University of Toronto
Department of Economics

Working Paper 754

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

By Stephen Ayerst, Loren Brandt and Diego Restuccia

August 18, 2023

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

Stephen Ayerst† Loren Brandt‡ Diego Restuccia§

August 2023

Abstract

The expansion in farm size is an important contributor to agricultural productivity in developed countries where more productive farms are larger, but in less developed economies the allocation of factor inputs to more productive farms is hindered. How do distortions to factor-input allocation 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 47%, accounting for 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 static 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, Marco Bassetto, Anmol Bhandari, Dan Cao, Julieta Caunedo, V. V. Chari, Klaus Desmet, Mark Huggett, David Lagakos, Toshi Mukoyama, Richard Rogerson, Kjetil Storesletten, Mike Waugh, Kirk White, Zoltan Wolf, and seminar participants at Georgetown, UBC, ASU Herrendorf memorial conference, the US Census, Minnesota, and Melbourne. 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), contributing to a more efficient allocation of resources across producers. Business dynamism is also observed in the agricultural sector in developed countries where land consolidation and farm exit are important drivers of productivity (Key, 2019). In sharp contrast, the allocation of resources across producers 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 allocation 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, most 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 farm production to the household and liberalizing output and input markets—that led to substantial improvements in productivity and growth. Reforms have been highly uneven across regions (north and south Vietnam) 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. We start by showing differences in factors related to productivity in north and south Vietnam, where measured productivity is over twice as high in the South. First, we show that farm inputs tend to be more correlated with farm productivity in the South than in the North. The elasticity of land and labor use with respect to productivity is three to four times higher in the South than in the North. Second, we show that crop productivity and selection into crops differs across the regions. While farms in both regions primarily grow rice, farmers in the South are much more likely to grow perennial cash crops, such as coffee. Third, the productivity of young farmers in the South grows faster than young farmers in the North. Additionally, the typical farmer in the South grows over twice as fast as in the North.

We develop a dynamic model of farms to understand the differences in productivity between south and north Vietnam. The model's structure 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 labor use depend on the productivity distribution.

Our main departure from the existing literature is to endogenize the productivity process of farms. Farm productivity depends on four components that reflect empirical differences observed in the data. The first is a permanent farmer-specific productivity component. The second is a random productivity component that varies between periods. The third is an ability component that depends on endogenous investments of farmers to improve productivity, reflecting evidence that less distorted farmers invest more in farm improvements and experience faster productivity growth. Finally, there is a crop-specific productivity component that depends on the farmer's endogenous crop choice. 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 A.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, or insecure property rights) that affect farmers' input choices relative to their first best. Institutional distortions depend on farmer productivity, farmer crop choice, and a random stochastic component that varies over time. In addition, government land-use regulations, such as restrictions on crop choice, 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 government restrictions on crop choice captured by our data (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 reflects 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 rich 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, productivity growth, 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 an allocation of factors less sensitive to farm productivity; perennial crops are more distorted compared to rice and other crops; and a higher share of farms face government-imposed crop restrictions. Imposing the North distortions on the South leads to: 1) productivity falling by 47% relative to the benchmark economy, representing over 70% of the observed TFP gap between the North and the South; 2) a reduction in the average growth rate of farmers' output of 2.0 percentage points, or half of the observed gap between the North and the South; 3) a reduction in farm TFP dispersion of 9 percentage points, more than half the 16 percentage points lower dispersion in the North compared to the South; and 4) a drop in the share of perennial farmers from 33% to 8%, which is similar to the observed 5% in the North. In our quantitative analysis, the key institutional feature is the larger elasticity of distortions with respect to farm productivity in the North, reflecting the weaker relationship between factor inputs and productivity in the North.

To understand the channels through which larger measured distortions in the North account for the North-South productivity gap, we examine the separate contributions of factor misallocation, the endogenous farm productivity distribution, and crop choice. 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. The channels of dynamic misallocation (crop choice and farm productivity) are the source for two-thirds of the productivity loss in the model, more than doubling the impact of static factor misallocation.

Our work connects with several strands in the literature. We relate to the broad literature on resource misallocation across production units for understanding aggregate productivity (Restuccia and Rogerson, 2008; Guner et al., 2008; Hsieh and Klenow, 2009), particularly in agriculture (Adamopoulos and Restuccia, 2014) where important aspects of land institutions are associated with the misallocation of land and other productive inputs, hampering agricultural productivity (Chen et al., 2023, 2022; Adamopoulos et al., 2022; Bolhuis et al., 2021). We differ from this literature in two important dimensions. First, we take advantage of Vietnam's unique historical context to quantify differences in misallocation between north and south Vietnam, rather than quantifying the level of misallocation relative to a hypothetical undistorted economy. Second, we assess the broader effects of misallocation by quantifying the impact of differences in measured distortions on crop choice and farm investment. 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., 2023; 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. We also contribute to a growing literature investigating economic growth and regional convergence in Vietnam (Benjamin and Brandt, 2004; Le, 2020; Ayerst et al., 2020).

The paper is organized as follows. Section 2 summarizes the institutional context. Section

3 summarizes the data, construction of key moments, and differences between the north and south Vietnam. Section 4 describes the model. 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 out productivity-augmenting investments in land and water and acquire information on new technologies and markets through their involvement in extension services. We examine the effect of these investments on farm productivity growth, and their correlation with measures of market distortions discussed in the next section. In both the North and the South, more productive 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 Productivity in North and South Vietnam

We provide an overview of our dataset and construction of the main variables used in our analysis. We use the constructed dataset to examine productivity differences between north and south Vietnam in (*i*) the misallocation of factors of production, (*ii*) crop productivity and selection into crops, and (*iii*) farm dynamism. These empirical differences act as the foundation for the model that we develop in the next section.

3.1 Data and Variable Construction

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 a brief overview of the data construction, a more detailed documentation is provided in Ayerst et al. (2020). Our variables of interest are output (value added) $y_{f,t}$, land $l_{f,t}$, and labor $n_{f,t}$ at the farm-year (f,t) level.

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 and reported values and quantities when sales are unavailable.[1] Finally, we set observations with negative value added to zero, which account for around 2.5% of our sample and are primarily in the North.

[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 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).

Production inputs are land and labor. Land is constructed as the cultivated area of plots owned and rented by households excluding land used for activities unrelated to crop production (e.g., forestry, animal husbandry), left fallow for more than 48 months, or rented out. Labor is constructed as the sum of hired labor and the labor supplied by household members. Using information on family members that hire out in agriculture, we construct wages controlling for individual characteristics (e.g., age, education, sex), the year of work, and regional differences that we use to convert household expenditure on hired labor into a quantity of hired labor.

Following Hsieh and Klenow (2009), we construct the farm-level TFP and the farm-level wedge for farm f as:

$$\text{TFP}_{f,t} = \frac{y_{f,t}}{(\ell_{f,t}^{\alpha} n_{f,t}^{1-\alpha})^{\gamma}}, \qquad \text{and} \qquad \text{Wedge}_{f,t} = \frac{y_{f,t}}{\ell_{f,t}^{\alpha} n_{f,t}^{1-\alpha}}. \tag{1}$$

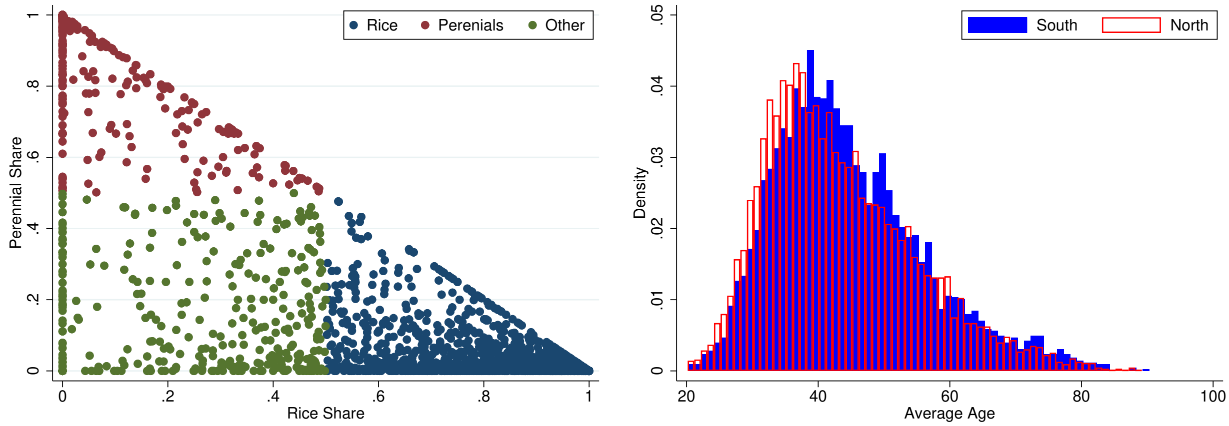
We use the term wedge, instead of TFPR (total factor revenue productivity), to highlight that the wedge is a model-based measure of misallocation of land and labor across producers. In an undistorted economy, marginal products and hence wedges are equalized across producers. The variables in (1) are consistent with the model that we develop in the next section.

We categorize households as either rice farms, perennial farms, or other (annual) crop farms based on their most valuable crop grown over the survey.[2] We categories farmers as a rice or perennial farmer if more than 50% of their output value, across all years, is in rice or perennial crops. We do not impose strict annual cutoffs because of inter-cropping, crop-rotation, and the fact that farms may devote some of their land to other crops. However, cropping tends to be concentrated in these categories.[3]

[2]In the survey, perennials include: fruits, coffee, tea, cocoa, cashew nuts, sugarcane, pepper, 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.

[3]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. In addition, more than 90% of rice and perennial farm-year observations would have the same classification if classified year by year. Differences are more common in the case of other crop farmers.

Figure 1: Farm Crop Type and Age Distribution



(a) Crop Composition

(b) Age Distribution

Figure 1a summarizes the empirical distribution of households across farm types. Unsurprisingly, most households are classified as rice farms as this is the most widely produced crop in Vietnam. The remaining households are split between perennials and other crop farms. Other crop farms grow, on average, around 30% of value in rice, but have a higher production value in other annual crops.

We construct a measure of farm age for each household. Farm age is taken as an average of individual household member age weighted by the number of days working in crop production. Figure 1b reports the histogram of constructed farm ages in the North and South. Most farms are between 20 and 60 years old and the South is slightly older than the North.

Finally, to minimize the influence of outliers, we winsorize the final set of variables at the top and bottom 2% for both the North and South in each year.[4]

Land quality differences. The productivity gap between the North and South is unlikely to be explained by differences in land suitability. Appendix A.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 (GAEZ) analyzed in Adamopoulos and Restuccia (2022).

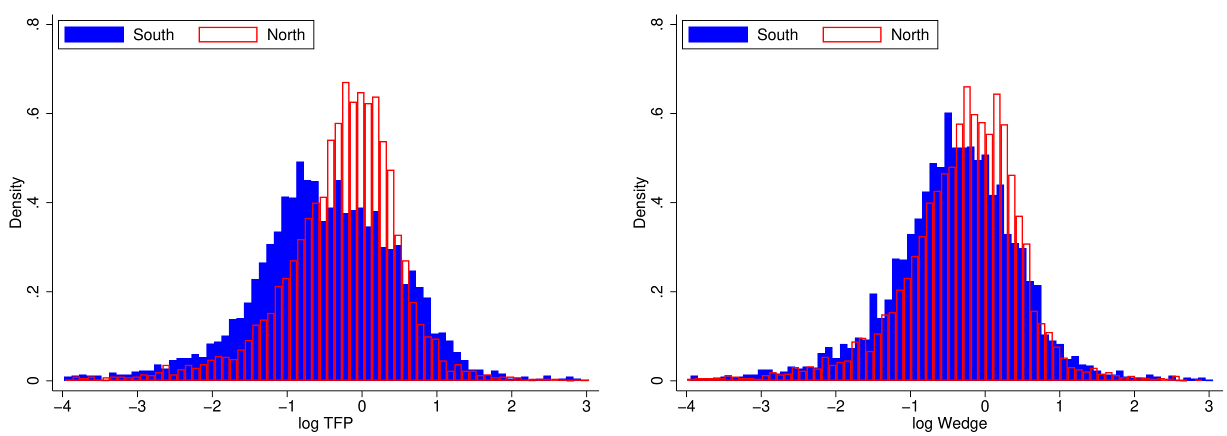
[4]We winsorize rather than trim the data since trimming disproportionately affects the share of perennial farmers in the final dataset. Other than for the crop share, trimming the data implies similar moments compared to winsorizing.

We find minimal differences in land quality between the two regions; if anything, land quality is slightly higher in the North than the South.

3.2 Misallocation

The first source of productivity differences between north and south Vietnam that we examine is the misallocation of factors of production. Figure 2 plots histograms of the TFP and the wedge distributions, where both variables are normalized by the mean in each region (North / South) year. Both TFP and wedges tend to be more dispersed in the South, as documented in Ayerst et al. (2020).

Figure 2: TFP and Wedge Distribution



(a) TFP Distribution (b) Wedge Distribution

Notes: Histogram of TFP and wedges for farm-year observations in north and south Vietnam. TFP and wedges are normalized by the mean in each region (North / South) year. We calculate TFP and wedges according to (1) with $\gamma = 0.70$ and $\alpha = 0.50$.

Wedges are important because of their impact on the allocation of resources in the economy. In an undistorted economy, resource allocation is proportional to farm-level TFP. As a simple measure of misallocation, we regress farm-level land and labor inputs on measured TFP to examine the allocative efficiency in each region. Table 1 reports the results.

The results show that a one log point higher TFP results in three-to-four times higher factors allocated to the farm in the South compared with the North, pointing to substantially

Table 1: Farm Allocations

	(1) log Land	(2) log Labor
log TFP (South)	0.554***	0.382***
	(0.0312)	(0.0228)
log TFP (North)	0.152***	0.122***
	(0.0200)	(0.0169)
North FE	Yes	Yes
R^2	0.208	0.132
Observations	10526	10526

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 fixed effects for the region (North/South) of the household.

larger misallocation in the North. It is also important to note that while the allocative efficiency of the South is much higher than that in the North, both economies face severe misallocation. In the hypothetical undistorted economy, the elasticity between land or labor and TFP is $1/(1-\gamma) = 3.33$ based on the model and calibrated parameters.

3.3 Crop Productivity Differences

The second source of productivity differences between north and south Vietnam that we examine is crop selection. Table 2 provides summary statistics by farm type for north and south Vietnam. An observation is a farm-year and only includes farm-years where TFP can be calculated. Farm-level TFP and wedges are normalized to one in each region-year such that reported values capture relative productivity and distortions within regions.

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. Among perennials, coffee is the most important. 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

Table 2: Crop Differences in North and South Vietnam

A. Vietnam, South

Crop Type	Mean						
	Output	Land	Labor	TFP	Wedge	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	Wedge	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

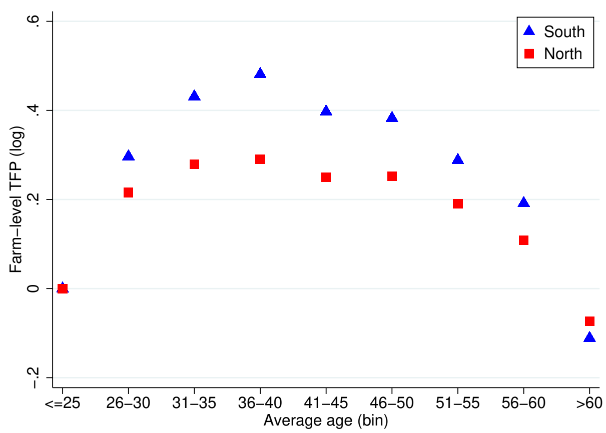
Notes: Observations are at the farm-year level. Output is reported as the log of total agricultural output using real prices common across farms to sum different crops. Land is reported in acres. Labor is reported in number of effective worker days. TFP and wedges are reported in logs and constructed following the equation (1). TFP and wedges 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 $\text{TFP Growth}_t = 100 \times (\text{TFP}_{f,t} - \text{TFP}_{f,t-2})/[0.5(\text{TFP}_{f,t} + \text{TFP}_{f,t-2})]$ before TFP is normalized where t is the calendar year.

North also tend to be smaller in terms of land and labor inputs and output, and experience lower growth. These differences motivate our main quantitative experiment.

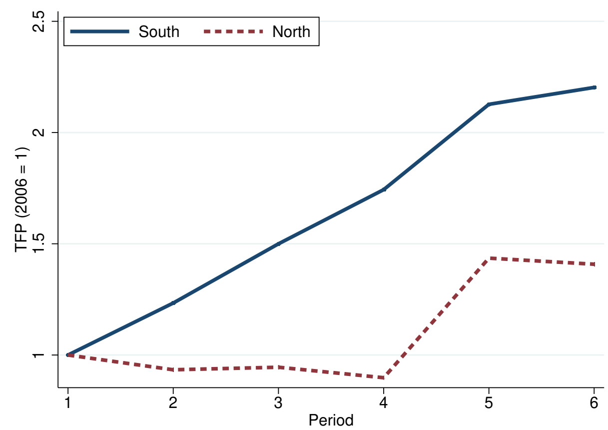
3.4 Farm Dynamism

The third source of productivity differences between north and south Vietnam that we examine is differences in farm dynamism. Following Hsieh and Klenow (2014), we compare life cycles of farms in north and south Vietnam by constructing synthetic life cycles using a measure of farm age. We use the average age of household members weighted by their time spent working in crops since this measure most closely aligns with our model. We find similar life cycle profiles using the age of the head of households and the simple average age of household members (Appendix A.2). Figure 3a reports the TFP of farms in different age bins obtained from regressions of farm-level TFP on fixed effects for the age bin. We remove region-by-year variation from farm-level TFP so that the life cycle is not contaminated by time trends.

Figure 3: Farm Dynamism in North and South Vietnam



(a) Farm-level productivity



(b) Aggregate productivity

Notes: Farm-level TFP and age are constructed as described in the text. Panel (a) reports the estimated of age bin fixed effects c_j^R (for $R \in \{South, North\}$) from the regression $\log \mathrm{TFP}_{f,t} = \sum_{j \in \mathcal{A}} c_j^R 1_{age_{f,t} \in j} + \Gamma^R + \Gamma_t + \varepsilon_{f,t}$ where Γ^R and Γ_t are region and year fixed effects. The coefficient estimates are normalized such that the youngest bin has value zero. Panel (b) reports aggregate TFP calculated as $Y_t/(L_t^{\alpha} N_t^{1-\alpha})^{\gamma}$ for a balanced panel of households active in cropping activities. Aggregate TFP is normalized to one in the first period (2006). Parameter values are $\gamma = 0.7$ and $\alpha = 0.5$.

In both the North and the South, farm productivity increases rapidly until age 40 and then begins to level off before declining at much older ages. These trends are much more pronounced in the South where the initial increase in productivity is steeper. Hsieh and Klenow (2014) similarly find that firms in less distorted economies experience more productivity growth over their life cycle, but do not find the decline at old ages. Prior to extensive mechanization, we expect that the decline is likely driven by declining physical abilities of older household members working in agriculture and by selection of households with older members active in agriculture.

On its own, differences in the life cycles of farmers do not imply differences in regional productivity. Differences in average growth, unequal growth across age groups, or selection can complicate the relationship between the productivity life cycle and aggregate productivity. We use the panel dimension of the data to address this issue. Figure 3b reports the evolution of aggregate TFP in the North and the South for a fixed group of households. We further restrict the sample to only include households for which we can construct TFP in all periods in order to limit the influence of unrelated factors, such as selection out of cropping. Figure 3b shows that the productivity of a typical farm in the South increases substantially faster than in the North over the same time horizon. The differences translate into annualized growth rates of around 8.2% in the South and 3.4% in the North.

Investment and productivity. We develop a model in the next section in which farmers invest in improving productivity to capture the productivity increase of young farmers. While we do not observe a single comprehensive measure of investment, the survey asks households about key farm investments and their participation in extension services on new technology and farming methods. First, we construct a variable $\text{Inv}_{f,t}$ that takes value one if farm f has made any cash or labor investments in irrigation or soil and water conservation. Second, we construct a variable $\text{Ext}_{f,t}$ that takes a value of one if farm f participates in extension services providing information on (a) new seeds, varieties, or breeds; (b) fertilizer use; (c) irrigation; (d) pest infestation and blight; or (e) market conditions.

Table 3 reports the relationship between investment and extension services and farm characteristics, following closely König et al. (2022). Panel A reports the relationship between investment and extensions services and farm-level productivity and wedges. Panel B reports the relationship between farm-level TFP growth and previous investment and extension services. We include the farm's previous productivity to control for more productive farms growing slower in the data, which our model also replicates.

Investment and extension services are positively related to farm-level TFP and negatively related to farm-level wedges, indicating that more distorted farms are less likely to take steps

Table 3: Investment and Farm Characteristics

A: Investment and Extension Services

	South		North	
	(1)	(2)	(3)	(4)
	$\text{Inv}_{f,t}$	$\text{Ext}_{f,t}$	$\text{Inv}_{f,t}$	$\text{Ext}_{f,t}$
$\log\ \text{TFP}_{f,t}$	0.240***	0.119***	0.253***	0.212***
	(0.0173)	(0.0180)	(0.0235)	(0.0206)
$\log \text{Wedge}_{f,t}$	-0.242***	-0.0910***	-0.225***	-0.196***
	(0.0187)	(0.0196)	(0.0236)	(0.0202)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
R^2	0.116	0.211	0.197	0.234
Observations	4387	4387	6139	6139

B: Farm TFP Growth

	South		North	
	(1)	(2)	(3)	(4)
	$g_{f,t}$	$g_{f,t}$	$g_{f,t}$	$g_{f,t}$
$\text{Inv}_{f,t-2}$	14.94***		-3.961*	
	(3.047)		(2.195)	
$\text{Ext}_{f,t-2}$		15.28***		8.567**
		(4.079)		(3.709)
$\log\ \text{TFP}_{f,t-2}$	-40.01***	-40.18***	-55.80***	-56.02***
	(1.429)	(1.465)	(1.745)	(1.747)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
R^2	0.263	0.262	0.320	0.320
Observations	3485	3485	4883	4883

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the household level are included in parentheses. $\text{Inv}_{f,t}$ takes value one if the household reports any cash or labor investment in irrigation or soil and water conservation. Farm TFP Growth is calculated over a two-year period as $g_{f,t} = 100 \times (\text{TFP}_{f,t} - \text{TFP}_{f,t-2})/[0.5(\text{TFP}_{f,t} + \text{TFP}_{f,t-2})]$ (before TFP normalization), where t is the calendar year.

to improve productivity. In the South, both investment and extension services are associated with faster growth while the relationship is weaker in the North where only extension services are associated with faster growth. These results continue to hold if we separate investment in cash and in labor and if we include the intensive margin of investment.

4 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.

4.1 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 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.

The investment function reflects the empirical evidence showing that farmers invest to improve productivity and that younger farms experience rapid productivity improvements while the productivity of middle-aged farms experience relatively slow or flat productivity growth (Section 3). In Appendix C.5, we show a model extension that incorporates the productivity decline of older farms. This extension does not significantly alter our results.

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 perennials involves extensive investment and seasons in which the plot does 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 A.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 ability $a_{f,t} = \lambda^0$. In Appendix C.5, we consider a model extension where entrant ability depends on predecessor ability, capturing, for example, transfers of skills from the old to the young. The extension has a negligible effect on the final results.

4.2 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, such that farm revenues net of the tax are $(1-\tau^i_{f,t})y^i_{f,t}$. While we model the wedge as a tax on revenues, this is isomorphic to modeling wedges on factors inputs, which could capture, for example, land transaction restrictions or lack of market access to intermediate inputs. Since the relationship between land and labor and productivity is relatively similar (Table 1), we choose to model wedges parsimoniously on output. The model distortions we define directly corresponds to the wedge defined in Section 3, Equation (1), as $\text{Wedge}_{f,t} = 1/(1-\tau^i_{f,t})$.

Higher values of $\tau^i_{f,t}$ imply that farmers operate smaller farms than they would in the absence of the distortion. Distortions are distributed according to

$$\ln(1-\tau^i_{f,t}) = (1-\gamma)\left[\ln\bar{\tau} + \ln\varphi^i - \rho\ln\left(s^i_{f,t}\right) + \varepsilon_{f,t}\right], \tag{2}$$

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 $\varepsilon_{f,t}$ 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_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 $\tau^i_{f,t}$ 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 $\tau^i_{f,t}$ that farmers face implying that restricted rice farmers are otherwise identical to unrestricted rice farmers.

Timing. The timing of each period is: (i) new farmers enter; (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; and (vi) farmers exit.

4.3 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

$$\Omega_z^i = \begin{cases} \omega + (1-\omega)\frac{(\eta^i \bar{V}_z^i)^\theta}{\sum_{i' \in \mathcal{I}} (\eta^{i'} \bar{V}_z^{i'})^\theta} & \text{for } i = R \\ (1-\omega)\frac{(\eta^i \bar{V}_z^i)^\theta}{\sum_{i' \in \mathcal{I}} (\eta^{i'} \bar{V}_z^{i'})^\theta} & \text{for } i \neq R \end{cases}. \tag{3}$$

See Appendix B 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

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

where $\mu_{E,z}^i$ is the entry rate of farmers. 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 $\mu_{E,z}^i = 1 - \xi$.

Aggregate output. Production of the agricultural good is given by

$$Y = \left[\frac{\left(\int_{v,\varepsilon} e^{\gamma\varepsilon+(1-\rho\gamma)v} d\Phi_v(v) d\Phi_\varepsilon(\varepsilon)\right) \int_z \sum_i \sum_h (\varphi^i)^\gamma (z\kappa^i\lambda^h)^{1-\rho\gamma} \mu^i_{z,h} \Omega^i_z d\Phi_z(z)}{\left(\int_{v,\varepsilon} e^{\varepsilon+v(1-\rho)} d\Phi_v(v) d\Phi_\varepsilon(\varepsilon)\right)^\gamma \left(\int_z \sum_i \sum_h \varphi^i (z\kappa^i\lambda^h)^{1-\rho} \mu^i_{z,h} \Omega^i_z d\Phi_z(z)\right)^\gamma} \right] \times$$
$$N_F^{1-\gamma} (L^\alpha N_W^{1-\alpha})^\gamma, \qquad (4)$$

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 raised 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 directly depend on the average level of distortions $\bar{\tau}$, which is canceled out by general equilibrium effects. The rest of the expression describes inputs of farm managers N_F, land L, and farm workers N_W to aggregate output, where aggregate output has constant returns in all three factors.

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.

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 described in Section 3.

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 (2) in the model, the parameters related to distortions can be estimated by regressing the measured farm-level wedges on TFP and fixed effects for the farm's crop type. Table 4 summarizes the estimated distortions in south and north Vietnam.

Table 4: Estimating Distortions in the South and the North

	National	South	North
	(1) log Wedge	(2) log Wedge	(3) log Wedge
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.

We use the estimated coefficients to parameterize the institutional distortions in the South. We set the elasticity $\rho = 0.856$ of wedges to TFP to the value of the coefficient on TFP from the regression, which is consistent with the model-implied elasticity. Since output wedges are a weighted average of wedges in inputs, the value of ρ falls between the elasticity implied by separate regressions of land and labor on farm TFP in Section 3 (Table 1), which are 0.83 and 0.89 respectively, suggesting distortions affect farm inputs proportionally. 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 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 the wedge is equal to one and hold $\bar{\tau}$ constant in the counterfactual economies. This is a conservative assumption since allowing the value of $\bar{\tau}$ to adjust so that the wedge 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 5. 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.

Table 5: Calibration Moments

	Model	Data
Avg Growth (%)	6.25	6.23
Std Growth	77.1	75.2
Std log TFP	1.00	1.00
Reg coefficient: growth on log TFP	-34.1	-34.4
Top 10% Land Share (%)	35.3	41.2
Relative Measured TFP	(1.00 , 1.21 , 0.75)	(1.00 , 1.20 , 0.75)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(49.1 , 33.1 , 17.8)

Notes: For Relative Measured TFP and Farm Share by Crop, we report moments first for rice farms, followed by those for perennials and then other crop farms. Farm share by crop is calculated based on (3). 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 $t+2$).

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 (υ, 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 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_υ.

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, $\ln \text{TFP}_{f,t} = k^P \times \text{Peren.}_{f,t} + k^O \times \text{Other}_{f,t} + \Gamma_t + \epsilon_{f,t}$ 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 3. 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 (3).

5.3 Parameters

Table 6 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 crop) farmers are 11% more (27% 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.

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

Table 6: Model Parameters

Parameter		Value
Discount Rate	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	1.69
Preference Shifter	η^i	(1.00 , 0.65 , 0.83)
Crop-Specific Productivity	κ^i	(1.00 , 1.43 , 0.34)
Permanent Productivity	σ_z	1.39
Random Productivity	σ_v	2.34
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

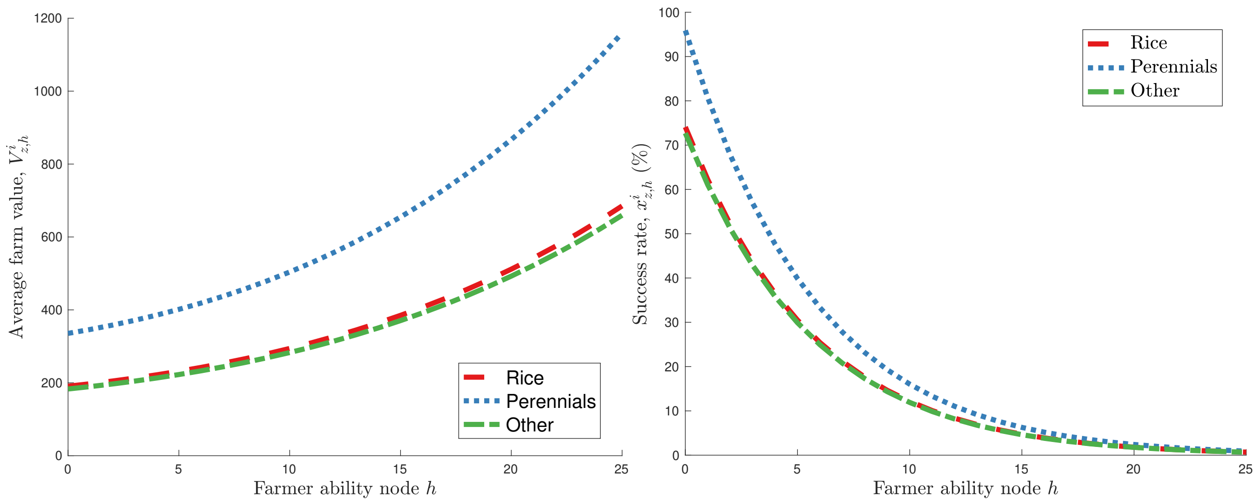
the manufacturing sector, which is closer to quadratic (e.g., Bento and Restuccia, 2017; Acemoglu et al., 2018; Ayerst, 2022). This implies that farm investment is comparatively less elastic to changes in profitability. The estimated curvature reflects, in part, that higher productivity farms tend to 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 the time that passes before they are mature and generate income.

5.4 Value and Policy Function

Figure 4 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 4: Value and Policy Function by Crop Specialization



(a) Value functions (b) Policy functions

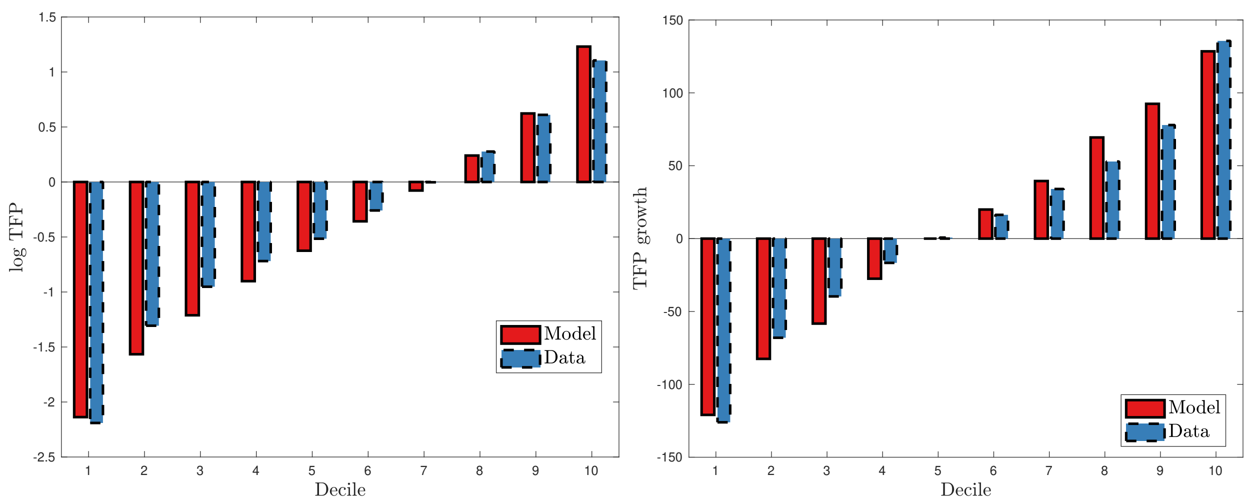
The value functions of the three crops reflect differences in relative profitability stemming from differences in productivity κ^i and distortions φ^i. Despite differences in κ^i and φ^i, the value of rice and other crop farmers are similar because the two parameters have offsetting effects on profitability. 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 5.

5.5 Other Moments and Goodness-of-Fit

Figure 5 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 5: Farm Productivity Distribution

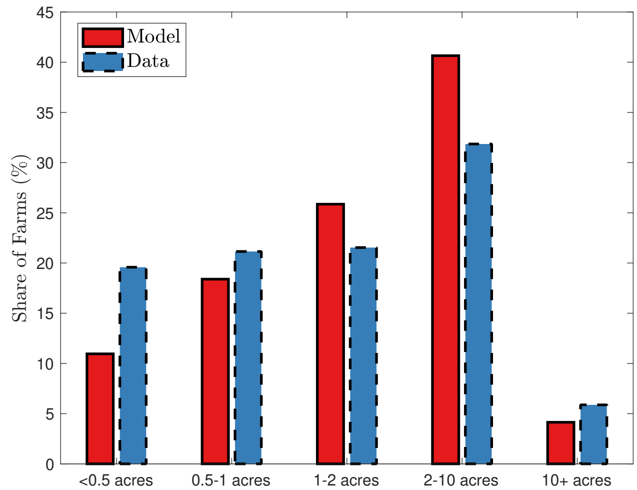


(a) TFP by decile (b) Growth by decile

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 6 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 6: Farm Land-Size Distribution



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

Table 7 compares other data moments with their 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 4 that are directly targeted. The third moment shows that the standard deviation of wedges 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 the farm-level wedge is smaller than that implied by the model, supporting our choice to model ε as transitory rather than permanent to 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 Hsieh and Klenow, 2009). This moment acts as a test of the goodness of fit of the joint TFP and wedge 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

Table 7: Other Model Moments

	Model	Data
Elasticity of Distortions*	0.86	0.86
Crop-Specific FE*	(0.000 , −0.142 , −0.035)	(0.000 , −0.142 , −0.035)
Std log Wedge	0.89	0.87
Autocorrelation Wedge	0.56	0.34
Gains from Reallocation (%)	66.2	62.0
Relative Output	(1.00 , 1.68 , 0.72)	(1.00 , 1.97 , 0.77)
Relative Land	(1.00 , 1.68 , 0.94)	(1.00 , 1.96 , 0.39)
Relative Labor	(1.00 , 1.68 , 0.94)	(1.00 , 1.65 , 1.07)
Std log Output	1.46	1.47
Std log Land	0.98	1.21
Std log Labor	0.98	1.07

Notes: Where applicable, we first report moments for rice farms, followed by those for perennials and then other crop farms. Moments with a * indicate moments that are directly targeted in the calibration.

data, the corresponding moments are based on the regressions reported in Appendix A.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 6, 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 explaining 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 government-imposed crop restriction ω; and (4) the random component of distortions σ_ε. Table 8 summarizes the values estimated for the first three of these for the counterfactual experiment. Other parameters, including the random component of distortions, are held fixed at the benchmark economy values.

Table 8: 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 4 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 4. 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 grow 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 productivity costs 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 those for north Vietnam compared to the benchmark economy. Table 9 compares the calibration moments and agricultural productivity in the benchmark and counterfactual economies as well as in 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) that impact the moments differ between regions. Nevertheless, Table 9 shows that the counterfactual economy is more similar to the data moments for north Vietnam than the benchmark economy.

Table 9: Comparing Counterfactual Moments with the Data for the North

	Benchmark	Counterfactual	Data
Productivity	1.00	0.53	0.42
Avg Growth (%)	6.25	4.29	2.62
Std Growth (%)	77.1	76.9	89.2
Std log TFP	1.00	0.91	0.84
Reg coefficient: growth on log TFP	−34.1	−40.0	−48.2
Top 10% Land Share (%)	35.3	30.5	38.3
Relative Measured TFP	(1.00 , 1.21 , 0.75)	(1.00 , 1.04 , 0.74)	(1.00 , 0.69 , 0.80)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(74.0 , 7.9 , 18.1)	(75.1 , 5.0 , 19.9)

Notes: Where applicable, moments are first reported for rice farms, followed by those for perennials and then other crop farms.

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 53% of the benchmark economy (South), implying that the model accounts for almost three quarters ($73\% \approx \log(0.53)/\log(0.42)$) of the productivity gap between the North and the South.

In addition, the model accounts for over half ($(6.25-4.29)/(6.25-2.62)$) of the gap in the average productivity growth rate of farmers between the South and the North. The model also accounts 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 accounts for around half of the gap in the standard deviation of log TFP and one-third of the regression coefficient of growth on log TFP. The similarity of the counterfactual economy and the north Vietnam data shows both that farm-level distortions correctly predict the directions of changes in moments in the data and that changes in farm-level distortions are important for explaining variation in these outcomes.

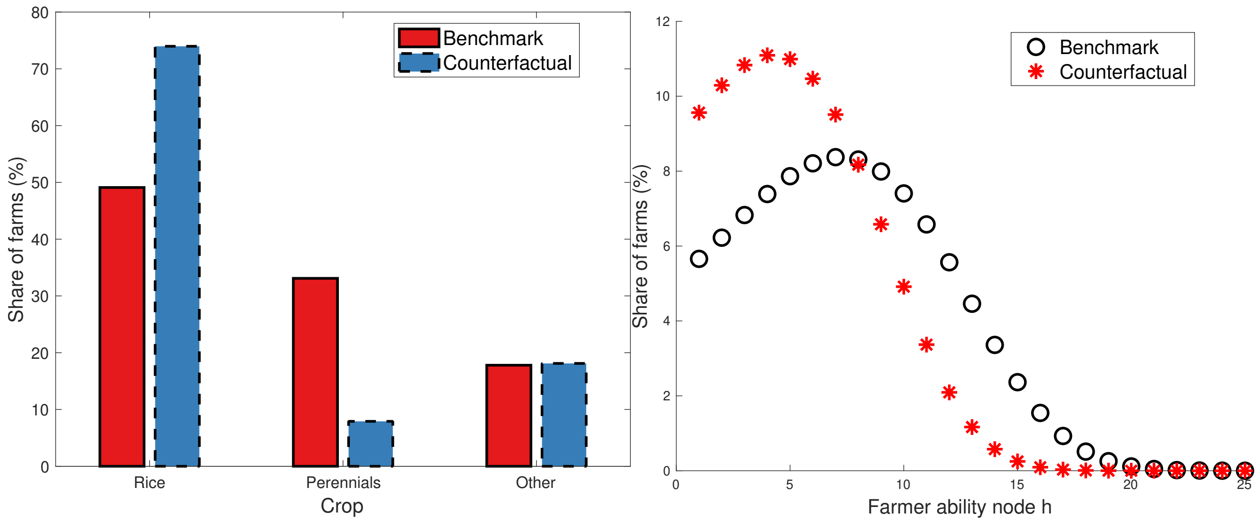
The model performs less well in two dimensions. First, it does not generate an increase in the standard deviation of growth, which is relatively unchanged in the counterfactual economy. This is because the standard deviation of growth is mostly driven by the idiosyncratic dispersion in productivity σ_v (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 4.8% ($= 35.5 - 30.5$) between the benchmark and counterfactual economies, which is slightly larger than the 2.9% ($= 41.2 - 38.3$ from Tables 5 and 9) 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 47%. What channels account for this productivity loss? Following equation (4), productivity in the model depends on factor misallocation, the crop distribution, and the ability distribution. Note that the change in output is equivalent to the change in productivity in our framework since aggregate inputs are held constant. Figure 7 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 productivity levels in the South.

To better understand the three components of productivity, we consider three experiments to decompose the relative contributions of factor misallocation, crop choice, and farm ability.

Figure 7: Farm Distributions in Benchmark and Counterfactual Economies



(a) Crop Distribution (b) Ability Distribution

Table 10 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 10: Output Loss by Channel

	Change in Output (%)
Factor Misallocation	−22.8
Crop Choice	−8.5
Farm Ability	−35.5
Sum of Channels	−66.8
Total	−46.8

Notes: The change in output is equivalent to the change in productivity since aggregate inputs are constant.

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 remain 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 22.8%, accounting for just under half of the productivity gap between the counterfactual and benchmark economies.

Partitioning the interaction effects proportionately to each channel, factor misallocation accounts for one-third ($\approx -22.8/-66.8$) 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 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 crop shares Ω^i_z are adjusted to match the counterfactual economy. We fix the within-crop ability distribution $\mu^i_{z,h}$ to that of the benchmark economy. However, the aggregate ability distribution, equal to $\mu^i_{z,h}\Omega^i_z\Phi_z(z)$, changes due to changes in the crop distribution. 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 -8.5%.

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. We adjust the farmer ability distribution $\mu^i_{z,h}$, conditional on crop i and permanent productivity z, to the counterfactual economy and hold the crop shares Ω^i_z fixed to that of the benchmark 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 35.5%, 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 over half of the productivity loss from the sum of channels, almost double the impact of factor misallocation.

The farmer ability channel also depends on the value of the elasticity of distortions ρ. At the extreme, farmer ability investment goes to zero as the elasticity of distortions ρ goes to one, because distortions at $\rho = 1$ eliminate any profit increase that farmers would receive from higher productivity. The impact of changes in ρ are asymmetric, with productivity changing more when ρ increases than when ρ decreases (see Appendix C.4). Consequently, the impact of increasing ρ from the South to the North value implies a large productivity loss despite the change in ρ being small when compared with the overall value of ρ.

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 11 summarizes the results.

Table 11: Output Loss from Individual Distortions

	ρ	φ	ω	(ρ, φ, ω)
Change in Output (%)	-44.9	-8.2	-1.7	-46.8

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 dampen the increase in profits associated with increasing farm productivity, which results in weaker incentives for farmers to invest in ability or select crops based on market factors as opposed to preferences. Our results point to a large productivity effect from seemingly small variation in correlated distortions ρ between the North and the South due to the asymmetric productivity effects from changes in ρ that are magnified as ρ approaches one as discussed earlier.

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 affect 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.

Government-imposed crop restrictions have 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 independent 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 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 prevent higher ability farmers from increasing production and disincentivize investment by farmers, magnifying the overall costs of misallocation. We evaluate the robustness of our results under alternative calibrations and model extensions.

Alternative calibrations. We consider two sets of exercises related to the calibrated ability distribution to examine the robustness of our results. Table 12 reports the productivity gap generated by the model using the alternative calibration parameters.

Table 12: Robustness of Main Results to Alternative Calibrations

	Relative Counterfactual Output (%)
Baseline	53.2
Alternative calibrations:	
Fix investment-cost curvature $\zeta = 2$	51.1
Avg growth target $6.23\% - 2\%$	57.1
Avg growth target $6.23\% - 4\%$	64.0

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 5. 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 12 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 3.9 percentage points compared with the baseline experiment. This implies that the model goes from accounting for 73% of the productivity gap to 65% of the productivity gap. Decreasing the targeted growth rate by four percentage points lowers the explanatory power of the model to 51% of the productivity gap between the North and the South.

Model extensions. We consider two model extensions and summarize the results; Appendix C.5 describes in more detail these extensions and results. First, we consider an extension of the model that replicates the hump shaped productivity life-cycle profile in Figure 3a. The extended model allows for farmer ability $a_{f,t}$ to also depend on a state variable that takes values *young* or *old*. Entrants start as *young* farmers and transition to *old* farmers over time, which is the absorbing state. We recalibrate the model to match the life-cycle profile found in the data. The relative counterfactual output is around 56%, similar to the value found in the baseline experiment.

Second, we consider an extension of the model that allows for entrants to draw ability from a distribution that depends on the ability of the predecessor (the farm that the entrant replaces). Intuitively, this could be thought of as capturing the passing of knowledge between generations. We recalibrate the model and repeat the same experiment as in the baseline model. We find slightly stronger productivity losses in this model extension due to the positive feedback of investment on the entrant productivity distribution.

7 Conclusion

We develop a 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 73% 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 captures the weaker relationship between factor inputs and productivity in the North relative to that 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.

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.

Bolhuis, M. A., Rachapalli, S. R., and Restuccia, D. (2021). Misallocation in indian agriculture. Technical report, National Bureau of Economic Research.

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.

Chen, C., Restuccia, D., and Santaeulàlia-Llopis, R. (2023). Land misallocation and productivity. *American Economic Journal: Macroeconomics*, 15(2):441–465.

Da-Rocha, J.-M., Restuccia, D., and Tavares, M. M. (2023). Policy distortions and aggregate productivity with endogenous establishment-level productivity. *European Economic Review*, 155:104444.

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.

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

On-line Appendix

A Data Details

A.1 Differences by Farm Type

Tables A.1 and A.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 A.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.

Figure A.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

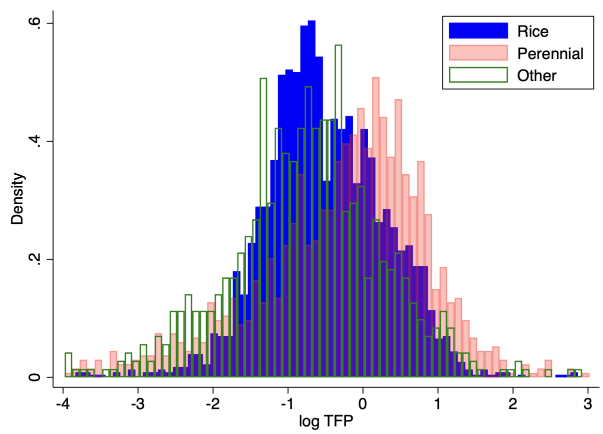
Table A.2: Farm Type Comparison in North Vietnam

	(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.

imply a discrete productivity cutoff in contrast with the data.

Figure A.1: Distributions of Farm TFP by Crop



A.2 Farm Life Cycle with Different Age Definitions

Table A.3 reports the productivity life cycle of farms in North and South Vietnam using three different measures of age. The baseline measure, discussed in the main text, constructs household age as the average of household members weighted by their time spent working on household crops. The Household Head measure constructs age as the age of the member identified as the household head. The Average measure constructs age as the simple average across household members. The productivity measure is normalized in each region and year such that the regressions do not capture time trends.

Table A.3: Farm Life Cycle

	(1) log TFP	(2) log TFP	(3) log TFP
Age (North)	0.0218***	0.0333***	0.0137**
	(0.00696)	(0.0117)	(0.00535)
Age (South)	0.0435***	0.0451**	0.0235***
	(0.0112)	(0.0221)	(0.00851)
Age^2 (North)	-0.000305***	-0.000374***	-0.000229***
	(0.0000740)	(0.000119)	(0.0000629)
Age^2 (South)	-0.000570***	-0.000542**	-0.000373***
	(0.000109)	(0.000219)	(0.0000921)
Age Definition	Baseline	Household Head	Average
R^2	0.0327	0.0141	0.0257
Observations	10203	9201	10520

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 a region fixed effect. log TFP is normalized at the region-by-year level. Household Head measures age as the age of the household member identified as the head of household. Average measures age as the average age of all household members. Column (2) excludes households where the head of household is older than 70.

We find that in all three cases the two main observations in the main text hold. First, household productivity life cycles in both the North and the South display a hump-shaped pattern where households quickly increase productivity when they are young and then decline at old ages. Second, the dynamics of farms in the South are much sharper than in the North, where productivity tends to be flatter over the farm's life cycle.

A.3 Differences in Land Quality

Table A.4 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 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 A.4 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.

Table A.4: 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.

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 4 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.

B 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:

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

C Other Quantitative Results

C.1 Sensitivity of Calibration Moments to Parameters

Table C.5 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. The table also shows 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.5: Sensitivity of Moments to Calibrated Parameters (%)

	ψ	ζ	λ	κ^P	κ^O	σ_z	σ_v
Land Share	-0.1	0.4	0.3	0.2	-0.1	0.9	0.9
Reg Coeff	0.5	-3.0	-1.3	-0.4	0.3	-7.5	8.6
Avg Growth	-2.2	5.9	5.8	0.1	0.0	-0.1	-2.4
TFP - Peren	-0.1	-0.4	0.3	3.0	0.0	0.0	-0.1
TFP - Other	0.0	0.2	0.1	0.0	3.1	0.1	0.1
Std TFP	-0.4	1.7	1.0	0.2	-0.1	3.9	3.9
Std Growth	0.0	0.0	0.1	0.0	0.0	0.0	7.4

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.6. 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.6: Moments Calibrated to North Vietnam

	Model	Data
Avg Growth (%)	2.63	2.62
Std Growth	82.5	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
Relative 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)

Notes: Where applicable, moments are first reported for rice farms, followed by those for perennials and then other crop farms.

The parameters in the re-calibrated model are summarized in Table C.7. 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 smaller payoff. The lower return to investment through λ explains the lower farm dynamism in the North compared with the South.

Table C.7: Parameters Calibrated to North Vietnam

Parameter		North	South
Discount Rate	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	1.32	1.51
Crop Preference Elasticity	θ	1.87	1.69
Preference Shifter	η^i	(1.00 , 0.57 , 0.87)	(1.00 , 0.65 , 0.83)
Crop-Specific Productivity	κ^i	(1.00 , 0.31 , 0.44)	(1.00 , 1.43 , 0.34)
Permanent Productivity	σ_z	1.27	1.39
Random Productivity	σ_v	2.56	2.34
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

Notes: Where applicable, parameters are first reported for rice farms, followed by those for perennials and then other crop farms.

Aggregate productivity. The re-calibrated model generates a productivity gap between north and south Vietnam that matches closely the data. Following equation (4) for aggregate output, aggregate total factor productivity in the calibrated economy is calculated as:

$$\frac{A^{North}}{A^{South}} = \frac{Y^{North}/(L^{North})^{\alpha\gamma}}{Y^{South}/(L^{South})^{\alpha\gamma}} = 41.4\%,$$

whereas this ratio is 42% in the data. This implies that the parsimonious re-calibrated model is able to account for the bulk of productivity differences between north and south Vietnam.

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). 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 $\sigma_\varepsilon = 0$. Setting σ_ε as in the benchmark economy does not substantially alter the results, which implies a relatively small role for measurement error potentially captured in the random component of distortions σ_ε.

Table C.8: Comparison with Undistorted Economy

	Benchmark	Undistorted Economy
Productivity	1.00	6.96
Avg Growth (%)	6.25	0.73
Std Growth (%)	77.1	76.8
Std log TFP	1.00	0.89
Reg coefficient: growth on log TFP	−34.1	−41.0
Top 10% Land Share (%)	35.3	87.0
Relative Measured TFP	(1.00 , 1.21 , 0.75)	(1.00 , 1.23 , 0.65)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(48.0 , 46.8 , 5.2)

Notes: Where applicable, moments are first reported for rice farms, followed by those for perennials and then other crop farms.

The undistorted economy is around 7 times as productive as the benchmark economy. Table 7 shows that the gains from removing static misallocation in the benchmark economy is around 66% 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, differences in the productivity distribution alone do not account for the remainder because of complementarities between the channels.

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 correlated 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.

C.4 Asymmetric Effects of Elasticity of Distortions

The baseline experiment shows that increasing the elasticity of distortions ρ from 0.86 in the South to 0.96 in the North can explain a large share of the productivity gap between the two regions, despite the increase being relatively small. Mechanically, the large productivity cost from increasing ρ is driven by the disincentivizing effect of correlated distortions on investment (Farm Ability). As ρ increases farms invest less because the incremental increase in profits becomes smaller. At the extreme, when $\rho \to 1$ farms have no incentive to invest because the entirety of additional profits is absorbed by higher distortions. This leads to an increasing impact of ρ on productivity that is maximized as ρ gets closer to one.

Table C.9 shows the asymmetric impact of increasing and decreasing ρ on productivity through each channel. The values are reported as log changes in productivity (rather than percent changes) for comparability. The difference in the effects is mainly driven by the farm

Table C.9: Increasing and Decrease Elasticity of Distortions

	Increase $\rho = \rho + \Delta\rho$	Decrease $\rho = \rho - \Delta\rho$
Factor Misallocation	-0.21	0.18
Crop Choice	-0.02	0.02
Farm Ability	-0.39	0.24
Total	-0.59	0.46

Notes: Values report the log change in output. The elasticity of distortions is set to the South benchmark value, $\rho = 0.86$, and the change in ρ is set equal to $\Delta\rho = \rho^{North} - \rho^{South} = 0.1$. All other parameters are set to the benchmark calibration values.

ability channel due to the disincentivizing effect of ρ on ability investment.

C.5 Model Extensions

We consider two model extensions to capture the full farm life-cycle dynamics of productivity and the potential intergenerational transmission of ability.

C.5.1 Farm Life Cycle Dynamics

In Section 3, we show that farm productivity is hump shaped over the life cycle with the productivity of young farms increasing quickly and then deteriorating as the farm reaches older ages. Our baseline model focuses on the initial buildup of farm productivity through investments in farm ability but does not account for the decline in productivity of older farmers. We show that the main model results are relatively unchanged if we extend the model to incorporate this feature.

Model. We extend the model to allow for life cycle dynamics following a similar structure of aging as in Acemoglu et al. (2018). Farmers initially enter as young age ($j = Y$) farmers and then with probability ϕ transition to old age ($i = O$) farmers. Old age acts as an absorbing state that all farmers eventually reach (if they do not exit), albeit at different points of time. Young farmers operate as described in the main text while old farmers have ability $a_{f,t} = 1$, regardless of their previous ability h or investment. Farm ability is now given by

$$a^j_{f,t} = 1_{j=Y}\lambda^h + 1_{j=O}. \tag{C.1}$$

The age structure allows us to capture the dynamics observed in the data, in a reduced form. Intuitively, the transition to old age could capture the deterioration of physical abilities of older farmers.

The model is otherwise as described in the main text. The equilibrium characterization is similar with the exceptions that the value function now accounts for the possibility of transitioning to old age and the type distribution now accounts for farms in old age.

Quantitative analysis. We consider an alternative calibration of the model to focus more on the farm productivity life cycle. The preference shifters η^i, preference curvature θ and all other parameters follow the baseline calibration procedure. We re-calibrate the jointly chosen parameters $\{\lambda, \psi, \zeta, \kappa^i, \sigma_z, \sigma_v\}$ as well as the transition probability ϕ to target a new set of moments. In addition to the baseline moments, we add two moments: (i) the average productivity of 36-40 year old farmers is 0.45 log points higher than the average productivity of 25 and younger farmers and (ii) the average productivity of 65 and older farmers is equal to average productivity of 25 and younger farmers (both from Figure 3). We also remove the moments on the average TFP growth of farms. Table C.10 reports the parameter estimates.

The estimated transition to old age is around 3% indicating that farms spend an average of 33 years at young age. The main difference relative to the baseline parameters is that the estimated ability step size increases from 1.51 in the baseline to 1.68 in the extended model, which is necessary to offset some of the negative growth from aging. Figure C.2 reports the relationship between farm TFP and age. The figure highlights the same hump-shaped dynamics as in the data.

The shape of farm dynamics is by construction since key features of the life-cycle pro-

Table C.10: Parameters with Farm Life Cycle Targets

Parameter		Value
Transition to Old	ϕ	0.03
Investment Level	ψ	3.76
Investment Curvature	ζ	2.79
Ability Step Size	λ	1.68
Crop Preference Elasticity	θ	1.72
Preference Shifter	η^i	(1.00 , 0.64 , 0.83)
Crop-Specific Productivity	κ^i	(1.00 , 1.54 , 0.36)
Permanent Productivity	σ_z	1.35
Random Productivity	σ_v	2.34

Notes: Where applicable, parameters are first reported for rice farms, followed by those for perennials and then other crop farms.

ductivity profile are targeted in the calibration. The purpose of the recalibration is to assess its impact on the main results. The main experiment adjusts distortions in the benchmark economy, calibrated to south Vietnam, to match distortions in north Vietnam. Table C.11 compares moments in the extended model benchmark economy, calibrated to the South, with the counterfactual economy, which adjusts distortions to match the values in the North.

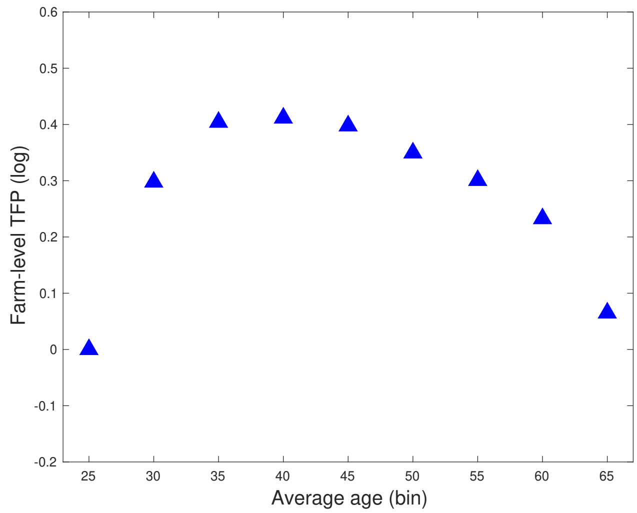
Table C.11: Comparing Counterfactual Moments with Life Cycle Targets

	Benchmark	Counterfactual	Data
Productivity	1.00	0.56	0.42
Avg Growth (%)	3.46	2.32	2.62
Std Growth (%)	78.1	77.4	89.2
Std log TFP	0.99	0.90	0.84
Reg coefficient: growth on log TFP	−33.9	−41.2	−48.2
Top 10% Land Share (%)	35.5	30.5	38.3
Relative Measured TFP	(1.00 , 1.22 , 0.75)	(1.00 , 1.09 , 0.74)	(1.00 , 0.69 , 0.80)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(74.3 , 7.6 , 18.1)	(75.1 , 5.0 , 19.9)

Notes: Where applicable, moments are first reported for rice farms, followed by those for perennials and then other crop farms.

The table highlights that the results in the extended model are in line with the baseline model. Aggregate productivity drops by 44%, only slightly less than in the baseline experiment (47%). We also find similar dynamics when comparing the other moments with the

Figure C.2: Farm Productivity Life Cycle



Notes: Age bins are $\{\leq 25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, > 60\}$ and plotted according to the oldest age in the group and 65 for the oldest group. The average of farm-level log TFP is calculated using simulated data (as described in Section 5) for 100,000 farms.

baseline experiment. The counterfactual economy is almost able to entirely replicate the farm crop distribution and accounts for around half of the change in the standard deviation of TFP, the regressions coefficient of growth on TFP, and the top 10% land share (with the same caveat as the baseline experiment).

The table also shows the average growth rate in the benchmark calibration economy and the counterfactual economy. Unlike the main text, this is no longer a moment that is directly targeted in the calibration. As discussed in Section 6.5, the growth rate of productivity is potentially related to factors unrelated to ability investment in the model. Table C.11 provides an extreme view on the magnitude of these other factors since it attributes none of the non-life cycle growth to farmer investment. The growth rate lies within the bounds considered in the robustness exercise in Section 6.5. Overall, the evidence in Table C.11 is reassuring about the robustness of the main results.

C.5.2 Entrant Ability

In the baseline model, entrants start at the lowest ability node and then progress to higher nodes through investment. In practice, we might expect that some ability is passed on through generational learning, such that some entrants are more productive than others. We show a simple extension of the model that incorporates this feature and find that the quantitative results remain relatively unchanged.

Model. Rather than entering with ability $a = \lambda^0$, we allow entrants to draw ability $a = \lambda^h$ where $h \in \{0, 1, ..., \tilde{h}\}$ is drawn from distribution $m(h, \tilde{h})$ and $\tilde{h}$ is the ability node of the exiting farmer the entrant replaces. We include $\tilde{h}$ as the upper bound to capture the intuition that entrants are learning from the previous generation of (exiting) farmers and note that the distribution would be the same if we instead had entrants learn from active farms, since exit is random. Since our goal is to show the robustness of the baseline results, we set the distribution of entrant productivity to be uniform between 0 and $\tilde{h}$, where we expect that this would tend to overstate the persistence in ability over time.

The model is otherwise as described in the main text. The equilibrium characterization is similar with the exceptions that the type distribution now accounts for entry into higher nodes. Additionally, we assume that entrants draw their predecessors permanent productivity z and preferences η^i and that entrants can only deviate from the crop choice of their predecessor by accepting ability $h = 0$, for tractability. However, quantitatively, this assumption has little impact on the results.

Quantitative analysis. Given that the model parameters are the same as in the main text, the calibration procedure is unchanged. The main difference with the baseline parameters is that the step size of ability improvements λ increases. This is necessary to match the same average growth rate since entrants now start at higher nodes where investment, and growth, would otherwise be lower (see Figure 4).

The main experiment adjusts distortions in the benchmark economy, calibrated to south Vietnam, to match distortions in north Vietnam. Table C.12 compares moments in the extended model benchmark economy, calibrated to the South, with the counterfactual economy, adjusting distortions to match values in the North.

Table C.12: Comparing Counterfactual Moments with Entrant Ability

	Benchmark	Counterfactual	Data
Productivity	1.00	0.52	0.42
Avg Growth (%)	6.24	4.66	2.62
Std Growth (%)	77.1	76.9	89.2
Std log TFP	1.01	0.92	0.84
Reg coefficient: growth on log TFP	−34.1	−39.6	−48.2
Top 10% Land Share (%)	35.2	30.5	38.3
Relative Measured TFP	(1.00 , 1.20 , 0.75)	(1.00 , 1.00 , 0.74)	(1.00 , 0.69 , 0.80)
Farm Share by Crop (%)	(49.1 , 33.1 , 17.8)	(74.1 , 7.8 , 18.1)	(75.1 , 5.0 , 19.9)

Notes: Where applicable, moments are first reported for rice farms, followed by those for perennials and then other crop farms.

The results show that allowing for entrants with higher ability results in a slightly greater productivity loss (48%) than the benchmark calibration (47%). The larger productivity loss arises because the transfer of ability to entrants creates a positive spillover in which investment further shifts the productivity distribution through improving entrants' productivity. Otherwise, the results are consistent with the baseline results and show the same conclusions.