rice tends to overperform relative to the other crops. Farms in the North also tend to be smaller, in terms of land and labor inputs and output, and lower growth. These differences motivate our main quantitative experiment.

Land quality differences. Our main exercise involves quantifying the model to examine the productivity gap between north and south Vietnam. The productivity gap between the North and South is unlikely to be explained by differences in land suitability. Appendix B.3 examines the potential role of land quality differences between the South and the North using potential yield data from the Global Agro-Ecological Zones analyzed in Adamopoulos and Restuccia (2022). We find minimal differences in land quality between the South and the North and if anything, the North features slightly higher land quality than the South.

5 Calibration

We calibrate a benchmark economy with distortions to match characteristics of south Vietnam that we observe in the data. Our main experiment in the next section adjusts distortions in the benchmark economy to match the higher measured distortions in the North.

5.1 Calibration Strategy

The model has eleven parameters that are common to all crops $\{L, N, \beta, \xi, \gamma, \alpha, \theta, \lambda, \psi, \zeta, \rho\}$, three sets of crop-specific parameters $\{\eta^i, \kappa^i, \varphi^i\}$, and three distributions $\{\Phi_z, \Phi_v, \Phi_\varepsilon\}$. We calibrate the crop-specific parameters to match the three farm types in Section 4.

Preliminaries. A period is set to one year. The discount factor is set to $\beta = 0.96$ to match a discount rate of 4%. The total mass of households is set to $N = 2$ such that there is a unit mass of farm managers and workers. The mass of land is set to $L = 2.77$ corresponding to an average farm size of 2.77 acres in south Vietnam. The span-of-control parameter is set to $\gamma = 0.7$, implying the profit share of farm managers is 30%, which reflects the combined return to the farm manager's labor on the farm and their management expertise (see, for example, Adamopoulos et al., 2022). The land share of output is set to $\alpha = 0.5$ based on the land share from Ayerst et al. (2020). Finally, the survival rate is set to $\xi = 0.955$ to match an exit rate of households from cropping of 1.2% and the implied inter-generational transfer of the farm of 3.3% in the data, which is based on the minimum and average ages of the head of household of 20 and 50 years old.

The three distributions describe the distributions of the permanent farmer productivity Φ_z, the idiosyncratic component of productivity Φ_v, and the idiosyncratic component of distortions Φ_ε. The permanent farmer productivity z takes five values while the idiosyncratic component of productivity v and distortions Φ_ε take fifteen values. We parameterize all three distributions with a log-normal distribution and dispersion parameters $\{\sigma_z, \sigma_v, \sigma_\varepsilon\}$ and with node ranges between two standard deviations above and below the mean value. For computation, we restrict the maximum farmer ability to λ^{99} and note that fewer than one in ten thousand farmers are above the 25th node of the ability distribution in the stationary equilibrium. The results are not sensitive to the number of grid points used for z, v, ε, or a.

Direct calibration of distortions. Following equation (1) in the model, the parameters related to distortions can be estimated by regressing measured farm-level TFPR on TFP and fixed effects for the farm's crop type. Table 2 summarizes the estimated distortions in south and north Vietnam.

We use the estimated coefficients to parameterize the institutional distortions in the South. We set the elasticity $\rho = 0.856$ of distortions to TFP to the value of the coefficient on TFP from the regression, which is consistent with the model-implied elasticity. Note that the estimated ρ is higher in the North. We normalize the crop-specific distortion of rice to one, $\varphi^R = 1$ and set the crop-specific distortions for perennial farmers to $\varphi^P = 1.61$ and for other crop farmers to $\varphi^O = 1.12$ using the relationship $\varphi^i = \exp(-\text{Coefficient}^i/(1-\gamma))$ implied by the model. The estimated coefficients indicate that distortions disincentivize production

Table 2: Estimating Distortions in the South and the North

	National	South	North
	(1)	(2)	(3)
	log TFPR	log TFPR	log TFPR
log TFP	0.907***	0.856***	0.964***
	(0.00473)	(0.00717)	(0.00491)
Perennials	−0.0989***	−0.142***	0.117***
	(0.0167)	(0.0178)	(0.0301)
Other	0.00250	−0.0350	0.0262
	(0.0142)	(0.0230)	(0.0183)
Year FE	Yes	Yes	Yes
R^2	0.907	0.905	0.917
Observations	10526	4387	6139

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the household level are included in parentheses. All regressions include year fixed effects.

by rice farmers (through higher τ) compared to perennial or other crop farmers. We set the standard deviation of the idiosyncratic component of distortions ε to match the standard deviation of the residual from the regression. The final parameter characterizing distortions is the level $\bar{\tau}$, which does not impact misallocation since it affects all farmers equally. However, the level $\bar{\tau}$ does affect profits and, consequently, the incentives for farmers to invest. We set $\bar{\tau}$ such that the average value of TFPR is equal to one and hold the value of $\bar{\tau}$ constant in the counterfactual economies. This is a conservative assumption since allowing the value of $\bar{\tau}$ to adjust such that TFPR is equal to one in the counterfactual economy results in lower productivity and slower farm growth relative to our baseline experiment.

We set the government-imposed restriction $\omega = 23\%$ for south Vietnam. In the data, farmers with multiple plots may report that only some plots face restriction while crop restrictions are a binary variable in the model. To construct the data moment, we take a land-weighted average of crop-restrictions for each farmer and then average this value over all farmers in south Vietnam in the dataset. This implies, for example, a farmer in the data with one-third of their land restricted is captured in the model by three farmers with one farmer having the entirety of their crop choice restricted and the other two being unrestricted. The comparison between the North and the South remains similar with alternative constructions of this moment. For example, the share of farmers that face crop restrictions on any plot is 40.5% in the South and 61.5% in the North, whereas the share of farmers with crop restrictions on greater than 50% of the land is 34.3% in the South and 51.6% in the North.

Jointly calibrated parameters. The remaining parameters are calibrated in two stages to match the moments in Table 3. The first stage exploits the fact that the preference shifters η^i can always be set such that the model exactly matches the farm share by crop in the data, regardless of the other parameter values. As a result, holding the distribution of crops fixed, the parameters $\{\lambda, \psi, \zeta, \kappa^i, \sigma_z, \sigma_v\}$ with κ^R normalized to one, are jointly calibrated to minimize the sum of squared errors between data moments and model moments constructed from simulated data (described below). The second stage involves calibrating the parameters describing the crop-specific preferences η^i, where η^R is normalized to one, and the preference curvature θ. As in the first stage, the crop-specific preferences η^i are always set such that the crop share matches the data. The value of the shape parameter θ on the distribution of farmer preferences is chosen to minimize the magnitude of the crop-specific preferences, given by $\sum_i (\eta^i - 1)^2$. In this regard, the final moment is chosen to treat the crop-specific preference as a residual and minimize the out-of-model factors that affect crop choice.

Other than the farm share by crop, the model moments are calculated using simulation data generated for 10,000 farmers in the stationary equilibrium. We initialize the simulation data for the 10,000 farmers using the stationary distribution of crop types i, permanent abilities z, and ability nodes h. We then allow the productivity of the farmers to evolve as in the stationary equilibrium—accounting for farmers transitioning to higher ability and random shocks (v, ε)—and allow for exit and entry of farmers for 103 periods. Finally, we drop the first 100 periods and construct model moments over the final three-period window following the same procedure in the data. For comparability, we winsorize the top and bottom

Table 3: Calibration Moments

	Model	Data
Avg TFP Growth (%)	6.25	6.23
Std TFP Growth	76.5	75.2
Std log TFP	1.01	1.00
Reg coefficient: growth on log TFP	−34.2	−34.4
Top 10% Land Share (%)	35.5	41.2
Relative Meas. TFP	(1.00 , 1.21 , 0.74)	(1.00 , 1.20 , 0.75)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(49.1 , 33.1 , 17.8)

Notes: Order of moments corresponds to rice, perennial and other crop farms when applicable. Farm share by crop is calculated based on the equilibrium value. All other moments are calculated using a simulation of 10,000 farms. Avg Growth and Std. Growth are calculated for growth over a two-year (two-period) interval (i.e., growth is calculated from t to $t + 2$).

2% of the simulation data, remove exiting farms from the relevant statistics (e.g., growth), and calculate moments using only data for the first (t) and last ($t + 2$) periods such that there is a year gap in the growth statistics.

5.2 Calibration Moments

Our theory describes the evolution and distribution of productivities and how these relate to farm crop decisions and the institutional environment. We leverage the micro data to construct moments that describe the joint distributions of TFP and growth as well as crop-specific differences across farms to discipline the calibration in the context of the measured distortions. We discuss the construction of the moments and closely related parameters below for intuition. Additional details on the sensitivity of the model moments to the calibrated parameters are provided in Appendix C.1.

Avg TFP growth. The moment reports the average growth of farm-level TFP. In the data, farm-level growth is calculated over a two-year period using measured TFP as $g_{f,t} = (TFP_{f,t} - TFP_{f,t-2})/(0.5 * (TFP_{f,t} + TFP_{f,t-2}))$. The reported moment is the average over all farm-years. In the simulated data, we similarly construct the growth in TFP from t to

$t+2$ and report the average value over all farms that remain active into $t+2$. The moment is closely related to the costs and benefits of the ability improvement technology (λ, ψ, ζ) which dictates persistent farm-level growth.

Std TFP growth. The moment reports the standard deviation of $g_{f,t}$ calculated in the previous moment across all active farms in both the empirical and simulated data. The moment is closely related to the idiosyncratic component of productivity through σ_v.

Std log TFP. The moment reports the standard deviation of $\text{TFP}_{f,t}$ across all active farms in both the empirical and simulated data. The moment acts as a residual measure of farm-level TFP differences to discipline the dispersion in the permanent component of productivity through σ_z but is also related to the parameters that dictate the productivity distribution, such as the ability improvement technology (λ, ψ, ζ) and the idiosyncratic component of productivity through σ_v.

Reg coefficient: TFP growth on log TFP. The moment measures the regression coefficient from farm-level TFP growth on log TFP. The empirical specification is given by $g_{f,t} = \nu \ln \text{TFP}_{f,t-2} + \Gamma_t + \epsilon_{f,t}$ where Γ_t is a year fixed effect and ν is the reported moment. The moment is calculated similarly using the simulated data, without the time fixed effect. The moment helps discipline the curvature of the cost function since higher curvature ζ implies investment is less elastic to incremental profitability. Correlated distortions lead to a flattening of profitability at higher ability levels implying that higher curvature ζ increases relative investment by higher ability farms. Less negative estimates of ν then correspond to higher values of ζ. The moment is also closely related to other parameters that dictate the productivity distribution, such as $(\lambda, \sigma_z, \sigma_v)$.

Top 10% land share. The moment is calculated as the share of land held by the 10% largest farmers (by land size) in the empirical and simulated data. A farmer's land size is closely related to both productivity and distortions. The moment helps discipline the steepness of productivity increases through ability improvement. Consequently, the moment is also closely related to the ability improvement technology (λ, ψ, ζ). Intuitively, a more concentrated distribution of land implies a more skewed productivity distribution and leads to a larger ability improvement (higher λ) by fewer farmers (higher costs ψ).

Relative measured TFP. In the data, the moment is calculated by regressing farm level TFP on crop fixed effects, $\ln \mathrm{TFP}_{f,t} = k^P \times \mathrm{Peren.}_{f,t} + k^O \times \mathrm{Other}_{f,t} + \Gamma_t + \epsilon_{f,t}$ where Γ_t is a year fixed effect. The empirical estimates $\exp(k^P)$ and $\exp(k^O)$ are then used as targets where the value for rice is normalized to one. In the simulated data, the corresponding moments are calculated as average TFP of perennial and other crop farmers normalized by the average TFP of rice farmers.

Farm share by crop. In the data, the moment is calculated as the number of farmers that qualify as Rice, Perennial, and Other farmers based on the definition in Section 4. In the model, the moment is calculated analytically as the share of farmers that choose to grow each crop i. For a given share of government-imposed crop restrictions ω and farm values $\bar{V}^i_{z,h}$, the crop-specific preferences η^i can be used to match directly the empirical distribution of crops using equation (2).

5.3 Parameters

Table 4 summarizes the calibrated model parameters. The ability improvement λ and the crop-specific productivity component κ need to be scaled by a factor of $1-\gamma$ to be converted into TFP values. This implies that the increase in TFP from improving ability is 13% ($\lambda^{1-\gamma} \approx 1.13$) and that perennial (other) farmers are 8% more (26% less) productive than rice farmers, all else equal. Relative to Adamopoulos and Restuccia (2020), we find larger differences in the crop-specific component of productivity despite smaller differences in measured TFP. This is driven by more productive farmers selecting into cash crops in Adamopoulos and

Restuccia (2020). In contrast, we find substantial overlap in the productivity distribution of farm types leading us to model crop selection at the time of entry.

Table 4: Model Parameters

Parameter		Value
Discount Factor	β	0.96
Survival Rate	ξ	0.955
Land	L	2.77
Span-of-Control	γ	0.7
Land Share	α	0.5
Investment Level	ψ	3.79
Investment Curvature	ζ	2.81
Ability Step Size	λ	1.51
Crop Preference Elasticity	θ	1.85
Preference Shifter	η^i	(1.00 , 0.66 , 0.83)
Crop-Specific Productivity	κ^i	(1.00 , 1.28 , 0.38)
Permanent Productivity	σ_z	1.47
Random Productivity	σ_v	2.32
Elasticity	ρ	0.856
Crop-Specific Distortion	φ^i	(1.00 , 1.61 , 1.12)
Random Distortion	σ_ε	0.92
Crop Restriction	ω	0.23

In addition to productivity, the crop-specific distortions φ^i tend to incentivize farmers to grow crops other than rice, where recall that higher values of φ corresponds to lower distortions τ. However, the allocation of farmers across crops is partly offset by the government-imposed crop-restrictions ω, which increases the share of rice farmers.

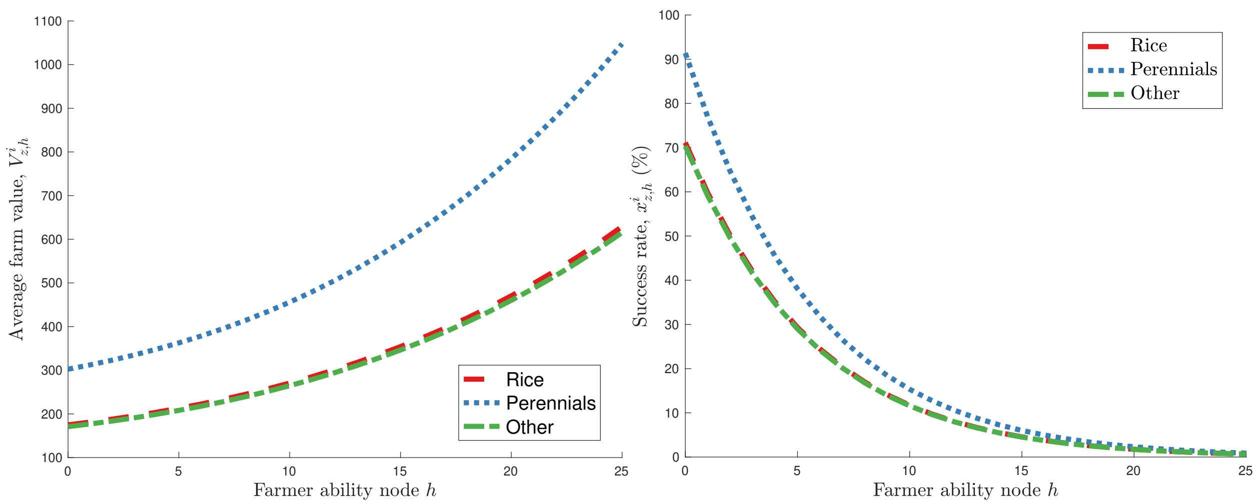
The estimated curvature ζ on the ability investment is higher than typically found in the manufacturing sector, which is closer to quadratic (e.g., Bento and Restuccia, 2017; Acemoglu et al., 2018; Ayerst, 2022). This implies that farm investment is comparatively less elastic to changes in profitability. The estimated curvature reflects, in part, that higher productivity farms tend to grow slower than lower productivity farms in the data. Smaller values of ζ would imply that productivity growth drops off more steeply for high productivity (high ability) farms than is observed in the data. We examine the sensitivity of the results to using quadratic curvature in Section 6.5.

The estimated crop-specific preference shifter η^i implies that, on average, households face a larger utility cost for perennial farms than the other crops. This is consistent with perennials requiring substantial investment by households in addition to time before maturing and generating income.

5.4 Value and Policy Function

Figure 5 plots the value and policy functions for farmers in each of the three crop types. The value and policy functions are averaged across idiosyncratic shocks (v, ε) and plotted for a common permanent productivity $z = 1$.

Figure 5: Value and Policy Function by Crop Specialization



(a) Value functions (b) Policy functions

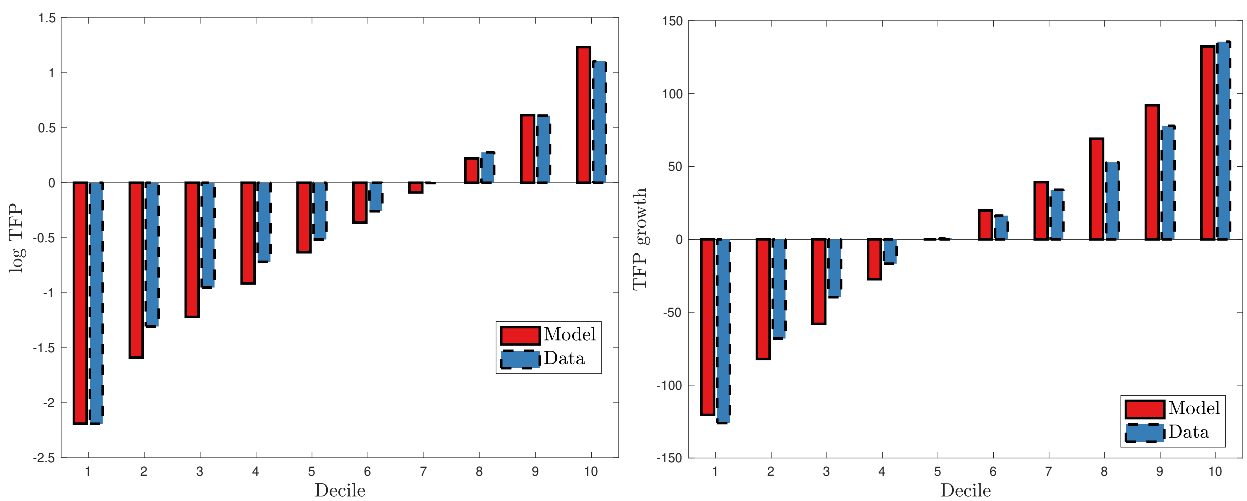
The value functions of the three crops reflect differences in relative profitability stemming from differences in productivity κ^i and distortions φ^i. Despite differences in κ^i and φ^i, the value of rice and other crop farmers are similar because the two parameters have offsetting effects on profitability. The success rate x, and the corresponding investment in ability, are driven by the incremental increase in farm value that farmers receive from successfully improving ability. More productive and less distorted farmers invest more in improving ability because of the complementarity in profits between ability, other sources of productivity (i.e., the permanent farmer component z or the crop-specific component κ^i), and lower distortions. However, differences in the policy functions across crops decrease at higher abilities because distortions become a limiting factor that discourage further investment.

A key feature of the policy function is that the success rate of improving ability declines as farmers become more productive. The correlation of distortions with productivity implies that the incremental increase in profitability is lower than that of investment costs as farmers improve ability. Models of firm dynamics (e.g., Klette and Kortum, 2004) typically assume that profits and investment costs grow at the same rate in order for Gibrat's law to hold in equilibrium. In contrast, we find that more productive farms tend to grow more slowly as evident by the negative relationship between TFP growth and farm productivity in Table 3.

5.5 Other Moments and Goodness-of-Fit

Figure 6 compares the median TFP and TFP growth by percentile in the empirical and simulated data. Despite only targeting the dispersion of productivity and growth in the calibration, the simulated distribution fits the empirical distribution well.

Figure 6: Farm Productivity Distribution

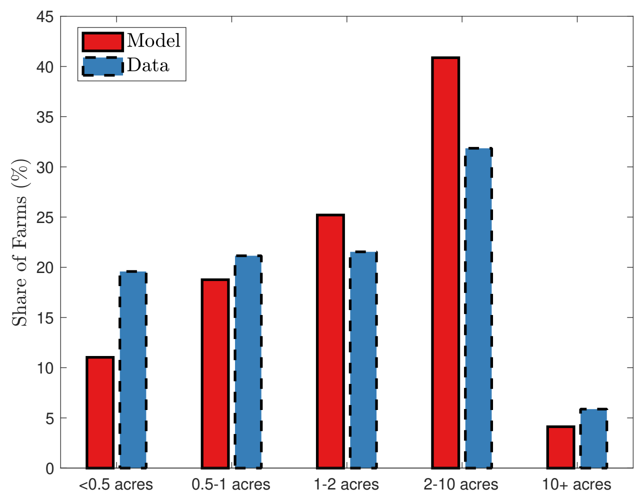


(a) TFP by decile (b) Growth by decile

Notes: Panel (a) reports log TFP for the median of each decile, i.e., the percentiles 5, 15, etc. Panel (b) similarly reports TFP growth for the median farm of each decile.

Figure 7 provides a comparison of the farm land size in the simulated model data and the data. The calibration sets the aggregate quantity L of land to match the average farm size in the model, but does not target the distribution of farms by land size. While the model overstates the share of farmers in the 2-10 acres category and understates the share of very small farmers (less than 0.5 acres), overall the distribution of farm sizes across size bins in the simulated data closely matches the empirical data.

Figure 7: Farm Land-Size Distribution



Notes: Share of farms in each farm land-size class. Land size refers to cultivated land by the farm.

Table 5 compares other data moments with the corresponding moment constructed in the simulated data. The first set of moments validate the modeled distortions. The first two moments are the regression coefficients from Table 2 that are directly targeted. The third moment shows that the standard deviation of TFPR implied by the calibrated distortions closely matches with the data. This provides a check on the choice to target the idiosyncratic component of distortions σ_ε to the regression residual. The fourth moment shows that the empirical autocorrelation of farm-level TFPR is smaller than that implied by the model, supporting that ε varies over time rather than being more permanent for the farmer. The fifth moment considers a simple experiment in both the empirical and simulated data. Holding the distribution of productivities $s^i_{f,t}$ fixed, we calculate the potential gain in aggregate productivity from moving to the efficient allocation of land and labor (as in, for example

Table 5: Other Model Moments

	Model	Data
Elasticity TFP,TFPR*	0.86	0.86
Crop-Specific FE*	(0.000 , −0.142 , −0.035)	(0.000 , −0.142 , −0.035)
Std log TFPR	0.89	0.87
Autocorrelation TFPR	0.57	0.34
Gains from Reallocation (%)	68.3	62.0
Relative Output	(1.00 , 1.69 , 0.73)	(1.00 , 1.97 , 0.77)
Relative Land	(1.00 , 1.69 , 0.96)	(1.00 , 1.96 , 0.39)
Relative Labor	(1.00 , 1.69 , 0.96)	(1.00 , 1.65 , 1.07)
Std log Output	1.48	1.47
Std log Land	0.98	1.21
Std log Labor	0.98	1.07

Notes: the order of moments corresponds to rice, perennial and other crop farms when applicable. Moments with a * indicate moments that are directly targeted in the calibration.

Hsieh and Klenow, 2009). The result supports the fit of the joint TFP and TFPR distribution in the model and data, which determines the gains from reallocation.

The second set of moments show that the model replicates the distribution of output, land, and labor across both crops and farms. Relative output, land, and labor are calculated in the simulated data as the average of the respective outcomes for each farm type. In the data, the corresponding moments are based on the regressions reported in Appendix B.1. The moments are indirectly related to the distribution of productivities across crops and farms in the calibration. The final set of moments compares the standard deviations of the three outcomes in the data and model. Similar to Figure 7, the moments show that the calibrated model is able to replicate the overall size distribution of farms in the data.

6 Quantitative Analysis

The agricultural sector in north Vietnam is comparatively more distorted than south Vietnam. We also observe considerably less farm dynamism in north Vietnam and slower growth in agricultural productivity. We assess the importance of institutional distortions in accounting for these differences by imposing distortions that reflect conditions in north Vietnam on the benchmark economy, which is calibrated to match south Vietnam. Appendix C.2 reports a fully recalibrated model to the north Vietnam data.

6.1 Counterfactual Distortions

The benchmark economy has four parameters related to distortions: (1) the elasticity of distortions to farm-level productivity ρ; (2) crop-specific distortions φ^i; (3) the random component of distortions σ_ε; and (4) the government-imposed crop restriction ω. Table 6 summarizes the values estimated for the counterfactual experiment. Other parameters, including the random component of distortions, are held fixed at the benchmark economy values.

Table 6: Counterfactual Distortions

		Benchmark (South)	Counterfactual (North)
Elasticity	ρ	0.86	0.96
Crop-Specific Distortion	φ^i	(1.00 , 1.61 , 1.12)	(1.00 , 0.68 , 0.92)
Crop Restriction (%)	ω	23	43

Notes: Distortions are ordered for Rice, Perennial, Other crop farm types. Crop-specific distortions are implied by the coefficient estimates in Table 2 as $\varphi^i = \exp(-\text{Coefficient}^i/(1-\gamma))$.

We follow the same procedure as in the benchmark economy calibration and base the distortions on the regression coefficients in Table 2. Relative to south Vietnam, distortions are more correlated with farm-level productivity, reflecting the greater difficulty productive farmers face to expand their farm size, and tend to discourage perennial crops. We set $\omega = 43\%$ to reflect the share of farmers reporting crop restriction in north Vietnam in our data. Finally, we hold the idiosyncratic component of distortions, governed by σ_ε, fixed at the benchmark economy level as this is the parameter most likely to capture measurement error. Since we compare two distorted economies rather than quantifying the level of misallocation, measurement error is less likely to be a factor since it is reflected in both measures of distortions. Appendix C.3 reports the overall cost of misallocation relative to the undistorted economy.

6.2 Comparison with the Data

We start by examining whether the counterfactual economy moves the model moments closer to the north Vietnam moments compared to the benchmark economy. Table 7 compares the calibration moments and agricultural productivity in the benchmark and counterfactual economies as well as the north Vietnam data. We do not expect the counterfactual model to replicate fully the data moments for the North since other factors (e.g., average farm size) differ across regions and could impact the moments. Nevertheless, Table 7 shows that the counterfactual economy is more similar to the data moments for north Vietnam than the benchmark economy.

Table 7: Comparing Counterfactual Moments with the Data for the North

	Benchmark	Counterfactual	Data
Productivity	1.00	0.54	0.42
Avg Growth (%)	6.25	4.46	2.62
Std Growth (%)	76.5	76.3	89.2
Std log TFP	1.01	0.93	0.84
Reg coefficient: growth on log TFP	−34.2	−39.3	−48.2
Top 10% Land Share (%)	35.5	30.6	38.3
Relative Measured TFP	(1.00 , 1.21 , 0.74)	(1.00 , 1.01 , 0.74)	(1.00 , 0.69 , 0.8)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(74.8 , 7.2 , 17.9)	(75.1 , 5.0 , 19.9)

Notes: When applicable, moments are ordered for Rice, Perennial, Other crop farm types.

Our main result is the implied productivity gap between the counterfactual and benchmark economy, which is a measure of how much of the observed productivity gap can be explained by differences in the distortions between the North and South. We find that aggregate TFP in the counterfactual economy (North) is 54% of the benchmark economy (South), implying that the model accounts for just over two thirds ($71\% \approx \log(0.54)/\log(0.42)$) of the productivity gap between the North and the South.

In addition, the model explains around half ($50\% \approx 1.8/3.6$) of the gap in the average productivity growth rate of farmers between the South and the North. The model is able to account almost entirely for differences in the farm share by crop in the data as well as around half of the relative measured TFP of perennial farmers. The model also explains around half of the gap for the standard deviation of log TFP and around one third of the regression coefficient of growth on log TFP. The results are reassuring as the theory shows how the interactions between distortions and crop decisions affect productivity and the model fits the direction of the data in these dimensions.

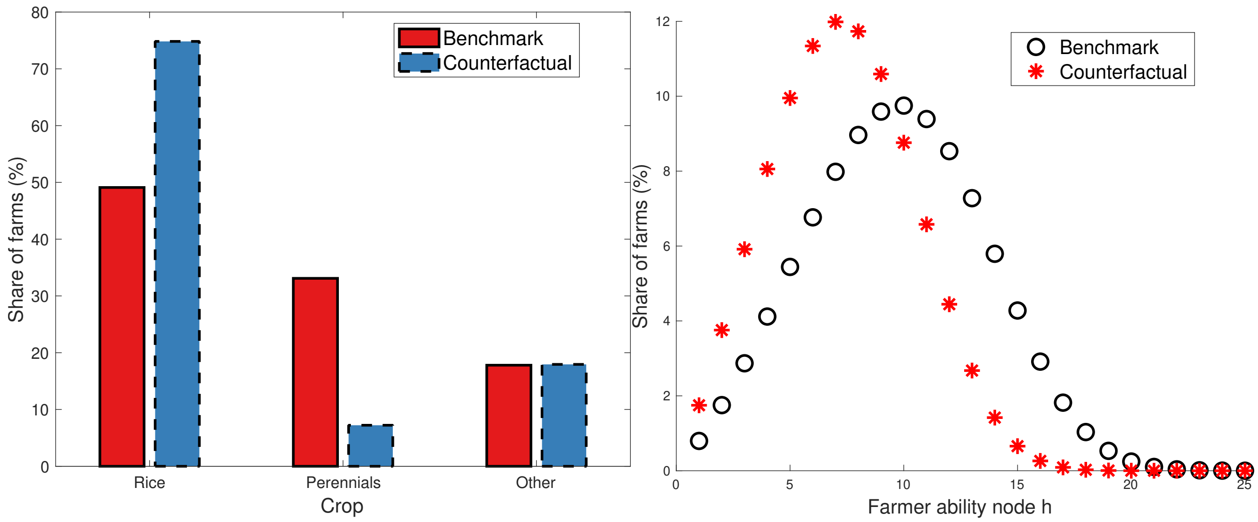
The model performs less well in two dimensions. First, it does not generate an increase in the standard deviation of growth, which is relatively unchanged in the counterfactual economy. This is because the standard deviation of growth is mostly driven by the idiosyncratic dispersion in productivity σ_v (Appendix C.1), which is held fixed in the counterfactual economy. Second, the top 10% land share declines relative to the benchmark economy, which is consistent with the difference in the North and South. That is, the top 10% share declines by around 4.9% ($= 35.5 - 30.6$) between the benchmark and counterfactual economies, which is slightly larger than the 2.9% ($= 41.2 - 38.3$ from Tables 3 and 7) gap between the South and North in the data.

6.3 Drivers of the North-South Productivity Gap

Differences in measured distortions between north and south Vietnam produce a productivity loss of 46%. What features of the model account for this productivity loss? Following equation (3), productivity in the model depends on factor misallocation, the crop distribution, and the ability distribution. Figure 8 compares the crop and ability distributions in the benchmark and counterfactual economies. Consistent with evidence in north and south Vietnam (Ayerst et al., 2020), the figure shows that the ability distribution has more mass at higher productivities in the South.

To better understand the three components of productivity, we consider three simple

Figure 8: Farm Distributions in Benchmark and Counterfactual Economies



(a) Crop Distribution

(b) Ability Distribution

experiments to decompose the relative contributions of factor misallocation, crop choice, and farm ability. Table 8 summarizes the loss in aggregate productivity from changing channels individually from the benchmark economy to match the counterfactual economy. The sum of the losses does not equal the total gap between the benchmark and counterfactual economy because of interactions between the channels. For example, changes in the ability or crop distributions also affect the potential scope for factor misallocation through their effect on the productivity distribution. We discuss each channel and its calculation below.

Table 8: Output Loss by Channel

	Change in Output (%)
Factor Misallocation	-21.7
Crop Choice	-7.3
Farm Ability	-35.8
Sum of Channels	-64.8
Total Output Loss	-46.3

Factor misallocation. We calculate the loss from factor misallocation as the change in aggregate output when distortions, $\tau^i_{f,t}$, are adjusted to match the counterfactual economy but the crop and ability distributions are held fixed at the benchmark distributions. Starting from the distribution of farm-level productivities $s^i_{f,t}$ in the benchmark economy, we recalculate the distortions $\tau^i_{f,t}$ that farmer f would receive with the counterfactual correlation ρ and crop-specific distortions φ^i. We find that factor misallocation lowers agricultural output by -21.7%, accounting for just under half of the gap between the counterfactual and benchmark economies.

Partitioning the interaction effects proportionately to each channel, factor misallocation accounts for one-third ($\approx -21.7/-64.8$) of the resulting productivity loss in the counterfactual economy. Factor misallocation has a negative interaction with the other two channels explaining why the sum of the losses from the individual channels is larger than the total loss in productivity. This is because factor misallocation has a larger effect on aggregate productivity when the productivity distribution is more dispersed. All else equal, the scope for factor misallocation is smaller in the counterfactual economy where TFP dispersion is smaller.

Crop distribution. We calculate the loss from the crop distribution as the change in aggregate output when the share of crops are adjusted to match the counterfactual economy. We assume that the ability distribution of farmers conditional on crop i and permanent farmer productivity z is the same as in the benchmark economy implying that the change in the crop distribution also affects the ability distribution. Consequently, average ability falls since perennial farmers are, on average, higher ability than rice farmers and the experiment redistributes around 25% of farmers from perennials to rice. We find that the change in the crop distribution has a relatively small contribution to the overall gap between the counterfactual and benchmark economies compared with the other channels, nevertheless the output loss from the change in the crop distribution is a non-trivial -7.3%.

Farmer ability. We calculate the loss from farmer ability as the change in aggregate output when the ability distribution is adjusted to match the counterfactual economy. As with the crop distribution, we assume that farmer ability is conditional on the farmer's permanent productivity z and crop i, such that the ability distribution does not fully match the counterfactual economy. The ability distribution in the counterfactual economy results from lower investment by farmers due to more correlated distortions, which makes higher ability levels less profitable. We find that the change in farm ability generates a loss in agricultural output of -35.8%, accounting for around three quarters of the productivity gap between the counterfactual and benchmark economies. The effect of crop choice and farm ability together, representing the broader effects of misallocation, account for two-thirds of the sum productivity loss, double the impact of factor misallocation.

6.4 Role of Individual Distortions

We also examine the role of the individual distortions, rather than channels. We measure the impact on output of individual distortions from unilaterally changing ρ, φ^i, or ω in the benchmark economy to match the North. Table 9 summarizes the results.

Table 9: Output Loss from Individual Distortions

	ρ	φ^i	ω	$(\rho, \varphi^i, \omega)$
Change in Output (%)	-44.9	-6.8	-1.5	-46.3

The main driver of the gap between the benchmark and counterfactual economies is correlated distortions. Correlated distortions have a large impact on factor misallocation by reallocating resources from high productivity to low productivity farms. Correlated distortions also flatten the increase in profits associated with increasing farm productivity, resulting in less incentives for farmers to invest in ability or select crops based on market factors as opposed to preferences.

The crop-specific distortions have a more moderate effect on the productivity gap between the benchmark and counterfactual economies. Crop-specific distortions increase factor misallocation by reallocating resources across different farm types. Crop-specific distortions also affects the relative incentives for farmers to invest in improving ability since it changes the relative profitability of crops. Finally, crop-specific distortions affect the crop distribution through changing the relative market value of farm types.

The government-imposed crop restriction has the smallest impact on productivity. Part of the reason is that crop restrictions are implemented before farmers make crop choices implying that some farmers would choose to grow rice regardless of the restriction. Since around half of farmers grow rice in the benchmark economy, this reduces the impact by a comparable amount. As a back-of-the-envelope calculation, the change in productivity is approximately equal to reducing the productivity of 7% (the change in ω times the share of perennial farmers) of farmers by 21% (the measured productivity of of perennials farmers relative to rice farmers). This calculation highlights the limited impact of crop restrictions on aggregate productivity despite the relatively large measured differences across crops.

6.5 Robustness

The results show that the interactions between distortions and farm dynamics lead to large productivity differences between north and south Vietnam. Higher distortions both prevent higher ability farmers from increasing production and disincentivize investment by farmers, magnifying the overall costs of misallocation. We consider two sets of exercises related to the calibrated ability distribution to examine the robustness of our results. Table 10 reports the productivity gap explained by the model using the alternative calibration parameters.

Table 10: Robustness of Main Results to Alternative Calibrations

	Relative Counterfactual Output (%)
Baseline	53.7
Alternative calibrations:	
Fix investment-cost curvature $\zeta = 2$	52.4
Avg growth target $6.23\% - 2\%$	56.5
Avg growth target $6.23\% - 4\%$	61.5

Notes: Compares the agricultural output (productivity) effect of the counterfactual relative to the benchmark economies in the baseline and alternative calibrations of the model.

First, we consider a re-calibration of the model that fixes the ability investment curvature to $\zeta = 2$. The remaining parameters are re-calibrated to match the moments in Table 3. The productivity gap in the re-calibrated model implies a larger gap between the counterfactual and benchmark economies than in the baseline experiment.

Second, we consider a re-calibration of the model using alternative targets for the average growth rate in south Vietnam. Lowering the targeted growth rate results in a more compressed ability distribution relative to the benchmark calibration since the model requires that farms either grow by less or less frequently to match the moment. A concern in our baseline calibration is that part of the growth captured in the target represents economy-wide factors (e.g., technology improvements) unrelated to the ability improvements in the model. If these other factors are large, then the ability distribution may be more compact than assumed in our baseline calibration and the results overstated. To give a sense of the quantitative importance of this factor, we re-calibrate the model using targets for average growth rates that are two and four percentage points lower than in the baseline calibration. Table 10 shows that the lower growth targets result in smaller productivity gaps, consistent with the importance of the ability channel in the main results. Despite the relatively drastic changes in the targeted growth rate, the productivity gap explained by the model remains economically significant in both cases. Lowering the targeted average growth rate by two percentage points, around one third of the targeted value, increases the relative productivity of the counterfactual economy by only 2.8 percentage points compared with the baseline experiment. This implies that the model goes from accounting 71% of the productivity gap to 66% of the productivity gap. Decreasing the targeted growth rate by four percentage points lowers the explanatory power of the model to 56% of the productivity gap between the North and the South.

7 Conclusion

We extend a standard model of heterogeneous production to capture two important aspects of farm dynamics in developing countries: crop choices and productivity investments. Using panel farm-level data from Vietnam, we apply this framework to study the effect of differences in institutional distortions between north and south Vietnam. Through the lens of the model, measured distortions in the North relative to the South account for 71% of the productivity gap, which represents a substantial 46% productivity loss, and around half the difference in farm dynamics, as measured by farm productivity growth. Farm ability and crop choice (dynamic misallocation) account for almost two thirds of the productivity loss, with the remaining one third coming through the standard channel of factor misallocation. Decomposing the sources of the productivity loss, we find that the key institutional feature is the higher elasticity of distortions to farmer productivity in the North, which effectively captures the inability of farmers to expand farm size in the North relative to those in the South.

Our results in the context of Vietnam agriculture provide novel quantitative evidence of the broader effects of misallocation emphasized in Restuccia and Rogerson (2017), especially when the pattern of misallocation most heavily penalizes the more productive producers, effectively lowering the return to productivity investment and growth. A promising area for future work is to examine the effects of distortions on producer dynamics in other contexts, joining recent efforts assessing the role of size-dependent policies on innovation and growth (Aghion et al., 2021; Akcigit et al., 2022). It will also be insightful to study the dynamic consequences of misallocation in the context of episodes of reform in either agriculture (Chari et al., 2021; Chen et al., 2022; Beg, 2022) or industry (Asturias et al., 2023), as well as episodes of trade reform (Pavcnik, 2002).

Finally, more work is needed in identifying the specific channels of dynamic misallocation which can help facilitate a deeper understanding of the broader role of policies and reform, including the importance for productivity growth of technology adoption and diffusion, the adoption of improved managerial practices, and other productivity-enhancing investments at the producer level. We leave these exciting areas of research for future work.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation and growth. *American Economic Review*, 126(4):1374–1443.

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2022). Misallocation, selection, and productivity: A quantitative analysis with panel data from china. *Econometrica*, 90(3):1261–1282.

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. (2022). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. *The Review of Economic Studies*, 89(4):1629–1653.

Aghion, P., Bergeaud, A., and Van Reenen, J. (2021). The impact of regulation on innovation. Technical report, National Bureau of Economic Research.

Akcigit, U., Akgunduz, Y. E., Alp, H., Cilasun, S. M., and Quintero, J. M. (2022). Cost of size-dependent regulations: The role of informality and firm heterogeneity. Technical report, University of Chicago.

Akcigit, U., Alp, H., and Peters, M. (2021). Lack of selection and limits to delegation: firm dynamics in developing countries. *American Economic Review*, 111(1):231–75.

Asturias, J., Hur, S., Kehoe, T. J., and Ruhl, K. J. (2023). Firm entry and exit and aggregate growth. *American Economic Journal: Macroeconomics*, 15(1):48–105.

Ayerst, S. (2022). Distorted technology adoption. Technical report.

Ayerst, S., Brandt, L., and Restuccia, D. (2020). Market constraints, misallocation, and productivity in vietnam agriculture. *Food Policy*, 94:1–16.

Baily, M. N., Hulten, C., Campbell, D., Bresnahan, T., and Caves, R. E. (1992). Productivity dynamics in manufacturing plants. *Brookings papers on economic activity. Microeconomics*, 1992:187–267.

Beg, S. (2022). Digitization and development: Property rights security, and land and labor markets. *Journal of the European Economic Association*, 20(1):395–429.

Benjamin, D. and Brandt, L. (2004). *Agriculture and income distribution in rural Vietnam under economic reforms: a tale of two regions*, volume 842. World Bank Publications.

Bento, P. and Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.

Brandt, L., Le, D., Huong, G., Trang, C., Pham, G., Nguyen, N., and Luu, V. (2006). Land access, land markets, and their distributive implications in rural vietnam. *Department of Economics, University of Toronto, Toronto. Photocopy.*

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., Restuccia, D., and Santaeulàlia-Llopis, R. (2022). The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics*, 45:41–54.

Da-Rocha, J.-M., Tavares, M. M., and Restuccia, D. (2017). Policy distortions and aggregate productivity with endogenous establishment-level productivity. Technical report, National Bureau of Economic Research.

Davis, S. J., Haltiwanger, J. C., Schuh, S., et al. (1998). Job creation and destruction. *MIT Press Books*, 1.

Foster, L., Haltiwanger, J. C., and Krizan, C. J. (2001). Aggregate productivity growth: Lessons from microeconomic evidence. In *New developments in productivity analysis*, pages 303–372. University of Chicago Press.

Guner, N., Parkhomenko, A., and Ventura, G. (2018). Managers and productivity differences. *Review of Economic Dynamics*, 29:256–282.

Guner, N., Ventura, G., and Xu, Y. (2008). Macroeconomic implications of size-dependent policies. *Review of economic Dynamics*, 11(4):721–744.

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.

Hsieh, C.-T. and Klenow, P. J. (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics*, 129(3):1035–1084.

Key, N. (2019). Farm size and productivity growth in the united states corn belt. *Food Policy*, 84:186 – 195.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy*, 112(5):986—1918.

König, M., Storesletten, K., Song, Z., and Zilibotti, F. (2022). From imitation to innovation: Where is all that chinese r&d going? *Econometrica*, 90(4):1615–1654.

Le, K. (2020). Land use restrictions, misallocation in agriculture, and aggregate productivity in vietnam. *Journal of Development Economics*, 145:102465.

Lucas, R. E. (1978). On the size distribution of business firms. *Bell Journal of Economics*, 9:508–523.

Markussen, T. (2017). Land issues. In Tarp, F., editor, *Growth, Structural Transformation and Rural Change in Viet Nam: A Rising Dragon on the Move*, pages 139–157. Oxford University Press.

Markussen, T. and Tarp, F. (2014). Political connections and land-related investment in rural vietnam. *Journal of Development Economics*, 110:291–302.

Markussen, T., Tarp, F., and Van Den Broeck, K. (2011). The forgotten property rights: Evidence on land use rights in viet nam. *World Development*, 39:839–50.

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from chilean plants. *The Review of economic studies*, 69(1):245–276.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720.

Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–74.

Tarp, F. (2017). *Growth, Structural Transformation and Rural Change in Viet Nam: A Rising Dragon on the Move*. Oxford University Press.

Verhoogen, E. (2021). Firm-level upgrading in developing countries.

On-line Appendix

A Cropping Decision

Let $\bar{V}_z^i$ be the expected utility from consumption of choosing crop i. Then, the probability that household f chooses crop i is given by:

$$
\begin{aligned}
&= \Pr\left[\bar{V}_z^i b^i > \bar{V}_z^{i'} b^{i'} \forall i' \neq i\right], \\
&= \int_{\tilde{b}} \prod_{i' \neq i} \Pr\left[\bar{V}_z^i \tilde{b} > \bar{V}_z^{i'} b^{i'}\right] h(\tilde{b}) d\tilde{b}, \\
&= \int_{\tilde{b}} \prod_{i' \neq i} \exp\left\{-(\eta^{i'})^\theta \left(\frac{\bar{V}_z^i}{\bar{V}_z^{i'}}\tilde{b}\right)^{-\theta}\right\} \left[\theta(\eta^i)^\theta \tilde{b}^{-\theta-1} \exp\{-(\eta^i)^\theta (\tilde{b})^{-\theta}\}\right] d\tilde{b}, \\
&= \int_{\tilde{b}} \exp\left\{-\left(\sum_{i' \neq i}(\eta^{i'})^\theta \left(\frac{\bar{V}_z^i}{\bar{V}_z^{i'}}\right)^{-\theta}\right) \tilde{b}^{-\theta}\right\} \left[\theta(\eta^i)^\theta \tilde{b}^{-\theta-1} \exp\{-(\eta^i)^\theta \tilde{b}^{-\theta}\}\right] d\tilde{b}, \\
&= \int_{\tilde{b}} \exp\left\{-\left(\frac{\sum_{i' \neq i}(\bar{V}_z^{i'}\eta^{i'})^\theta}{(\bar{V}_z^i)^\theta}\right) \tilde{b}^{-\theta}\right\} \left[\theta(\eta^i)^\theta \tilde{b}^{-\theta-1} \exp\{-(\eta^i)^\theta \tilde{b}^{-\theta}\}\right] d\tilde{b}, \\
&= (\eta^i)^\theta \int_{\tilde{b}} \left[\theta \tilde{b}^{-\theta-1}\right] \exp\left\{-\left(\frac{\sum_{i' \neq i}(\bar{V}_z^{i'}\eta^{i'})^\theta}{(\bar{V}_z^i)^\theta} + \eta^i\right) \tilde{b}^{-\theta}\right\} d\tilde{b}, \\
&= (\eta^i)^\theta \frac{(\bar{V}_z^i)^\theta}{\sum_{i'}(\bar{V}_z^{i'}\eta^{i'})^\theta} \int_{\tilde{b}} \left[\theta \frac{\sum_{i'}(\bar{V}_z^{i'}\eta^{i'})^\theta}{(\bar{V}_z^i)^\theta} \tilde{b}^{-\theta-1}\right] \exp\left\{-\left(\frac{\sum_{i'}(\bar{V}_z^{i'}\eta^{i'})^\theta}{(\bar{V}_z^i)^\theta}\right) \tilde{b}^{-\theta}\right\} d\tilde{b}, \\
&= \frac{(\bar{V}_z^i \eta^i)^\theta}{\sum_{i'}(\bar{V}_z^{i'}\eta^{i'})^\theta}.
\end{aligned}
$$

B Data Details

We provide some statistics on farm differences by crop type, evidence of farm investment, and evidence on similar land quality between south and north Vietnam.

B.1 Differences by Farm Type

Tables B.1 and B.2 report cross-crop differences for output, land, labor, TFP, and TFP growth by farm type. The results are consistent with the summary statistics presented in the main text. In the South, perennials farmers tend to be larger in terms of both output and inputs, more productive, and higher growth. Other crop farmers tend to be smaller, at least in terms of output, and less productive. In the North, perennial farmers perform comparatively worse than rice farmers.

Table B.1: Farm Type Comparison in South Vietnam

	(1) log Output	(2) log Land	(3) log Labor	(4) log TFP	(5) TFP Growth
Perennials	0.679***	0.675***	0.499***	0.180***	6.028***
	(0.104)	(0.0836)	(0.0687)	(0.0608)	(1.723)
Other	−0.267***	−0.0935	0.0700	−0.287***	−2.973
	(0.103)	(0.100)	(0.0759)	(0.0609)	(2.610)
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	4406	4406	4406	4387	3485

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the household level are included in parentheses. All regressions include year fixed effects.

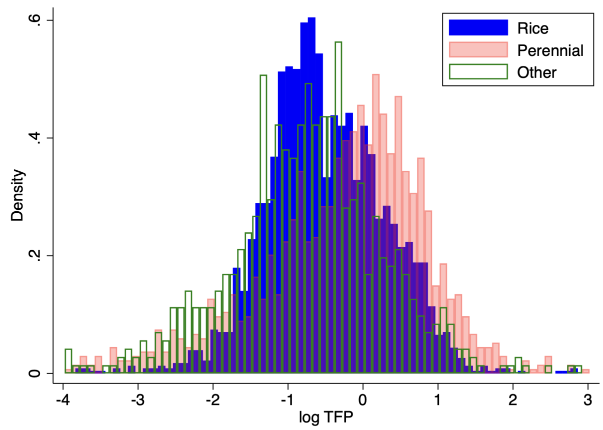
Table B.2: Farm Type Comparison in North Vietnam

	(1) log Output	(2) log Land	(3) log Labor	(4) log TFP	(5) TFP Growth
Perennials	−0.583***	−0.478***	−0.413***	−0.376***	−5.541
	(0.148)	(0.114)	(0.109)	(0.0993)	(4.859)
Other	−0.222***	−0.108	−0.113**	−0.222***	−7.064***
	(0.0686)	(0.0772)	(0.0527)	(0.0453)	(2.427)
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	6348	6348	6348	6139	5034

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the household level are included in parentheses. All regressions include year fixed effects.

Figure B.1 reports the TFP distribution by crop for south Vietnam. The figure highlights a key empirical observation that motivates how we model selection into different crops: a substantial productivity overlap between the three farm types. That is, while perennial farmers are on average more productive than the other farm types, there is a significant mass of perennial farmers that are less productive than the typical rice or other crop farmers. In contrast, selection based on farmer ability (as in Adamopoulos and Restuccia, 2020) would imply a discrete productivity cutoff in contrast with the data.

Figure B.1: Distributions of Farm TFP by Crop



B.2 Investment, Farm Productivity, and Distortions

We examine the relationship between farm-level productivity and growth and measures of investment. While we do not observe investment as broadly defined as in the model, there are two sets of variables that capture farm investment. The first set of variables relate to cash expenditure (Inv_t^{Cash}) and labor (Inv_t^{Lab}) investment in irrigation and soil and water conservation.[4] The second set of variables relate to extension services utilized by households to improve aspects of their farm. We construct a dummy variable Ext_t that takes the value one if the household used extension services relating to: (a) new seeds, varieties or breeds; (b) use of fertilizer; (c) irrigation; (d) pest infestation, blight problems; or (e) market information.

[4]To avoid dropping zeros, we transform the cash and labor investment expenditure to be equal to the log of one plus the investment.

Table B.3 reports how these measures of investment correlate with farm-level TFP growth ($g_{f,t}$). In the South, both investment and extension services are associated with faster growth, consistent with the model. The relationship is weaker in the North however where only extension services are associated with faster growth.

Table B.3: Investment and TFP Growth

	South			North		
	(1)	(2)	(3)	(4)	(5)	(6)
	$g_{f,t}$	$g_{f,t}$	$g_{f,t}$	$g_{f,t}$	$g_{f,t}$	$g_{f,t}$
$\log(1 + \text{Inv}^{Cash}_{t-2})$	1.895***			−0.790		
	(0.485)			(0.513)		
$\log(1 + \text{Inv}^{Lab}_{t-2})$		5.418***			−1.710**	
		(1.168)			(0.860)	
Ext_{t-2}			15.28***			8.567**
			(4.079)			(3.709)
$\log\ \text{TFP}_{t-2}$	−39.99***	−40.00***	−40.18***	−55.85***	−55.74***	−56.02***
	(1.435)	(1.432)	(1.465)	(1.749)	(1.743)	(1.747)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.261	0.262	0.262	0.320	0.320	0.320
Observations	3485	3485	3485	4883	4883	4883

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the household level are included in parentheses. TFP growth $g_{f,t}$ is calculated as $(\text{TFP}_{f,t} - \text{TFP}_{f,t-2})/(0.5(\text{TFP}_{f,t} + \text{TFP}_{f,t-2}))$

We also examine how investment decisions correlate with farm-level productivity and distortions. Tables B.4 summarizes the results. $\text{Inv}^{Cash}_t > 0$ and $\text{Inv}^{Lab}_t > 0$ take value one if the household invests in period t and zero otherwise. Both higher TFP and lower TFPR households are more likely to invest and engage in extensions services.

B.3 Differences in Land Quality

Table B.5 compares the quality of land across Vietnamese provinces using the FAO's Global Agro-Ecological Zones data analyzed in Adamopoulos and Restuccia (2022). We follow Adamopoulos and Restuccia (2022) by measuring land quality as the average potential

Table B.4: Investment and Farm Characteristics

	South			North		
	(1) $\text{Inv}_t^{Cash} > 0$	(2) $\text{Inv}_t^{Lab} > 0$	(3) Ext_t	(4) $\text{Inv}_t^{Cash} > 0$	(5) $\text{Inv}_t^{Lab} > 0$	(6) Ext_t
log TFP_t	0.122***	0.221***	0.119***	0.0819***	0.255***	0.212***
	(0.0133)	(0.0163)	(0.0180)	(0.0164)	(0.0234)	(0.0206)
log TFPR_t	−0.113***	−0.224***	−0.0910***	−0.0707***	−0.228***	−0.196***
	(0.0142)	(0.0179)	(0.0196)	(0.0166)	(0.0235)	(0.0202)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.0475	0.108	0.211	0.0610	0.197	0.234
Observations	4387	4387	4387	6139	6139	6139

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the household level are included in parentheses. $\text{Inv}_t^{Cash} > 0$ and $\text{Inv}_t^{Lab} > 0$ take value one if the household reports any cash or labor investment in irrigation or soil and water conservation.

yield of land (across cells) within the province. We focus on two measures: an average of 27 crops and wet rice, the most prevalent crop in Vietnam. We use the rainfed, low input potential yield which most closely reflects the land quality without human intervention, see Adamopoulos and Restuccia (2022) for details and discussion.

Panel A of Table B.5 describes land quality differences between the North and South for all provinces in the country. Panel B focuses on only the twelve provinces that are included in the VARHS dataset. Panel C adjusts the mean values of land quality for the relative frequency of observations in our final dataset.

Comparing Panel A and Panel C shows that after adjusting the means for the relative frequency of observations there is little difference between our final dataset and the average province in the North and South. The observed differences in land quality are not large enough to explain the productivity gap that we observe between farms in the North and South. Taking the production function in Section 3 implies that the impact of land quality on TFP requires differences to be scaled by a factor $\alpha\gamma = 0.35$. This would further reduce the potential impact of any differences between the North and the South.

Table B.5: Comparison of Land Quality

A. All Provinces

	Mean Avg.	Sd Avg.	R9010 Avg.	Mean Rice	Sd Rice	R9010 Rice
North	88.8	0.4	3.2	1.6	0.8	6.6
South	87.3	0.3	2.0	2.1	0.5	3.2
Total	88.0	0.4	2.2	1.9	0.7	5.3

B. In Final Dataset

	Mean Avg.	Sd Avg.	R9010 Avg.	Mean Rice	Sd Rice	R9010 Rice
North	67.4	0.6	3.7	1.1	1.0	12.4
South	94.9	0.3	2.1	1.7	0.5	4.1
Total	81.2	0.5	3.5	1.4	0.9	11.3

C. In Final Dataset (observation-weighted means)

	Mean Avg.	Sd Avg.	R9010 Avg.	Mean Rice	Sd Rice	R9010 Rice
North	87.2	0.6	3.7	1.7	1.0	12.4
South	86.3	0.3	2.1	1.8	0.5	4.1
Total	86.8	0.5	3.5	1.8	0.9	11.3

Notes: Values calculated using provinces as unit of observation. “Avg.” refers to statistics calculated on the average potential yield of 27 common crops. “Rice” refers to statistics calculated on the average potential yield of wet rice. “Sd” is the standard deviation of the log variable. “R9010” is the ratio between the 90th and 10th percentile observations. Panel C constructs the mean values using the relative frequency of farm-year observations in our data as weights.

C Other Quantitative Results

We perform sensitivity analysis of calibration moments to parameter values. We also conduct a full calibration analysis to data for the North and provide results when removing distortions in the calibrated benchmark economy.

C.1 Sensitivity of Calibration Moments to Parameters

Table C.6 summarizes the changes in moments to a 10% change in the model parameters, highlighting that the moments are highly interconnected with the set of parameters implying that no individual moment identifies an individual parameter. Nevertheless, the table shows that the chosen moments for calibration are informative about the values of parameters in the calibration. The relationship between the moments and parameters is discussed in detail in Section 5.

Table C.6: Sensitivity of Moments to Calibrated Parameters (%)

	ψ	ζ	λ	κ^P	κ^O	σ_z	σ_v
Top 10% Land Share	−0.1	0.3	0.2	0.2	−0.1	0.9	0.8
Reg Coeff: growth on log TFP	0.7	−3.0	−1.3	−0.3	0.4	−7.4	8.8
Avg Growth	−2.2	4.8	6.0	0.0	0.0	−0.8	−1.6
Rel TFP perenials	−0.1	−0.3	0.4	3.1	−0.1	0.0	0.0
Rel TFP other	0.1	−0.1	0.1	−0.1	3.8	0.4	0.4
Std log TFP	−0.4	1.7	1.0	0.2	−0.1	3.8	3.9
Std Growth	0.0	0.0	0.0	0.0	0.0	−0.1	7.5

Notes: Percent change in the moments from a 10% change in each parameter relative to the benchmark calibration value. For λ the change is calculated only on the value above one.

C.2 Calibration to the North

The baseline experiment applies distortions set to match north Vietnam to the benchmark economy, calibrated to match south Vietnam. We show that the counterfactual economy moves towards the north Vietnam data moments, relative to the benchmark economy. An alternative approach is to re-calibrate the model to match the full set of moments from north Vietnam and then use this to compare with south Vietnam. We explore this approach in this section.

Calibration moments and parameters. The calibration follows the same procedure as in the baseline calibration. We adjust the total stock of land to be $L = 1.10$ to reflect the smaller average farm size in the North. The jointly calibrated parameters are selected to target the same moments as in the baseline calibration, where the values for the North are reported in Table C.7. The one difference is that we set the preference curvature θ to the value calibrated in South Vietnam and use the crop-specific preferences η^i to match the farm crop share.

Table C.7: Moments Calibrated to North Vietnam

	Model	Data
Avg TFP Growth (%)	2.62	2.62
Std TFP Growth	82.2	89.2
Std log TFP	0.89	0.84
Reg coefficient: growth on log TFP	−49.8	−48.2
Top 10% Land Share (%)	27.6	38.3
Rel Measured TFP	(1.00 , 0.69 , 0.80)	(1.00 , 0.69 , 0.80)
Farm Share by Crop (%)	(75.1 , 5.0 , 19.9)	(75.1 , 5.0 , 19.9)

The parameters in the re-calibrated model are summarized in Table C.8. There are several differences between the North and South parameters that provide insight into the productivity differences between the regions.

Overall, the parameter values in the re-calibrated model are relatively similar to those in the baseline calibration. This reflects the overall ability of the benchmark economy to match the North data moments when the North distortions were imposed. The main difference between the North and South parameters is in the ability investment function, (λ, ψ, ζ). Relative to the South, investment in the North is substantially cheaper but also has a lower payoff. The lower return to investment through λ explains the lower farm dynamism in the North compared with the South.

Aggregate productivity. The re-calibrated model replicates the productivity gap between north and south Vietnam. Following (3), aggregate total factor productivity in the

Table C.8: Parameters Calibrated to North Vietnam

Parameter		North	South
Discount Factor	β	0.96	0.96
Survival Rate	ξ	0.955	0.955
Land	L	1.1	2.77
Span-of-Control	γ	0.7	0.7
Land Share	α	0.5	0.5
Investment Level	ψ	1.07	3.79
Investment Curvature	ζ	2.75	2.81
Ability Step Size	λ	1.29	1.51
Crop Preference Elasticity	θ	1.87	1.85
Preference Shifter	η^i	(1.00 , 0.57 , 0.87)	(1.00 , 0.66 , 0.83)
Crop-Specific Productivity	κ^i	(1.00 , 0.31 , 0.49)	(1.00 , 1.28 , 0.38)
Permanent Productivity	σ_z	1.3	1.47
Random Productivity	σ_v	2.57	2.32
Elasticity	ρ	0.96	0.86
Crop-Specific Distortion	φ^i	(1.00 , 0.68 , 0.92)	(1.00 , 1.61 , 1.12)
Random Distortion	σ_ε	0.81	0.92
Crop Restriction	ω	0.43	0.23

calibrated economy is calculated as:

$$\frac{A^{North}}{A^{South}} = \frac{Y^{North}/(L^{North})^{\alpha\gamma}}{Y^{South}/(L^{South})^{\alpha\gamma}} = 37.8\%.$$

We correct the above TFP comparison for differences in the average permanent productivity z and idiosyncratic productivity v that arise from differences in the calibrated dispersion σ_z and σ_v (due to the log normal distribution).

C.3 Undistorted Economy

The undistorted economy represents a hypothetical first-best economy that could be achieved if all institutional distortions were removed. In practice, it is unclear whether this economy is achievable since some baseline distortions may be unavoidable or reflect other factors (e.g., measurement error). Additionally, the comparison with the undistorted economy takes the entirety of the measured distortions τ in the economy as a wedge on production, which may overstate the gains from reallocation if these wedges capture measurement error. With those caveats in mind, we find the undistorted economy useful as a benchmark to understand the full potential gains in productivity.

We calculate the undistorted economy by setting the parameters as in the baseline calibration and setting the government-imposed crop restrictions to $\omega = 0$, the elasticity of distortions to $\rho = 0$, the crop-specific distortions to $\varphi^i = 1$ for all crops i, and the random component of distortions $\sigma_\varepsilon = 0$.[5]

Table C.9: Comparison with Undistorted Economy

	Benchmark	Undistorted
Productivity	1.00	6.2
Avg TFP Growth (%)	6.25	0.82
Std TFP Growth (%)	76.52	76.08
Std log TFP	1.01	0.9
Reg coefficient: growth on log TFP	−34.16	−40.54
Top 10% Land Share (%)	35.49	86.51
Rel Measured TFP	(1.00 , 1.21 , 0.74)	(1.00 , 1.15 , 0.66)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(53.3 , 40.8 , 5.9)

The undistorted economy is around 6.2 times as productive as the benchmark economy. Table 5 shows that the gains from removing static misallocation in the benchmark economy is around 69% implying that the remaining gains are coming from improving the productivity distribution through higher investment in ability and selecting into more productive crops. That said, improving the productivity distribution alone would not yield the full 3.7x ($\approx$ 6.2/1.69) gain since the cost of resource misallocation would increase as the productivity distribution became more dispersed.

One other noticeable difference between the benchmark and undistorted economy is in the average growth rate. This can be understood through two channels. First, removing cor-

[5] Setting σ_ε as in the benchmark economy does not substantially alter the results relative to Table C.9. This implies a minimal role of measurement error since we interpret measurement error as being captured in the idiosyncratic component of distortions σ_ε.

related distortions causes investment in ability to become flat with respect to the farmer's ability because farmers are not disincentivized by larger distortions at higher abilities. All else equal, this causes higher ability farmers to invest more than in the benchmark economy. Second, removing distortions improves productivity and, consequently, the wage rate w and cost of land q, which results in lower profits for a given ability level. Lower profits disincentivize investment in ability for all farmers. The net impact is that lower ability farmers invest less in the undistorted economy while higher ability farmers invest more. This results in both more low ability farmers and more very high ability farmers in the undistorted economy. The productivity gains are then driven by these increases in the top end of the productivity distribution, which is consistent with the concentration of agricultural production in large, highly productive farms in advanced economies.