ESSAYS ON AGRICULTURAL MISALLOCATION

By

Tram Hoang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food, and Resource Economics – Doctor of Philosophy
Economics – Dual Major

2024

ABSTRACT

This dissertation is comprised of two essays, which explore the cost of factor misallocation in agriculture in Vietnam. The first essay addresses the measurement error in estimates of resource misallocation. The second essay is a policy evaluation of the Vietnam 2013 Land Law studying the impact of improved tenure security on land markets, land allocative efficiency, employment, and household welfare.

In the first essay, I examine misallocation by investigating how measurement errors in output and inputs affect the estimation of agricultural productivity loss associated with resource misallocation. I find that measurement errors account for a substantial part of the estimated total factor productivity (TFP) variations (30-45% at the national level). Correspondingly, failing to account for measurement errors would considerably overestimate the gains from resource reallocation. Based on the preferred Two-Stage least squares (2SLS) estimation of the production function, measurement errors in both output and inputs will lead to an overstatement of production gains by 2-3-fold if not adjusted in productivity estimation. The results are consistent regardless of whether the analysis is explored by analyzing household productivity variation across years or across households within local communes. The findings caution against relying on estimates unadjusted for measurement error of potential gains from reallocation in cost-benefit analysis of reallocation. Certain caveats and assumptions of the analysis are discussed in the essay.

The second essay investigates the impact of increased tenure security on land transactions and the ensuing productive efficiency, as well as its spillover effects on the labor market and overall household welfare. Vietnam's 2013 Land Law, which extends the lease term for usufruct rights for annual land from 20 years to 50 years, provides the opportunity for difference-in-differences (DID) identification. This involves the first difference between annual land and perennial land,

and the second difference between before and after the law was passed, to study the effect of increased land security. Plot-level data are available for the land transfer outcomes (lease out, lease in, sold, purchased). For the welfare outcomes, the impacts of the land law are estimated at the household level. Household outcomes include the household's food expenditure per capita as well as indicator variables regarding labor (wage labor, nonfarm wage labor, wage labor in agriculture, wage labor in commune, wage labor in province, and wage labor outside of province), and whether households have their own business. Plot-level DID results reveal that annual plots are 3 (or 6) percentage points more likely to be leased out (or sold) as a consequence of the law, while there is no significant effect on the likelihood of annual plots being leased in or purchased. This result is in line with the expectation that the heightened security generated by the law is a supply factor affecting the supply of land. As both rental and sale markets are found to transfer land from less productive to more productive farmers, the more active land markets incentivized by the law are expected to enhance land use efficiency. Household-level analysis shows that the passage of the law is associated with a shift from self-employed farm work to wage employment, especially agriculture-related wage work that is closer to home. Household food expenditures per capita are also found to increase due to the law. Given these findings, the study suggests that the law can be a low-cost tool in increasing land market participation with some effects on the labor market and improving welfare.

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my advisor Dr. Songqing Jin for his treasured wisdom, experience, and colossal amount of patience during the long course of my Ph.D. degree. I could not have finished this dissertation without his unwavering support. I would like to thank my Graduate Program Director and one of my committee members Dr. Nicole Mason-Wardell for her guidance and unrelenting advocacy. The other two members of my committee Dr. Mark Skidmore and Dr. Prabhat Barnwal have my gratitude for their advice and knowledge. Additionally, I would like to express my gratitude to Dr. Klaus Deininger of The World Bank for making this project possible.

On a personal note, I owe my friends Dr. Liao Zhang and David Jennett a debt of gratitude for their help and encouragement during my most difficult times. Their kindness meant the world to me. Finally, I would like to thank my father, mother, brother, all my aunts and uncles, and last but not least, my better half Dr. Patrick Buck, whose love and support have carried me along the entire way.

TABLE OF CONTENTS

ESSAY 1

MISALLOCATION OR MEASUREMENT ERROR: EVIDENCE FROM VIETNAM'S AGRICULTURE

1.1. Introduction

Optimization of resource allocation maximizes the total production of an economy endowed with finite productive resources and efficient productive individuals heterogeneous in total factor productivities (TFP) (Restuccia and Rogerson 2008; Hsieh and Klenow 2009). Misallocation represents a departure from this optimal allocation, resulting in income loss and increased dispersion in output per worker. In practice, large variations in productivity have been observed across sectors and between establishments in narrowly defined industries within a sector; moreover, the variation tends to be greater in poor countries than in rich countries (Restuccia and Rogerson 2008, 2013; Hsieh and Klenow 2009; McMillan, Rodrik, and Verduzco-Gallo 2014; Porzio 2016; Bento and Restuccia 2017).

Motivated by these observations, a substantial and growing body of literature on resource misallocation has emerged (Restuccia and Rogerson 2008; 2013; Hsieh and Klenow 2009; Banerjee and Moll 2010; Bartelsman, Haltiwanger, and Scarpetta 2013; Hopenhayn 2014; and Bento and Restuccia 2017). The emerging evidence from these empirical studies suggests that misallocation has played a significant role in explaining the low level of aggregate productivity in low-income countries and the world's income gap.

The agricultural sector holds special significance and importance in the developing world (Gollin, Parente, and Rogerson 2002). According to estimates from the International Labor Organization (ILO), approximately 58.8% of the labor force was engaged in the agriculture sector in low-income countries in 2022, with figures of 56% for Eastern and Southern Africa and 45% for Western and Central Africa. Misallocation within this sector is often attributed to market constraints and local restrictions that hinder the efficient distribution of productive resources (Banerjee and Moll 2010; Restuccia and Rogerson 2017), and constraints and imperfections are notoriously prevalent in the agricultural land and labor markets in developing countries.

Given the significance of the agricultural sector in these economies, a considerable portion of the literature has focused on misallocation within agriculture (see for example, Gollin, Lagakos, and Waugh 2014; Adamopoulos and Restuccia 2014; Restuccia 2016 for the assessments of the potential income gap resulting from such misallocation). Previous studies reveal significant output gains through reallocation, with estimates ranging from 57% in China and 80% in Vietnam to 140% in Ethiopia and 260% in Malawi (Ayerst, Brandt, and Restuccia 2020; Adamopoulos et al. 2022; Chen, Restuccia, and Santaeulàlia-Llopis 2022, 2023).

An empirical challenge in assessing misallocation lies with the measurement of factor elasticities and TFP, which hinge on accurate measures of inputs and production outputs. If the input and outputs are measured with errors, the estimated productivity gain resulting from resource reallocation could be biased (Gollin and Udry 2021). In reality, measurement errors are a ubiquitous part of data analysis and pose a significant estimation challenge.

In agriculture, inaccurate estimates or imperfect reporting are significant sources of measurement errors in both inputs and outputs (Abay, Bevis, and Barrett 2021). One traditional method to reduce errors in production input data is to require respondents to maintain continuous production diaries (Deininger et al. 2012). To reduce the widespread measurement errors of land size, researchers rely on methods such as the compass-and-rope approach (Dillion et al. 2019) or, more recently, the use of Global Positioning System (GPS) data (Carletto et al. 2013; 2015; Kilic 2017). The crop-cut method is one approach to address the measurement error of output data (Abay et al. 2019; Desiere and Jolliffe 2018; Gourley et al. 2019).

There is a growing body of recent literature that explores the impact of measurement error on estimates of the relationship between farm size and productivity (Desiere and Jolliffe 2018; Abay et al. 2019; Ayalew et al. 2024). Abay (2020) investigates the relationship between measurement error and marginal returns to modern agricultural inputs. Findings from these studies suggest a correlation between output and input use and their respective measurement errors, highlighting the possibility of non-classical measurement error. Cohen (2019) argues that even GPS measurements may still be subject to classical measurement errors, primarily due to 'position error' in satellites and human errors in GPS device operation (Bogaert, Delincé, and Kay 2005; Bogaert, Delincé, and Kay 2005; Keita and Carfagna 2009).

Despite the widespread recognition of measurement errors in agriculture and the increasing literature dedicated to addressing this issue in studies examining the relationship between farm size and productivity, the issue of measurement error has largely been overlooked in the emerging literature on resource misallocation (e.g., Ayerst, Brandt, and Restuccia 2020; Adamopoulos et al. 2022; Chen, Restuccia, and Santaeulàlia-Llopis 2022, 2023). There are a few exceptions, such as Bils, Klenow, and Ruane (2021), Aragon, Restuccia, and Rud (2021), and Gollin and Udry (2021). Bils, Klenow, and Ruane (2021) use data from manufacturing sectors in both India and the U.S. to identify the measurement error stemming from the rates of revenue and input growth in response to productivity shocks. They find that measurement error contributes to a greater dispersion in revenues per input in the U.S. and the potential gains from reallocation undergo a more significant reduction in the adjustment process compared to India.

Using agricultural data from Tanzania and Uganda, Gollin and Udry (2021) find that the misallocation diminishes significantly after accounting for measurement errors in their study. Apart from the aforementioned measurement and reporting errors, Gollin and Udry (2021) identify two other sources of measurement errors in agricultural production data. One is linked to the stochastic nature of agricultural production, which is associated with the vagaries related to weather, pests, and crop diseases. The other is related to shocks occurring late in the production season after farmers have already made their production decisions (Gollin and Udry 2021). These late-season shocks could include adverse weather events, pests, or disease shocks that occur sufficiently late in the growing season that farmers are unable to effectively respond to them. They find that potential output in an efficiently allocated scenario is overestimated by a factor of 2.6 in Tanzania and an even higher factor of 3.7 in Uganda. Similar to the findings of Bils, Klenow, and Ruane (2021), they find that this overestimation is more pronounced in less wealthy countries where measurement errors exhibit greater variability.

The method of Gollin and Udry (2021) heavily relies on a structure of plot data. They use data from multiple plots growing the same crop managed by the same individual or household within a season of a year, creating a panel. This panel structure ensures that market distortions are held constant so that the distortion-induced variance is eliminated. Then by employing a normalization that involves TFP and factor-specific productivity, they infer the variances in measurement errors in outputs and inputs to correct misallocation accordingly. They then proceed to infer productivity and measurement error-induced variances.

In contrast, Aragon, Restuccia, and Rud (2021) present several arguments against the use of plot-level data in misallocation calculation[1]. They provide empirical evidence suggesting that plot-level data tend to overestimate the impact of measurement errors, leading to estimates that do not agree with the literature at large.

[1] According to Aragon, Restuccia, and Rud (2021), the main reasons why using plot-level data is likely to exaggerate the measurement error include (1) a much higher level of productivity dispersion to begin with, and (2) the practical issue to separate inputs by plots.

In this essay, I follow Aragon, Restuccia, and Rud (2021) and utilize panel data at the household level from Vietnam to identify within-household variance in measurement errors. As robustness checks, I also estimate the within-commune variance in measurement errors. Additionally, I differentiate the estimates for all crops versus rice crops, as well as for the Southern region versus the Northern region. My method relies on three key assumptions.

The first assumption is that the measurement errors are classical measurement errors, orthogonal to their true value and with each other. This assumption is crucial for effectively separating and identifying the variance of measurement error.[2] The second assumption posits that productive households operate efficiently within the constraints they face. The final assumption relates to intermediate inputs and facilitates the identification of a subset of measurement error variances. It assumes minimal change within a household over time and little variance in the shadow prices and elasticities for intermediate inputs within a commune. By holding intermediate input use constant, I can identify variances of measurement errors in intermediate inputs and output. The assumptions can be empirically tested against conditions derived from the household model.

Finally, using crop data from Vietnamese farming households, I aim to infer the actual output gap after adjusting for measurement error. I find that nationwide, up to 45% of the variation in the standard TFP estimate comes from measurement error and allocation-unrelated elements. Using the production residual as an estimate for TFP, the raw estimate of potential gains from reallocating is 139% of the observed revenue for all crops. Heterogeneity analysis between the Northern and Southern regions and between all crops and rice investigates regional comparative advantages in different crops.

[2] One limitation of this study is that it cannot deal with non-classical measurement errors. Previous studies such as Gollin and Urdy (2021) and Aragon, Restuccia, and Rud (2021) make the same assumption on measurement errors in their studies.

The essay makes three contributions. First, it contributes to the scant literature that directly addresses measurement error in misallocation analysis, and it is the first such study using data from Vietnam[3]. Second, as opposed to the usual requirement for high-quality data to address measurement error, this method is cost-saving on data by achieving misallocation adjustment for measurement error using household-level survey data. While plot- and parcel-level data tend to overestimate the impact of measurement error, household-aggregate data reduce both the magnitude and the dispersion of measurement error (Aragon, Restuccia, and Rud 2021). The household-level analysis also allows for convenient interpretation and easy placement within the tradition of the literature on misallocation. Regarding methodology, a key assumption in the essay involving intermediate inputs is observationally inspired, testable, and not too restrictive. Third, by exploring the difference between the South and the North, and between all crops and rice crops only, the findings of this study are of policy relevance. Given the historical difference in property rights between the North and South regions, whether and to what extent resource misallocation differs between the two regions is of academic and policy significance.

The remainder of the essay is as follows. Section 1.2 outlines the theoretical model. Section 1.3 describes the data set. Section 1.4 presents the estimation strategy. Section 1.5 presents the results, and section 1.6 concludes the essay.

1.2. Theoretical model

1.2.1. Production function and measured total factor productivity

The model will assume a Cobb-Douglas production function of the form:

$$Y_{ht}^{o} = e^{\epsilon_{Yht}} e^{\omega_{Yht}} \boldsymbol{e^{\beta W_{ht}}} (L_{ht})^{\alpha_{Lht}} (N_{ht})^{\alpha_{Nht}} (M_{ht})^{\alpha_{Mht}}$$

[3] Ayerst, Brandt, and Restuccia (2020) investigates agricultural misallocation in Vietnam without adjusting for measurement error.

where Y_{ht}^{o} is the observed output of household h at time t, $J \in \{L, N, M\}$ are land, labor, and intermediate inputs (seeds, saplings, fertilizers, pesticides/herbicides, energy, irrigation, maintenance, and other), respectively, used in the household's crop production. Output elasticities of factor inputs α_{Jht} are allowed to vary across households and time to capture the differences in land and other input's quality and intensity. $\boldsymbol{W_{ht}}$ are observable household, land characteristics, and late-seasoned shocks. The parameter ω_{Yht} denotes the total factor productivity (TFP) and is unobservable to the researcher but known to the household. Finally, the classical error term ϵ_{Yht} is unobservable to the researcher as well as unknown to farmers. Rewrite

$$Y_{ht}^{o} = Y_{ht} e^{\epsilon_{Yht}}$$

where Y_{ht} denotes the household's true output. Measurement errors in inputs J are similarly modeled:

$$J_{ht}^{o} = J_{ht} e^{\epsilon_{Jht}}$$

where J_{ht}^{o} is the reported input, J_{ht} is the true value of input J of household h in time t and ϵ_{Jht} is the corresponding measurement error in factor input J.

All the components of the vector $(\epsilon_{Yht}, \epsilon_{Jht})$ are subject to classical measurement error assumptions and orthogonal with each other. The production is re-expressed with lowercase to represent their logarithms.

$$\begin{aligned} y_{ht}^{o} &= \epsilon_{Yht} + \omega_{Yht} + \boldsymbol{\beta W_{ht}} + \sum_{J} \alpha_{Jht} j_{ht} \\ &= \epsilon_{Yht} + \omega_{Yht} + \boldsymbol{\beta W_{ht}} + \sum_{J} \alpha_{Jht} j_{ht}^{o} - \sum_{J} \alpha_{Jht} \epsilon_{Jht} \end{aligned} \tag{1}$$

An estimation of the production function estimates the coefficients β and α_J the expected values of α_{Jht}. The production residuals provide estimates for the household TFPs that are calculated as:

$$\ln\widehat{TFP}_{ht} = y^o_{ht} - \widehat{\boldsymbol{\beta}}\boldsymbol{W_{ht}} - \sum_J \widehat{\alpha_J} j^o_{ht}$$

$$= \epsilon_{Yht} + \omega_{Yht} + \boldsymbol{\beta}\boldsymbol{W_{ht}} + \sum_J \alpha_{Jht} j^o_{ht} - \sum_J \alpha_{Jht}\epsilon_{Jht} - \widehat{\boldsymbol{\beta}}\boldsymbol{W_{ht}} - \sum_J \widehat{\alpha_J} j^o_{ht}$$

$$= \epsilon_{Yht} + \sum_J \widehat{\alpha_J}\epsilon_{Jht} + \omega_{Yht} + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\boldsymbol{W_{ht}} + \sum_J (\alpha_{Jht} - \widehat{\alpha_J}) j^o_{ht} - \sum_J (\alpha_{Jht} - \widehat{\alpha_J})\epsilon_{Jht}$$

$$= \epsilon_{Yht} + \sum_J \widehat{\alpha_J}\epsilon_{Jht} + \omega_{Yht} + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\boldsymbol{W_{ht}} + \sum_J (\alpha_{Jht} - \widehat{\alpha_J}) j_{ht}$$

In the limit, with consistent estimators of the production elasticities, the production residuals approach:

$$\ln\widehat{TFP}_{ht} \to \left(\epsilon_{Yht} + \sum_J \alpha_J \epsilon_{Jht}\right) + \omega_{Yht} + \sum_J (\alpha_{Jht} - \alpha_J) j_{ht}$$

The second and third terms in the limit ω_{Yht}, and $\sum(\alpha_{Jht} - \widehat{\alpha_J}) j_{ht}$ reflect the true factor-neutral and factor-specific productivities, and are both assumed known to farmers. Together, they are informative about household productivities and form the true TFP, that is $\ln TFP_{ht} = \omega_{Yht} + \sum_J (\alpha_{Jht} - \alpha_J) j_{ht}$. The first term $\epsilon_{ht} \equiv \epsilon_{Yht} + \sum_J \alpha_J \epsilon_{Jht}$ with $E(\epsilon_{ht}) = 0$ and $Var(\epsilon_{ht}) = Var(\epsilon_{Yht}) + \sum_J \alpha_J^2 Var(\epsilon_{Jht})$ is the aggregate measurement error in output and inputs. As far as allocative efficiency is concerned, ϵ_{ht} provides no information on efficiency gain through reallocation of factor resources and only serves as noise.

One is often interested in the distribution of measured TFP because it is a tool to characterize efficient factor allocation. With the classical measurement error and orthogonal assumptions, the variance of measured TFP can be decomposed into two parts:

$$Var(\ln\widehat{TFP}_{ht}) = Var\,(\epsilon_{ht}) + Var\,\left(\omega_{Yht} + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\boldsymbol{W_{ht}} + \sum (\alpha_{Jht} - \widehat{\alpha_J}) j_{ht}\right) \quad (2)$$

The first term shows that measurement error ϵ_{ht} becomes an additional source of variation in the production residual that does not represent actual variation in productivity. For allocative efficiency purposes, it is irrelevant. To see how this added variance presents a problem, section 1.2.2 characterizes the efficiency allocation and outcome, followed by the implications of measurement error on perceived distortion and estimates of potential income gain through factor reallocation in section 1.2.3. Then section 1.2.4 details how the measurement error components ϵ_{Yht} and ϵ_{Jht} of ϵ_{ht} can be measured and removed from the TFP variance estimate in (2).

1.2.2. Characterization of efficient output and allocation of factors

In this section, to systematically characterize the optimal allocation, I will assume away heterogeneity in output elasticities. Furthermore, in this exercise, the time subscript t is implied and dropped for convenience purposes. Consider a Cobb Douglas production technology in year t of the form $Y_h = \exp \omega_h L_h^{\alpha_L} N_h^{\alpha_N} M_h^{\alpha_M}$. In the efficient allocation, the optimal inputs and output solve the following problem:

$$\underset{L_h, X_h}{\text{Max}} \sum_h \exp \omega_{Yh} \prod_J J^{\alpha_J} \text{ s.t } \sum_h J_h = \bar{J}$$

The first order condition implies that efficient allocation requires equalized marginal product of each factor across households, i.e.

$$\frac{Y_h^*}{J_h^*} = \frac{Y_g^*}{J_g^*} = \frac{\bar{Y}}{\bar{J}}$$

Let $s_h \equiv \frac{J_h^*}{\bar{J}}$ define the optimal share of resources for each household from the total pool of resources. It follows that $s_h = \frac{Y_h^*}{\bar{Y}} = \frac{\exp \omega_{Yh} \prod_J (s_h \bar{J})^{\alpha_J}}{\bar{Y}}$ and therefore is constant across factor inputs J. The social planner's solution for s_h is $s_h = \left(\frac{\prod_J \bar{J}^{\alpha_J}}{\bar{Y}}\right)^{\frac{1}{1-\sum_J \alpha_J}} \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)$.

Since household shares add up to 1, i.e.

$$\sum_h s_h = 1 \Rightarrow \left(\frac{\prod_J \bar{J}^{\alpha_J}}{\bar{Y}}\right)^{\frac{1}{1-\sum_J \alpha_J}} \sum_h \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right) = 1 \Rightarrow \left(\frac{\prod_J \bar{J}^{\alpha_J}}{\bar{Y}}\right)^{\frac{1}{1-\sum_J \alpha_J}} = \frac{1}{\sum_h \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)},$$

this share can be alternatively expressed as $s_h = \frac{\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)}{\sum_h \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)}$. The idea of this expression is simple and intuitive: a household's optimal share of input factor is strictly increasing in its TFP and is proportionate to its productivity relative to other households.

1.2.3. Implications of measurement error

In the absence of measurement error, the first-order condition provides several ways to measure how far the existing allocation is from optimality. Productivities, $\frac{Y_h}{J_h}$, are proportional to the marginal product of factors, and are constant across households in efficient allocation. Similarly, the cross-factor ratios, $\frac{J_{ht}}{I_{ht}}$, are indicative of allocative efficiency. The general idea is that, in optimal allocation, productivities, and cross-factor ratios are equalized across households, and that a higher dispersion indicates a higher level of distortion, and thus factor misallocation.

In the presence of measurement error, however, these observations would be misguided, since measurement error increases the dispersion of observed factor productivities and observed cross-factor ratios alike. For illustration, let us look at the variance of the logarithm of the measured crop yields:

$$Var\left(\ln\left(\frac{Y_h^o}{L_h^0}\right)\right) = Var(y_h^o - l_h^o) = Var(\epsilon_{Yh} - \epsilon_{Lh}) + Var(y_h - l_h)$$

$$> Var(y_h - l_h) = Var\left(\ln\left(\frac{Y_h}{L_h}\right)\right)$$

Similarly, it can easily be proved that the presence of measurement error also generates more dispersion in other observed factor productivities as well as any factor ratio combination. Observing these measures and letting them inform us of the existing level of distortion, therefore, can exaggerate the misallocation problem. If measurement error varies greatly, it can create a lot of noise in the distribution of these distortion measures that otherwise may have low spreads.

Next, I will quantify the overstatement in potential gains from reallocation. If resources are to be optimally distributed, the observed output is:

$$Y_h^{o*} = \exp(\epsilon_h)\exp(\omega_{Yh})\left(\frac{\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)}{\sum_h \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)}\right)^{\sum_J \alpha_J} \prod_J \bar{J}^{\alpha_J}$$

$$= \exp(\epsilon_h)\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)\left(\sum_h \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)\right)^{-\sum_J \alpha_J} \prod_J \bar{J}^{\alpha_J}$$

$$= \exp(\epsilon_h)\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)\left(NE\left(\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)\right)\right)^{-\sum_J \alpha_J} \prod_J \bar{J}^{\alpha_J}$$

In expectation, optimal observed output turns out to be:

$$E(Y_h^{o*}) = E(\exp(\epsilon_h))\left(E\left(\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J} N^{-\sum_J \alpha_J} \prod_J \bar{J}^{\alpha_J}.$$

In practice, researchers use the production function residuals to generate TFP estimates and to generate estimates of optimal factor allocation and output. As shown previously in section 1.2.1, these TFP estimates contain measurement errors and are in fact "inconsistent" with the true TFP.

The estimated optimal output in an allocation planned according to the confounded TFP estimates is $\widehat{Y_h^{o*}} = \exp(\omega_{Yh} + \epsilon_h)\left(\frac{\exp\left(\frac{\omega_{Yh}+\epsilon_h}{1-\sum_J \widehat{\alpha_J}}\right)}{\sum_h \exp\left(\frac{\omega_{Yh}+\epsilon_h}{1-\sum_J \widehat{\alpha_J}}\right)}\right)^{\sum_J \widehat{\alpha_J}} \prod_J \bar{J}^{\widehat{\alpha_J}}$

A similar derivation as above will reveal the expected value of estimated output as:

$$E\left(\widehat{Y_h^{o*}}\right) = \left(E\left(\exp\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J} \left(E\left(\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J} N^{-\sum_J \alpha_J} \prod_J \bar{J}^{\alpha_J}$$

To compare the estimated and true optimal outputs, simply take their ratio:

$$\frac{E\left(\widehat{Y_h^{o*}}\right)}{E\left(Y_h^{o*}\right)} = \frac{\left(E\left(\exp\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J}}{E(\exp(\epsilon_h))}$$

Within the caveat of a decreasing returns to scale Cobb-Douglas production function, $1 - \sum_J \alpha_J > 0$. Apply Jensen's Inequality, it can be seen that

$$\left(E\left(\exp\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J} > \left(\exp\left(E\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J}$$

$$= \exp\left(\left(1-\sum_J \alpha_J\right) E\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)\right) = E(\exp(\epsilon_h)),$$

and therefore $E\left(\widehat{Y_h^{o*}}\right) > E\left(Y_h^{o*}\right)$, i.e optimal outputs are overestimated using the estimated TFP confounded by measurement error. To quantify the magnitude of optimal output overestimation, assume ϵ_h to follow a normal distribution. Then,

$E(\exp(\epsilon_h)) = \exp\left(E(\epsilon_h) + \frac{Var(\epsilon_h)}{2}\right) = \exp\left(\frac{Var(\epsilon_h)}{2}\right)$, and

$$\left(E\left(\exp\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)\right)\right)^{1-\sum_J \alpha_J} = \left(\exp\left(E\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right) + \frac{Var\left(\frac{\epsilon_h}{1-\sum_J \alpha_J}\right)}{2}\right)\right)^{1-\sum_J \alpha_J}$$

$$= \left(\exp\left(\frac{Var(\epsilon_h)}{2\left(1 - \sum_J \alpha_J\right)^2} \right) \right)^{1-\sum_J \alpha_J} = \exp\left(\frac{Var(\epsilon_h)}{2\left(1 - \sum_J \alpha_J\right)} \right)$$

$$\Rightarrow \frac{E\left(\widehat{Y_h^{o*}}\right)}{E(Y_h^{o*})} = \exp\left(\frac{Var(\epsilon_h)}{2\left(1 - \sum_J \alpha_J\right)} - \frac{Var(\epsilon_h)}{2} \right) = \exp\left(\frac{Var(\epsilon_h) \sum_J \alpha_J}{2\left(1 - \sum_J \alpha_J\right)} \right) \qquad (3)$$

It can be concluded that under our set of assumptions, using production residuals as TFP estimates to inform about the efficient allocation overstates the misallocation gap, the magnitude of which depends on the returns to scale $\sum_J \alpha_J$ of the production technology, as well as variance $Var(\epsilon_h)$ of the aggregate measurement error. Elasticities α_J can be consistently estimated in the production function. The remaining task is to calculate the variance of measurement error $Var(\epsilon_h)$, separating it from the actual productivity variance. In order to do so, next I turn to the household's problem.

1.2.4. Separation of productivities and measurement errors

Assume that households are efficient and maximize profit subject to their shadow input prices $p_{Jh} = (p_{Lh}, p_{Nh}, p_{Mh})$ relative to normalized output price $p_Y = 1$.

$$\text{Max}\, p_Y Y_h - \sum_J p_{Jh} J_h$$

The first-order condition of the problem $J_h = \frac{\alpha_{Jh} Y_h}{p_{Jh}}$ implies:

$$Y_h = \left(e^{\beta W_h} e^{\omega_{Yh}}\right)^{\frac{1}{1-\sum_J \alpha_{Jh}}} \prod_J \left(\frac{\alpha_{Jh}}{p_{Jh}}\right)^{\frac{\alpha_{Jh}}{1-\sum_J \alpha_{Jh}}}$$

The output and input solutions in logarithm are:

$y_h = \frac{1}{1-\sum_J \alpha_{Jh}} \left(\beta W_{ht} + \omega_{Yh} + \sum_J \alpha_{Jh} (\ln \alpha_{Jh} - \ln p_{Jh})\right)$ and $j_h = y_h + (\ln \alpha_{Jh} - \ln p_{Jh})$

It is not possible in this paper to empirically separate factor-specific productivities α_{Jh} from distortions in the corresponding market p_{jh}. Therefore, I combine them through the term $\omega_{Jh} \equiv \ln \alpha_{Jh} - \ln p_{Jh}$, then j_h is rewritten as $j_h = y_h + \omega_{Jh}$, where ω_{Jh} represents both a household's ability to use factor J and the idiosyncratic distortion they face in that factor market. I call ω_{Jh} factor J-specific productivity-distortion. A household's profit-maximizing solution for the logarithm of factor demand J is the sum of its total output y_h and factor J- specific productivity-distortion ω_{Jh}.

The observed factor demands and observed production outcome are simply the sum of their true value and their measurement errors.

$$y_h^o = \epsilon_{Yh} + y_h \tag{4}$$

$$j_h^o = \epsilon_{Jh} + y_h + \omega_{Jh}$$

This set of rules breaks down the observable factor demands and production output on the left-hand side of (4) into components that are known $(\omega_{Yh}, \omega_{Jh})$ and unknown $(\epsilon_{Yh}, \epsilon_{Jh})$ to farmers at the time of decision making. For the sake of convenience, some short-hand notations are defined as follows: variances of output and input measurement errors $\left(Var(\epsilon_{Yh}), Var\left(\epsilon_{Jh}\right)\right) = (\sigma_{\epsilon Y}^2, \sigma_{\epsilon J}^2)$, variances of output productivity and input-specific productivities-distortions $\left(Var(y_h), Var\left(\omega_{Jh}\right)\right) = (\sigma_Y^2, \sigma_J^2)$, and covariances between any two productivities ($\left(Cov\left(y_h, \omega_{Jh}\right), Cov\left(\omega_{Ih}, \omega_{Jh}\right)\right) = \left(\sigma_{YJ}, \sigma_{IJ}\right)$.

Since measurement errors are assumed to be independent of each other and of productivities, expressing (4) in terms of variances and covariances provides the following set of equations:

$$Var(y_h^o) = \sigma_{\epsilon Y}^2 + \sigma_Y^2$$

$$Var(j_h^o) = \sigma_{\epsilon J}^2 + \sigma_Y^2 + \sigma_J^2 + 2\sigma_{YJ} \tag{5}$$

$$Cov(y_h^o, j_h^o) = \sigma_Y^2 + \sigma_{YJ}$$

$$Cov(i_h^o, j_h^o) = \sigma_Y^2 + \sigma_{YI} + \sigma_{YJ} + \sigma_{IJ}$$

This system is short of identifying all variances of measurement error, with the number of unknowns exceeding the number of equations by $J + 1$. Further assumptions are needed in order to identify the key parameters of the system.

In the next section, I will present the data used in this study, and explore the potential of market distortions and factor misallocation, before moving on to outline the identification strategies I will apply on this dataset in section 1.4.

1.3. Data

This essay is based on three rounds 2012, 2014, and 2016 of household data from the Vietnam Access to Resources Household Survey (VARHS). The survey is collected by the United Nations University World Institute for Development Economics Research (UNU WIDER) and provides a household panel representative of the rural population in 12 provinces across all regions of the country in 2006. The subsequent waves of data follow up on previously selected households with additional households surveyed in 2008 and 2012.

While I draw on households' all crop farming activities, special focus is also given to the rice crops and those households that exclusively grow rice. A comparison between the aggregate and rice crops provides insights into the magnitude of measurement error impact on lower-level data and aggregate data usage. If what Aragon, Restuccia, and Rud (2021) assert is true, a higher dispersion of measurement error is expected in the rice crop function relative to the household-level aggregate crops.

Physical outputs and revenue are given for each specific crop production for different annual crops, perennial crops, fruits, and forestry, which is aggregated to the household total[4]. Despite that data

[4] The annual crops reported are rice, maize, potato, sweet potato, cassava, peanuts, soybean, vegetables, and other annual crops. The perennial crops include coffee, tea, cocoa, cashew nuts, sugar cane, pepper, rubber, medicinal trees, and other perennial crops. With the exception of aggregate categories like vegetables, other annual crops,

on factor inputs are scarcer in detail for each specific crop in the survey, household level and rice crop information are reported.

Data on factor demand used in this essay includes land use area, labor, and intermediate inputs (i.e. expenses on seeds, saplings, fertilizers, pesticides/herbicides, energy, irrigation, maintenance, and other costs). For production function identification, I control for self-reported land characteristics (distance from the household, land value, irrigation, land use rights certification, crop restriction) and quality (below, average, or above local average) aggregated from plot level, as well as household demographics including household head's age, gender, and educational level. Further controls are weather shocks (drought). Some other household shocks are employed as excluded variables (avian flu, change in commodity price, whether the household's head is sick).

The sample is summarized in Table 1.1. The main analysis is drawn from 8264 household observations in three separate years, 2,887 in 2012, 2861 in 2014, and 2,516 in 2016, which account for 3140 unique crop-growing households in the sample. Geographically, these households belong to 492 administrative communes in 138 districts drawn from 12 provinces from across the regions of Vietnam, which can be divided into North and South regions. On average, households in the sample grow more than two different crops a year. To examine misallocation and measurement error in the rice production, I also focus on the subsample of single rice-crop households[5]. This sample narrows the number of observations down to 2,755 across all three years. Table 1.2, which includes two panels, panel A for any household that grows crops, and panel B for rice-specialized households, gives insights into crop revenues and land, labor, and intermediate input factor demands. All four measures (the first column) are highly skewed to the right with the

fruits, and forestry products, where physical production is not available, physical production and revenue were reported for all of the listed crops.

[5] These households are highly specialized in rice production and the sample makes sure output and input observations are untampered with other crops.

mean several times the median. For starters, the sample averages $1,672 in crop revenue per household, almost 2.5 folds of its median with a huge 3,252 standard deviation[6]. Revenue from rice specialization is smaller in both mean ($1,003) and median ($374). On average, households use less than 1 hectare of land for crop growth, the median is only slightly more than half of that. The distribution has a long right tail suggesting most are small farming households. Labor use on farms is between a third and a half of the year at median and mean. Intermediate input use averages $582 a year with a median of $174. For rice crop specialized households, panel B in table 1.2 reports not only smaller output revenue but in all categories of input use as well due to smaller production scale. The mean land-use area is only half compared to the full sample, labor and intermediate input demands are also smaller. To make a meaningful comparison between the two sets of samples, next, I examine productivities and factor intensity.

Cross factor ratio (labor/land) reveals that on average, labor-land intensity is comparable between all crops and rice crops, spending 386 days per hectare of land. In the median, rice-specialized households use 100 more labor days per hectare of land than in the full sample, suggesting a slightly more labor-intensive technology for rice farming compared to other crops. Unsurprisingly, labor productivity is higher for rice crops in both mean and median, yielding $0.33-$0.80 more each day of labor than with other crops, earning more than $10.15 in revenue at the mean though only $6 at the median. The picture painted by land productivity is not as straightforward since land returns higher revenue in the mean but less in the median for all crops compared to rice crops, yielding between $2,270-$2,466 per hectare of land a year. Most strikingly, the distribution of returns to intermediate input is almost identical between all household crops and rice crop, yielding 4.62 times in revenue at the mean, 3.4 times at the median, and a 6.5 standard deviation. It is

[6] I deflate all VND values of 2012, 2014, and 2016 by factors of 1.3451, 1.4607, 1.4741, respectively, and convert all to 2010 US dollar at the exchange rate of $1 = 18,802VND.

suggestive that intermediate input use is quite robust to crop technologies for the households in the sample.

Figure 1.1 shows a visual presentation of the dispersion in factor productivities and intensity. In the absence of measurement error, efficient allocation of resources implies that the marginal product of factors, which is proportional to crop yield of the corresponding factor, is the same among households in any particular year as they would share the same factor price and face no household-specific distortions. A similar argument can be made about equalized labor-land intensity across households in optimal allocation. Figure 1.1 reveals a picture far from non-dispersed productivities and cross factor ratios.

The figure plots the distribution of all three factor productivities and labor-land intensity after logarithm transformation using the kernel estimate of the density of the dispersion. In each graph, controls are included for characteristics about households, land quality, and shocks as well as year dummies. Then further geographical controls are added, including regional fixed effects and then narrowing down to commune, and lastly, household fixed effects to account for regional and commune differences and finally the yearly household's deviation.

The same pattern exists in all four graphs and tells the same story. Take the first panel for example. This graph depicts the distribution of land productivity (in logarithm) where dispersion exists in all four specifications but decreases with further controls. Based on the theoretical framework in section 1.2, these dispersions reveal potential variations in household's shadow prices for land, suggesting the likely market distortions and misallocation of land. The variation appears smaller in more narrow geographical units as price differences become less drastic with a higher level of localization. The variance for logarithm for land productivity after controlling for observable characteristics and year fixed effects is measured at 0.32. Regional differences account for only 4%

of that variation while controlling for smaller geographic units at the commune level accounts for almost 30% of the variation. This suggests that while there are regional differences, differences across communes within regions can explain a significantly greater fraction of the market condition heterogeneity. That still leaves more than 70% to be explained by within-commune variation. The household fixed-effects specification addresses changes over time and contains the least amount of dispersion.

Including commune fixed effects eliminates 30% of the variation in measured crop yield per hectare of land, 35% of the variation in measured revenue per labor day, and 29% of the variation in measured output returns per dollar on intermediate input. Productivity of intermediate inputs also displays the least amount of within-commune-year dispersion, reporting a 0.16 variance compared to 0.23 in land productivity and 0.25 in labor productivity. The household-fixed effects specification has the least dispersion with a greatly reduced variance of 0.109 in productivity of intermediate inputs, less than the 0.116 and 0.15 equivalent variances in land and labor, respectively. This suggests that the households face limited changes in market conditions or distortions over time.

These observations motivate the key assumption that there is little change both within a household over time and little within-commune-year variance in the shadow prices as well as elasticities for intermediate inputs and that most of the observed variance of intermediate input productivity comes from the variation in its measurement error.

The last panel in figure 1.1 plots the distribution of the logarithm of the labor-land ratio in the data. The cross-factor ratio may be advantageous to output to land size or labor days because it involves only physical measures of land and labor inputs. It mitigates the concern about variations from demand-related factors in production measures. Having said that, the cross-factor ratio reveals a variance even higher (0.164) than that of either land or labor productivity. More variation seems to come from the factor input rather than output, potentially suggesting distortion in the land and labor markets.

These observations are suggestive of factor misallocation, but its extent and the magnitude of potential gains generated from reallocation remain unclear. If measurement error exists, it could be a driving force to introduce dispersion into measured productivities and give a skewed picture of allocative efficiency. Grasping a more accurate understanding of the extent of factor misallocation and its impacts on revenues requires identifying variances of measurement errors. In what follows, I will lay out the production function estimation and present my key assumption to identify the measurement error variance, and the test for that assumption. After TFP and the variance of measurement errors are estimated, the final step is to measure misallocation and adjust the estimated TFP and potential gains for measurement errors.

1.4. Estimation

1.4.1. Production function estimation

Cross-sectional estimation of the production function likely generates inconsistent coefficient estimates since TFP affects household input decisions, while fixed effects is a commonly used approach to control for unobservable time-invariant effects. However, it leaves the estimator susceptible to time-varying factors that the model fails to capture. More importantly, while addressing measurement error, the fixed effects estimation can exacerbate measurement error bias. The fixed effects estimates are reported as a reference point only. Another benchmark reported in the essay is the calculation made with the coefficients used in Ayerst, Brandt, and Restuccia (2020) since their study also investigates factor misallocation in Vietnam using the same dataset[7]. For my own analysis in this essay, the parameters β and the expected factor productivity α_J in equation 1 are estimated using Two-stage least squares (2SLS). The observations are at household level with the full sample and one subsample of households who specialize in rice. The estimation includes year fixed effects and the standard errors are clustered at the commune level.

Output values measured in 2010 US dollar are either the reported rice crop output in the subsample or the aggregate revenue across all crops in the survey in the full sample. Land area and labor supplied by household members used in crop production are available in the survey. Similar to the process performed by Ayerst, Brandt, and Restuccia (2020), I calculated the median provincial daily wage from a household's income and time worked in agriculture outside of the household's own farm and used that measure to approximate the amount of hired labor to work on household production. The labor input on the farm is then constructed from the amount of hired labor and self-supplied labor. Intermediate inputs are reported as the aggregate expenditure converted into 2010 US dollars on seeds, saplings, fertilizers, pesticides and herbicides, energy and fuel, maintenance, irrigation and other costs. Covariates used as controls for household observable characteristics are the household head's age, gender, and education level. Land quality controls are aggregated from plot level weighted by their area and include the furthest distance of land plot from household, land value, irrigation fraction, fraction of land with land use rights certificates, and perceived land quality compared with commune average (below, average, or above local average). Drought shocks reported by the households are also controlled and allowed to have heterogenous effects on crops depending on land quality through their interactions.

[7] This essay includes data only from the years 2012, 2014, and 2016, whereas in their paper, earlier rounds of data from 2006, 2008, and 2010 were used as well. However, the coefficients used in Ayerst, Brandt, and Restuccia (2020) were borrowed from the U.S. benchmark rather than estimated.

In order to implement the 2SLS estimator, instrumental variables are required to be correlated with the three factor inputs and satisfy the exclusion restriction. The first set of instruments draws from household shocks regarding avian flu, changes in commodity prices, and whether the household's head is sick. I argue that these shocks may place a restraint on household's cash and labor available that would affect intermediate input and labor use in crop farming, but otherwise have no direct effect on crop output. Another component of instruments, like the instruments used in Gollin and Udry (2021), involves the interaction between the share of weather shocks (drought) with the share of land of different qualities in the commune outside the household, proxied by the sample data from the commune's other households. It captures the varying effects of droughts on different soil quality. The idea is that having controlled for weather shocks' differential impacts based on *household*s' land quality distribution, weather shock effects on the rest of the *commune* bear no direct effect on households' return except through a shadow price change. As such, drought shocks are restricted to only six months out of the survey year so that only droughts that happened early in the crop season would have an impact on input allocation decisions.

Diagnostic tests (under-identification, weak-identification, and over-identification tests) are performed and reported at the bottom of table 1.4 after the results of the second stage estimation. The P-value for the LM-statistic is 0.0047, rejecting the null hypothesis in the under-identification test and suggests that the instruments are indeed relevant and correlated with the endogenous inputs. However, the low Wald F-statistic of 2.916 relative to the critical values at various levels indicates weak instruments. Importantly, the number of instrumental variables allows for the exclusion restriction assumption to be tested. The result indicates that the null hypothesis that all instruments are valid cannot be rejected, lending confidence to the model specification. The first stage is reported in table A1.1 in the appendix.

1.4.2. Identifying variances of measurement errors

The measures observed and recovered in this section are within-household (across year) and within-commune variances. The variation of deviations from the mean is lower than the variation coming from the total sample. Moreover, it allows more room for interpretations[8]. The demean transformation is represented in system (4HH) where ˜ represents the deviations from a household's average. Alternatively, in (4C) ˜ denotes the deviations from a commune's yearly average, with an extra subscript c for communes.

$$y_{ht}^{o} = \epsilon_{Yht} + y_{ht} \quad \rightarrow \quad \tilde{y}_{ht}^{o} = \tilde{\epsilon}_{Yht} + \tilde{y}_{ht} \qquad (4HH)$$

$$j_{ht}^{o} = \epsilon_{Jht} + y_{ht} + \omega_{Jht} + \ln \alpha_J - \ln p_J \qquad \tilde{j}_{ht}^{o} = \tilde{\epsilon}_{Jht} + \tilde{y}_{ht} + \widetilde{\omega}_{Jht}$$

Or expressed in communes' mean deviations:

$$y_{ht}^{o} = \epsilon_{Yht} + y_{ht} \quad \rightarrow \quad \tilde{y}_{hct}^{o} = \tilde{\epsilon}_{Yhct} + \tilde{y}_{hct} \qquad (4C)$$

$$j_{ht}^{o} = \epsilon_{Jht} + y_{ht} + \omega_{Jht} + \ln \alpha_J - \ln p_J \qquad \tilde{j}_{hct}^{o} = \tilde{\epsilon}_{Jhct} + \tilde{y}_{hct} + \widetilde{\omega}_{Jhct}$$

The key assumption is that there is little change within a household over time or alternatively, little within-commune variance in the shadow prices and elasticities for intermediate inputs. Mathematically, it assumes $\widetilde{\omega}_{Mht} \approx 0$ or $\widetilde{\omega}_{Mhct} \approx 0$. Furthermore, system of equations (5) becomes:

$$Var(\tilde{y}_{ht}^{o}) = \sigma_{\epsilon Y}^{2} + \sigma_{Y}^{2}$$

$$Var(\tilde{l}_{ht}^{o}) = \sigma_{\epsilon L}^{2} + \sigma_{Y}^{2} + \sigma_{L}^{2} + 2\sigma_{YL} \qquad (5HH)$$

$$Var(\tilde{n}_{ht}^{o}) = \sigma_{\epsilon N}^{2} + \sigma_{Y}^{2} + \sigma_{N}^{2} + 2\sigma_{YN}$$

$$Var(\widetilde{m}_{ht}^{o}) = \sigma_{\epsilon M}^{2} + \sigma_{Y}^{2}$$

[8] In Gollin and Udry (2021), the variation from shadow prices would be completely eliminated in the process of taking the deviation from the mean of plots under the same management in the same season. The household-level data does not afford the removal of any source of variation. However, the distribution of mean deviations does allow for a further assumption based on intermediate input use that is important for the identification of measurement error variances.

$$Cov(\tilde{y}^o_{ht}, \tilde{l}^o_{ht}) = \sigma^2_Y + \sigma_{YL} \approx Cov(\tilde{l}^o_{ht}, \tilde{m}^o_{ht})$$

$$Cov(\tilde{y}^o_{ht}, \tilde{n}^o_{ht}) = \sigma^2_Y + \sigma_{YN} \approx Cov(\tilde{n}^o_{ht}, \tilde{m}^o_{ht})$$

The last two equations of (5HH) provide a simple test for the assumption. If the assumption holds and most of that dispersion comes from measurement errors rather the heterogeneity in productivity or shadow price, then $Cov(\tilde{y}^o_{ht}, \tilde{l}^o_{ht}) \approx Cov(\tilde{l}^o_{ht}, \tilde{m}^o_{ht})$ and $Cov(\tilde{y}^o_{ht}, \tilde{n}^o_{ht}) \approx Cov(\tilde{n}^o_{ht}, \tilde{m}^o_{ht})$. That is the observed intermediate inputs are expected to covary with the other observed factor demands similarly to how measured output covary with them. On the contrary, if the assumption does not hold, and there are a lot of changes in technology, factor quality, and market conditions within a household over the years, those changes would covary with land in a different way from output.

Observations from figure 1.1 and table 1.2 motivated the assumption. It is now time to turn to more concrete evidence from the data. Table 1.3 reports on variances and covariances of observables. In panel A, the differences $Cov(\tilde{y}^o_{ht}, \tilde{l}^o_{ht}) - Cov(\tilde{l}^o_{ht}, \tilde{m}^o_{ht})$ and $Cov(\tilde{y}^o_{ht}, \tilde{n}^o_{ht}) - Cov(\tilde{n}^o_{ht}, \tilde{m}^o_{ht})$ are very close to 0. There are the most striking differences in the rice sample between the covariances with labor, and even here the differences are only around 0.01. These observations further strengthen the argument for little heterogeneity in market and productivity conditions for intermediate inputs and lend support to this key identification assumption.

Under (5HH), not every variance is identifiable. The solvable variables are variances of measurement error in output and intermediate input $(\sigma^2_{\epsilon Y}, \sigma^2_{\epsilon M})$, variance of true production output and variances of land and labor productivities-distortions $(\sigma^2_Y, \sigma^2_L, \sigma^2_N)$, and the covariance between land and labor productivities-distortions (σ_{LN}). The two variances of measurement error in land and labor $(\sigma^2_{\epsilon L}, \sigma^2_{\epsilon N})$ remain unidentifiable.

This estimate of measurement error variance provides a lower bound for the overall variance of measurement error. Adjusting for within-household variance instead of total sample variance coupled with under-identification ensures that misallocation is not over-adjusted for measurement error.

1.4.3. Misallocation gap and adjustment

Household TFP is estimated by the production residuals. Applying the TFP and coefficient estimates to the optimal household's allocation share $s_h = \frac{\exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)}{\sum_h \exp\left(\frac{\omega_{Yh}}{1-\sum_J \alpha_J}\right)}$ provides the complete estimates of all households' efficient allocation and production as well as potential gains from reallocation before adjustment for measurement error. The estimate for the variance of TFP is adjusted using equation (2) by subtracting from its variance the within-commune or within-household variance of aggregate measurement error $\widehat{Var}(\epsilon_{ht}) = \hat{\sigma}_{\epsilon Y}^2 + \hat{\alpha}_M^2 \hat{\sigma}_{\epsilon M}^2$. The potential gain from reallocating factors can also be adjusted using equation (3).

1.5. Results

1.5.1. Production function

The production estimates are presented in table 1.4. For comparison, column 1 reports the coefficients of land, labor, and intermediate input used in Ayerst, Brandt, and Restuccia (2020) paper, column 2 is the household fixed-effects results, and column 3 reports my preferred 2SLS estimates for the whole sample. The first stage is reported in table A1.1 in the appendix. The land, labor, and intermediate input coefficients are estimated to be 0.42, 0.19, and 0.33. Compared to ABR [9]and fixed-effects specifications, the 2SLS estimates are closer to constant returns to scale.

[9] The paper makes many references to Ayerst, Brandt, and Restuccia (2020) , which will be shortened as ABR for convenience.

Land accounts for a relatively higher share of revenue using 2SLS, whereas the share of intermediate inputs is comparable across the two models.

Land value and self-reported relative land quality prove to be good indicators of land quality, which show statistically significant positive correlations with revenue even if modest in magnitude. Particularly, doubling the value of land leads to a 1.4% increase in revenue, which translates into a $9.4 increase in median crop value. Drought weather shocks negatively affect crop revenues. The coefficients of demographic controls are not surprising but not significant. Households with older or higher-educated heads receive higher crop revenue, whereas female heads tend to have lower yields.

1.5.2. TFP dispersion and its relationship with output and inputs

Production estimation provides estimates for the unadjusted logarithm of TFPs using production residuals. In a distortion-free environment, TFP would have a strong positive correlation with all inputs and hence revenue as well. The more friction is introduced into factor and output markets, the weaker this relationship becomes; and in extreme cases, the direction could turn negative in instances where distortions are severe, which prevent more productive households from acquiring more inputs. Examining how TFP correlates with factor inputs and revenue can provide insight into the allocative efficiency of the factor markets. Insights can be gained even when TFP estimates are confounded by measurement errors as long as true productivities and measurement errors are orthogonal. Figure 1.2 features four graphs depicting the relationships and linear fits between the log TFP estimates (with 1% trimmed at the top and bottom) and inputs and revenue. In all four panels, productivity shows the strongest positive association with revenue with a 0.8 correlation and a 0.6 correlation with intermediate inputs. The relationship between land and labor is still positive, but the correlation is found to be around 0.3 The pattern remains the same in figure 1.3

in rice production. This suggests distortion in the land labor markets. This observation is of little surprise in the presence of, among other things, crop regulations where certain plots of land are restricted to rice growth to ensure food security. However, intermediate input use is found to be relatively efficient.

Table 1.5 reports on the dispersion of productivity, including the variance of log and the 90th-10th percentile log difference after trimming the top and bottom 1% of the residual prediction. ABR's coefficients while controlling for no other observable characteristics understandably result in the most dispersed TFP with a 0.28 log variance and a 1.26 difference between the 90^{th} and 10^{th} percentiles. When demographics, shocks, and input qualities are taken into account, the dispersion drops significantly. With full crops, both FE and 2SLS estimates report similar variances at 0.16 and 0.14 respectively, which is about half that using ABR.

1.5.3. Calibrating measurement error and true output and input variances

I now move on to calculating the measurement error induced variance within the raw TFP estimate. As noted in Section 1.4, the system of equations in my approach is not identified with the assumptions made. I am only able to solve for variance of measurement error in output revenue and intermediate inputs. Table 1.6 documents the solutions to identifiable variables, which also include variance of output productivity, covariance of output and land productivities, covariance of output and labor productivities, and covariance of land and labor productivities. Panel A is similar to a household fixed effects approach where household's data is demeaned by their average across years while panel B uses data points demeaned by year-commune average while.

The majority of dispersion in observed outputs is caused by actual output variation rather than measurement error, and there seems to be more variation in productivity in the South than the North. Within-household true output variance across years is 0.14 in the country, 0.11 in the North and 0.18 in the South in the full sample of households regardless of crops. This holds true for the restricted sample of rice-exclusive farming households as well, to a lesser magnitude, with variances of 0.08, 0.07, and 0.1 in the nation, North and Southern regions, respectively. The decrease in the spread of true output productivity nationwide and both regions suggests a more homogeneous set of rice-exclusive farming households in terms of productivity. The covariances results indicate that land and labor specific productivities are positively correlated and are each negatively correlated with true output. One interpretation is that land and labor use efficiency are complementary rather than supplementary, the better households can make use of land the better they can make use of labor. Furthermore, constraints in one factor market could be positively linked with that in the other factor market as well.

My measurement error variance results indicate that there are rather large dispersions in measurement error of intermediate inputs relative to the measurement error of output. For example, I found 0.12 a within-household variance in intermediate input measurement error in the full sample, 0.15 in the North, and 0.1 in the south. In the rice-crop restricted sample, these variances are even higher, 0.24 in the nation, 0.2 in the North, and 0.27 in the South. Compared to that, the within-household variances in output measurement error are a few times smaller in the modest range of 0.02-0.04. Indeed, considering the diverse items reported in the survey under the category of intermediate input, it is reasonable that there be more variation in measurement error of intermediate inputs. Measurement error in intermediate inputs is weighted by the square of its output elasticity, which is a range of 0.11-0.15. Measurement errors in land and labor are unidentified, leaving the aggregate measurement error variance estimate most likely under-computed. Panel B reports the same findings using within commune-year variances. As expected, I find much less within-household true output variation, since most of it comes from year-to-year shocks and may be price distortion differences, rather than actual households' factor use efficiency. There is barely any variation at all when I look at rice growers only.

1.5.4. Measurement error and potential reallocation gains

Table 1.7 combines the findings of TFP from table 5 and measurement error from table 1.6. After the estimates of the raw log TFP and measurement error variance are obtained, the next investigation is to learn how much measurement error in aggregate confounds the true TFP and calculate output gains from reallocating before and after adjustment for confoundment. In table 1.7, I again report two panels for two sets of estimates, within households across years in panel A, and years within commune-year in panel B.

The unadjusted log TFP estimates are the residuals from the production function, its variance report is repeated from table 1.5. The adjusted variance of log TFP is simply the difference between unadjusted log TFP and measurement error variances. Aggregate measurement error variance is the weighted sum of variances of output and intermediate input measurement errors.

Between the within-household and within-commune variances, measurement error variance is calculated to be between 0.04 and 0.06, accounting for 30-45% of TFP variance. Based on these findings, the country could more than double its crop output by reallocating. While this is not uncommon, this magnitude is more often seen in African agriculture such as a 186% increase in Uganda (Aragon, Restuccia, and Rud, 2021), and 259% in Malawi (Chen, Restuccia, and Santaeulàlia-Llopis 2023), and is far from what is seen for Vietnam's neighbor China with an increase of 53% (Adamopoulos et al., 2022). This adds to the evidence that allocative inefficiency may be overstated by raw TFP estimates.

Using ABR's coefficients and calculated output gains for benchmark and found only a 79% potential increase. This difference is neither surprising nor does it speak to which estimate is closer to the truth. Rather the estimates offer a possible range and point out room for improvement. The set of coefficients estimated in this essay using 2SLS translates into a distribution of TFP, 30-45% of whose variation is confounded by measurement errors, while ABR's coefficients estimate that only15-22% of TFP variation is measurement error induced. The adjustment for measurement error reduces all estimates of potential gains from reallocation. After adjustment, the 2SLS method estimates a much more modest 45-72% allocative efficiency gain, and the ABR ranges from 58%-65% potential gain. Depending on the set of coefficients and measurement error adjustment method/assumption, I find that the gains through factor reallocation are overestimated up to threefold in the national sample of households.

Both methods of adjustment using within-household variances and within-commune variances are found to perform consistently and yield similar results.

1.5.5. Heterogeneity analysis results

The TFP variance from rice production is drastically smaller, suggesting the rice specialists constitute a more homogeneous selection of farmers. Measurement error from the rice production is consistently responsible for a higher fraction of the TFP than from the aggregate crop production, with one exception (the within commune-year variances). This potentially suggests that the aggregation method does seem to have power against the impact of measurement error.

The historical and natural differences between the two regions translate into large differences in agricultural practices and efficiency. Both regions specialize in rice crops but the South is a larger rice producer by far. The comparative analysis between the North and the South examines their respective efficiency.

Regionally, the raw TFP has higher dispersion in the Southern sample than the Northern sample in the aggregate production. The unadjusted estimate of potential gains from reallocation is large in both regions for aggregate crops, more than doubling productions (131% and 123% in North and South, respectively). Unsurprisingly, efficiency is higher in rice production with smaller predicted gains from allocation; however, the potential gains are still estimated to be more than half of current production (66%

and 59% in North and South, respectively). In both productions, the unadjusted estimates find the South to be more efficient in resource allocation.

After adjusting for measurement error, the allocative gain estimate is drastically reduced in both regions and both crop productions. The potential gains from resource reallocation within each region are estimated in the 32%-51% range for all crops. However, a different pattern emerges. The North is subject to such large variation in measurement error that after adjustment, the region seems to gain a slight edge in the overall production, and the edge is higher using within-commune variance estimates. This finding is surprising and calls for a closer examination of the agricultural practices and conditions in the two regions. On the other hand, not only is rice crop confirmed to be the comparative advantage of the South in rice production, but the adjusted measures find that the South is greatly efficient at allocating resources in rice growing that reallocation can only increase production by 15% of production.

1.5.6. Crop-level analysis

In this section, I replicate Gollin and Udry's (2021) method and assumptions for identification. The results are included in the appendix. The structure of household survey data used in this study is not suited for a highly demanding plot-level data approach. The lack of seasonal and plot-level input information prevents the analysis to be performed at this level. Thus I divided crops into three categories: rice, maize, and all other crops, as output and input observations are available in the available dataset for these categories. The drawback of this procedure is that additional dispersion is being added rather than removed from the production residuals, stemming from crop technology heterogeneity, inherent factor quality requirement differences, varying seasons and timing within a year, etc. It is for these reasons that the results of this exercise should not be taken literally.

Table A1.3 summarizes the observed variances and covariances of observed outputs and inputs deviation from the household-year average. A quick look reveals nonnegligible differences between how output and intermediate inputs covary with land and labor, suggesting non-trivial variation in true intermediate input use across households' different crops, which invalidates my assumption on intermediate inputs.

Another concern of this approach is that since there can be multiple seasons growing different crops within any year, the sum of land use often far exceeds the total amount of land available to farmers. This creates a challenge in computing the output loss from misallocation since any plot of land can be counted multiple times if it is used repeatedly and spread across different categories of crops.

For the sake of the exercise, consider each year a single point in time, and the amount of land available is the sum of the area every time it is used. Every household has its own ratio of self-reported land use for crops to landholdings, where the median of this distribution is 2, with a 2.07 mean and 3.6 at the 90^{th} percentile. The total household's self-reported land use at the crop level averages 1.6 times their landholding. Gains from reallocating are calculated and reported using these four values. The higher land availability is assumed, the larger the output gap.

Table A1.4 shows that within the 0.39 variance of unadjusted log TFP from the set of IV coefficient estimates, measurement error causes 63.2% of that dispersion. Even in the most conservative estimate, the raw gain from reallocation comes at an unrealistic 548% and becomes an outlandish 815% in the more liberal assumption. After removing the 63% share of measurement error from log TFP variance, the gap is drastically adjusted down to anywhere from less than 1% to a seemingly reasonable 42%. It also means that the output gap is being exaggerated from 19 times to over 900 times. These estimates should not be taken literally for the many reasons discussed above, but they showcase once again how measurement error can lead to a severe overstatement of the misallocation gap.

1.6. Conclusion

Resource misallocation in the agricultural sector is the main cause of productivity and income differences across nations and regions and there are a large number of existing studies exploring the determinants and consequences of resource misallocation in the agricultural sector (Caselli 2005; Restuccia, Yang, and Zhu 2008; Restuccia and Rogerson 2017). However, few have seriously taken the measurement errors of output and inputs into account in their estimations of the productivity costs of resource misallocation even though measurement errors are known to be prevalent in developing countries. As a result, the costs of resource misallocation are likely to be overestimated because part of the estimated effects are indeed associated with measurement errors. Gollin and Udry (2021) is the first study developing approaches to decompose the causes of productivity gaps in the agricultural sector into a part that is truly due to resource misallocation and a part associated with measurement error. They found that failing to account for misallocation would substantially overestimate the productivity costs of misallocation.

Despite the difference in data structure and different empirical strategies, this study largely confirms the findings of recent studies (Gollin and Udry 2021, Aragon, Restuccia, and Rud 2021) that measurement error plays a substantial role in the estimation of productivity effects of resource allocation. More specifically, failing to account for the measurement errors associated with output and inputs would lead to a large overestimation of the negative effects of resource misallocation on productivity. The essay finds that measurement error accounts for 30%-45% of the variation in TFP, and failing to address the measurement error leads to doubling and tripling the estimated gains from reallocation. It suggests that after adjustment for measurement errors, potential gains from reallocation are drastically lowered and range from 46% to 71%. Further analysis shows that misallocation varies across regions where the property rights and market conditions are historically quite different. These findings further highlight the importance of taking measurement error into account for future studies quantifying misallocation and productivity inefficiency.

The essay is limited in several ways. The assumption of classical measurement error may be too restrictive given the emerging evidence of non-classical measurement error (Desiere and Jolliffe 2018; Abay et al. 2019; Abay 2020; Ayalew et al. 2024)[10]. The method in this paper only allows for the identification of within-household and within-commune variances rather than the overall variances; additionally, measurement errors in land and labor are unidentified, leading to an underestimation of measurement error impact. The variances of two factor inputs, land and labor, cannot be identified, and the recovered measurement error variances are within-household and within-commune, rather than total sample variance. Nevertheless, the conservative estimates significantly lower the risk of overadjustment.

This essay makes several contributions. It is an entry to the thin literature on the impact of measurement error on misallocation and the first to do so for Vietnam. The method developed in the paper is adaptable to household survey data in similar contexts, saving costs on high-quality data. The essay proposes a key assumption on intermediate input use that allows for testing and identification of the variance of measurement error. The addition of intermediate inputs is essential in this context since Vietnam is within the region with the highest consumption. The method contributes by not only adding intermediate inputs into the production function but also using it as a key identification instrument. The use of household-level data to address measurement error is cost-effective and may even be advantageous over plot-level data if measurement error can be aggregated out (Aragon, Restuccia, and Rud 2021). Finally, the regional differences in allocative

[10] Like Bils, Klenow, and Ruane (2021) and Gollin and Udry (2021), measurement error is necessarily assumed to be classical for identification in the paper. Furthermore, studies on non-classical measurement error find that they tend to correlate with extreme farm sizes. A highly homogeneous society, farm size and farming practices are less likely to be dichotomous in Vietnam, lending some support to the classical measurement error assumption.

efficiency between the North and the South are highlighted, drawing attention to the differential needs of the two agricultural economies.

Given the significant impact of measurement error on the estimate of gains from reallocation, researchers and policymakers should exercise caution when interpreting measures of potential gains from reallocation, especially when comparing them with the often large costs of reallocation.

FIGURES

Figure 1.1 Distribution of factor productivities and labor/land ratio, controlling for observable characteristics with top and bottom 1% trimmed

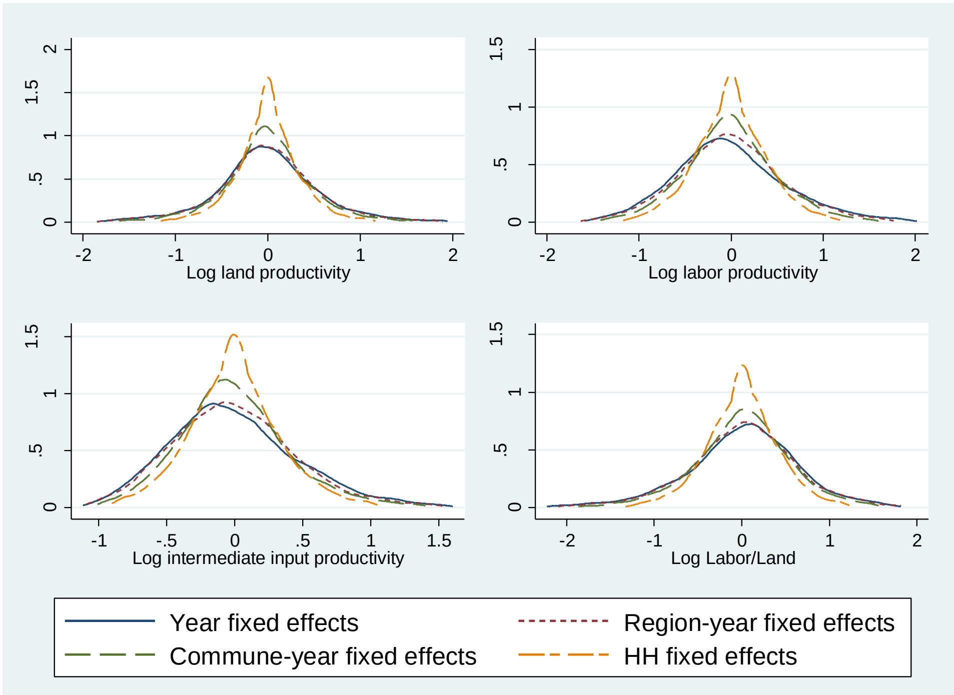


Figure 1.2 Output, Input and unadjusted log TFP in all crop sample

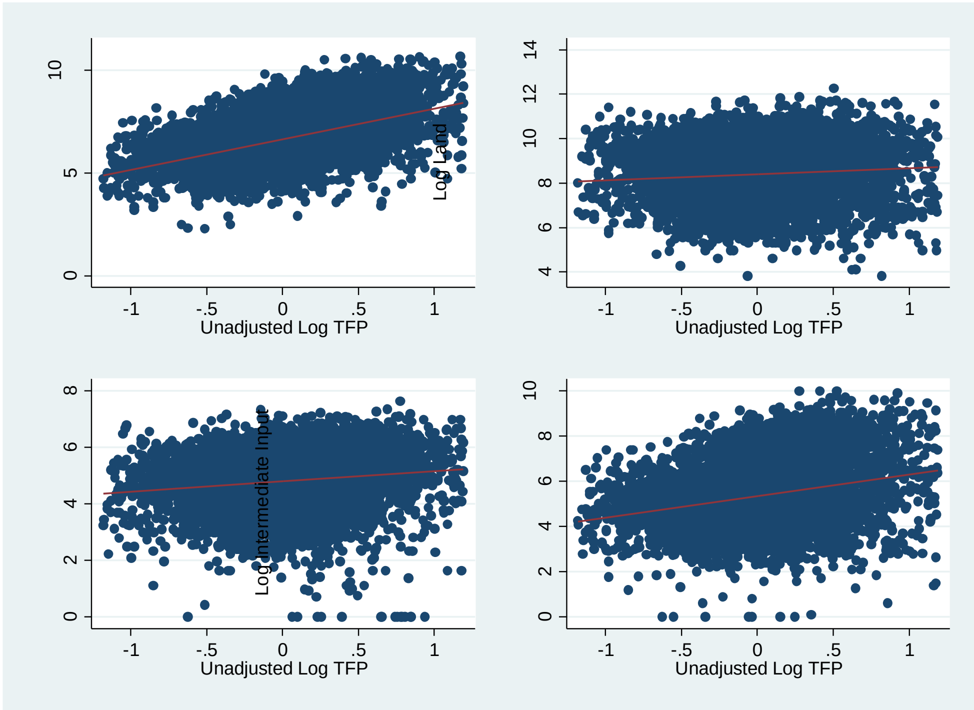
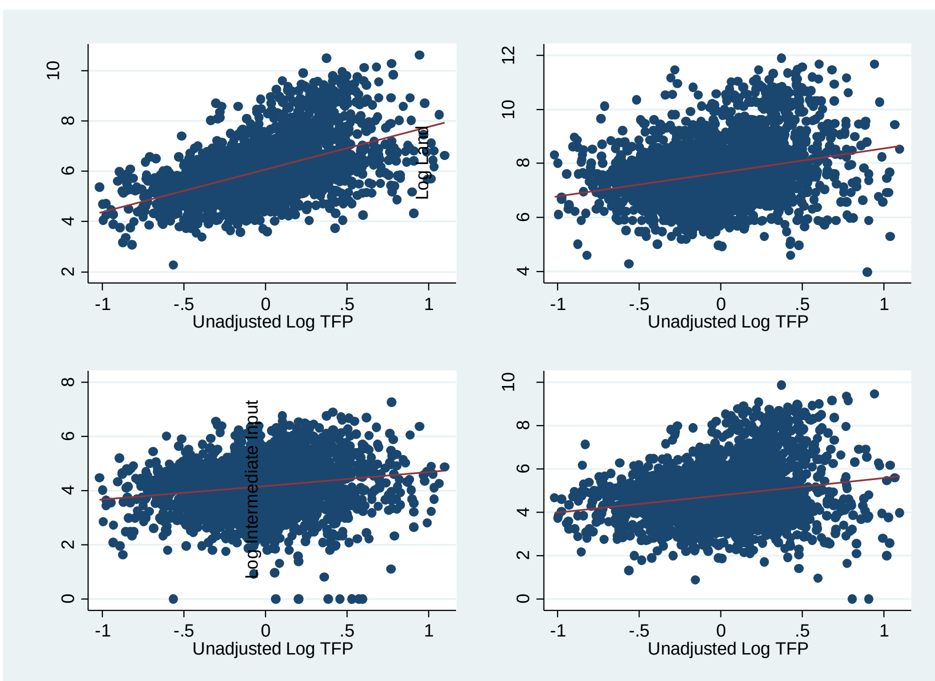


Figure 1.3 Output, Input and unadjusted log TFP in sole rice farming households



TABLES

Table 1.1 Sample

Sample size	Pooled	2012	2014	2016
Household-years	8264	2887	2861	2516
Households	3140	2887	2861	2516
Communes	492	481	464	456
Districts	138	138	136	135
Provinces	12	12	12	12
Mean Crops/Household-years	2.11	2.12	2.14	2.06

Source: VARHS by UNU WIDER. Statistics are calculated by the author.

Table 1.2 Output, inputs, and factor productivities

A. All crops				*B. Rice only HH*			
Output ($)		Labor/Land (Days/ha)		Output ($)		Labor/Land (Days/ha)	
Mean	1672.36	Mean	385.13	Mean	1003.31	Mean	386.29
Median	672.09	Median	271.07	Median	374.053	Median	337.06
STD	3252.15	STD	573.63	STD	2441.94	STD	306.17
Land Area (ha)		Land productivity ($/ha)		Land Area (ha)		Land productivity ($/ha)	
Mean	0.8915	Mean	2466.45	Mean	0.4796	Mean	2270.04
Median	0.47	Median	1981.6	Median	0.19	Median	2216.7
STD	1.5238	STD	3748.18	STD	1.0458	STD	1693.19
Labor (days)		Labor productivity ($/Day)		Labor (days)		Labor productivity ($/Day)	
Mean	175.69	Mean	9.35	Mean	131.78	Mean	10.15
Median	130	Median	5.67	Median	84	Median	6
STD	162.93	STD	18.71	STD	149	STD	19.69
Intermediate input ($)		Interm. input productivity		Intermediate input ($)		Interm. input productivity	
Mean	581.93	Mean	4.62	Mean	343.83	Mean	4.62
Median	173.94	Median	3.37	Median	102.81	Median	3.38
STD	1327.24	STD	6.46	STD	966.73	STD	6.47

Source: VARHS by UNU WIDER. Statistics are calculated by the author.

Table 1.3 Variances and Covariances of observed outputs and inputs

	All crops			Rice only HH		
	Nation	North Region	Southern Region	Nation	North Region	Southern Region
A. Within HH across year						
Var(Output)	0.1745	0.1459	0.2215	0.1113	0.0901	0.1321
Var(Land)	0.1107	0.1167	0.1008	0.0921	0.0886	0.0957
Var(Labor)	0.2257	0.2162	0.2413	0.2286	0.1979	0.2588
Var(Interm. Input)	0.2636	0.2553	0.2774	0.3186	0.2695	0.3667
Cov(Output, Land)	0.0506	0.0511	0.0499	0.0573	0.0538	0.0606
Cov(Output, Labor)	0.0851	0.0794	0.0946	0.0577	0.0552	0.0603
Cov(Output, Interm. Input)	0.1389	0.112	0.1833	0.0823	0.0662	0.098
Cov(Land, Labor)	0.0406	0.0462	0.0315	0.0459	0.0536	0.0383
Cov(Land, Interm. Input)	0.0463	0.0411	0.0549	0.0536	0.0514	0.0559
Cov(Labor, Interm. Input)	0.0831	0.072	0.1015	0.0696	0.0604	0.0786
Cov(Output, Land) - Cov(Land, Interm. Input)	-0.0043	-0.01	0.005	-0.0037	-0.0024	-0.0047
Cov(Output, Labor) - Cov(Labor, Interm. Input)	-0.002	-0.0074	0.0069	0.0119	0.0052	0.0183
B. Within commune-year						
Var(Output)	0.5663	0.4063	0.8286	0.4125	0.293	0.5383
Var(Land)	0.5486	0.4563	0.6999	0.4382	0.3222	0.5604
Var(Labor)	0.4533	0.3872	0.5617	0.4569	0.399	0.5181
Var(Interm. Input)	0.6516	0.4728	0.9444	0.6737	0.4961	0.8606
Cov(Output, Land)	0.3925	0.2973	0.5486	0.3712	0.2638	0.4843
Cov(Output, Labor)	0.3584	0.28	0.4868	0.2986	0.2313	0.3694
Cov(Output, Interm. Input)	0.5151	0.3468	0.7908	0.3896	0.275	0.5102
Cov(Land, Labor)	0.3173	0.2592	0.4125	0.304	0.2404	0.371
Cov(Land, Interm. Input)	0.3898	0.2878	0.5568	0.3721	0.2634	0.4865
Cov(Labor, Interm. Input)	0.3589	0.2735	0.4988	0.3121	0.2428	0.3852
Cov(Output, Land) - Cov(Land, Interm. Input)	-0.0027	-0.0095	0.0082	0.0009	-0.0004	0.0022
Cov(Output, Labor) - Cov(Labor, Interm. Input)	0.0005	-0.0065	0.012	0.0135	0.0115	0.0158

Table 1.4 Production Function

	ABR	FE	2SLS
Land in m2 (Log)	0.2	0.225***	0.420***
		(0.0213)	(0.114)
Labor (Log)	0.3	0.178***	0.190*
		(0.0162)	(0.107)
Intermediate input (Log)	0.3	0.429***	0.328***
		(0.0185)	(0.113)
Year = 2014		-0.0169	-0.0121
		(0.0167)	(0.0407)
Year = 2016		0.0335	-0.0196
		(0.0209)	(0.0366)
Head's age		0.00171	0.000751
		(0.00146)	(0.000708)
Female head		-0.0184	-0.0498
		(0.0447)	(0.0496)
Head's education		0.0168	0.0446*
		(0.0125)	(0.0248)
Furthest distance from land (Log)		0.000185	-0.0202
		(0.00832)	(0.0237)
Land value (Log)		0.000385	0.0213**
		(0.00240)	(0.00916)
Irrigated fraction		0.0123	0.473***
		(0.0360)	(0.172)
LURC fraction		0.0199	0.0680
		(0.0255)	(0.0422)
Crop restricted fraction		0.0444**	-0.160
		(0.0218)	(0.120)
Below average land quality		-0.0492**	-0.0734**
		(0.0207)	(0.0297)
Average land quality		-0.0188	0.0189
		(0.0298)	(0.0374)
Above average land quality		0.0190	0.0952***
		(0.0290)	(0.0362)
Missing land quality		-0.000401	-0.115
		(0.196)	(0.127)
Drought		-0.0858	-0.150
		(0.103)	(0.115)
Drought x Below average land quality		-0.0128	0.0495
		(0.0689)	(0.0663)

Table 1.4 (cont'd)

Drought x Average land quality	0.0733	0.101
	(0.0935)	(0.109)
Drought x Above average land quality	0.0260	0.104
	(0.145)	(0.195)
Constant	1.473***	0.246
	(0.164)	(0.983)
Observations	8,233	8,032
R-squared	0.500	0.831
Number of hhid	3,146	
Under-identification test		
	Kleibergen-Paap rk LM-statistic	15.01
	$\chi^2(4)$ P-value	0.0047
Weak identification test		
	Cragg-Donald Wald F-statistic	3.527
	Kleibergen-Paap rk Wald F-statistic	2.916
Stock-Yogo weak ID test critical values		
	5% maximal IV relative bias	12.20
	10% maximal IV relative bias	7.77
	20% maximal IV relative bias	5.35
	30% maximal IV relative bias	4.4
Over-identification test of all instruments		
	Hansen J-statistic	6.182
	$\chi^2(3)$ P-value	0.1031

Note: Standard errors are clustered at the commune level. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.5 Unadjusted TFP

	All crops		Rice crop	
	Variance of log	90-10 log difference	Variance of log	90-10 log difference
ABR	0.28	1.26	-	-
FE	0.16	1.04	-	-
IV				
Nation	0.14	0.94	0.10	0.79
Northern Region	0.11	0.84	0.10	0.75
Southern Region	0.18	1.03	0.11	0.84

Table 1.6 Estimates of variances and covariances of measurement error and productivity

	All crops			Rice only HH		
	Nation	North Region	Southern Region	Nation	North Region	Southern Region
A. Within HH across years						
Output ME	0.03	0.04	0.04	0.03	0.02	0.03
Interm. Input ME	0.12	0.15	0.1	0.24	0.2	0.27
True output	0.14	0.11	0.18	0.08	0.07	0.1
Total and land productivities	-0.09	-0.06	-0.13	-0.02	-0.02	-0.04
Total and labor productivities	-0.05	-0.03	-0.09	-0.02	-0.01	-0.04
Land and labor productivities	0.04	0.03	0.07	0.01	0.01	0.02
B. Within commune-year						
Output ME	0.05	0.06	0.04	0.02	0.01	0.03
Interm. Input ME	0.13	0.12	0.15	0.28	0.22	0.35
True output	0.52	0.35	0.79	0.39	0.28	0.51
Total and land productivities	-0.13	-0.05	-0.24	-0.02	-0.02	-0.03
Total and labor productivities	-0.16	-0.05	-0.2	-0.09	-0.05	-0.14
Land and labor productivities	0.09	0.05	0.16	0.02	0.03	0.03

Table 1.7 Log TFP adjustment and allocative gains

	All crops			Rice only HH		
	Nation	North Region	Southern Region	Nation	North Region	Southern Region
A. Within HH across years						
Unadjusted log TFP	0.14	0.11	0.18	0.10	0.10	0.11
Adjusted log TFP	0.10	0.06	0.13	0.04	0.04	0.04
ME	0.04	0.06	0.05	0.07	0.05	0.07
Share of ME in unadj log TFP	30.1%	49.3%	28.7%	64.5%	53.1%	63.6%
Unadjusted gains	138.9%	131.1%	122.9%	63.5%	66.4%	58.5%
Adjusted gains	71.6%	49.9%	50.7%	26.8%	37.3%	20.8%
Unadj.gains / Adj. gains	1.94	2.63	2.42	2.37	1.78	2.81
B. Within commune-year						
Unadjusted log TFP	0.14	0.11	0.18	0.10	0.10	0.11
Adjusted log TFP	0.08	0.04	0.12	0.04	0.05	0.03
ME	0.06	0.07	0.06	0.06	0.04	0.08
Share of ME in unadj log TFP	44.8%	64.1%	31.8%	60.8%	45.8%	74.5%
Unadjusted gains	138.9%	131.1%	122.9%	63.5%	66.4%	58.5%
Adjusted gains	45.9%	31.7%	44.6%	28.7%	40.9%	15.3%
Unadj.gains / Adj. gains	3.03	4.13	2.76	2.21	1.62	3.81

	ABR	FE		ABR	FE
A. Within HH across year			*B. Within commune-year*		
Unadjusted log TFP	0.28	0.16	Unadjusted log TFP	0.28	0.16
Adjusted log TFP	0.24	0.11	Adjusted log TFP	0.22	0.09
ME	0.04	0.05	ME	0.06	0.07
Share of ME in log TFP	14.5%	31.8%	Share of ME in log TFP	21.9%	45.1%
Unadjusted gains	78.6%	55.7%	Unadjusted gains	78.6%	55.7%
Adjusted gains	64.6%	37.8%	Adjusted gains	57.8%	30.9%
Unadj.gains / Adj. gains	1.22	1.47	Unadj.gains / Adj. gains	1.36	1.80

BIBLIOGRAPHY

Abay, Kibrom A. "Measurement errors in agricultural data and their implications on marginal returns to modern agricultural inputs." *Agricultural Economics* 51, no. 3 (2020): 323-341.

Abay, Kibrom A., Leah EM Bevis, and Christopher B. Barrett. "Measurement Error Mechanisms Matter: Agricultural intensification with farmer misperceptions and misreporting." *American Journal of Agricultural Economics* 103, no. 2 (2021): 498-522.

Abay, Kibrom A., Gashaw T. Abate, Christopher B. Barrett, and Tanguy Bernard. "Correlated non-classical measurement errors,'Second best'policy inference, and the inverse size-productivity relationship in agriculture." *Journal of Development Economics* 139 (2019): 171-184.

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia. "Misallocation, selection, and productivity: A quantitative analysis with panel data from China." *Econometrica* 90, no. 3 (2022): 1261-1282.

Adamopoulos, Tasso, and Diego Restuccia. "The size distribution of farms and international productivity differences." *American Economic Review* 104, no. 6 (2014): 1667-1697.

Aragón, Fernando M., Diego Restuccia, and Juan Pablo Rud. *Heterogeneity, measurement error, and misallocation in african agriculture: A comment*. University of Toronto, Department of Economics, 2021.

Ayalew, Hailemariam, Jordan Chamberlin, Carol Newman, Kibrom A. Abay, Frederic Kosmowski, and Tesfaye Sida. "Revisiting the size–productivity relationship with imperfect measures of production and plot size." *American Journal of Agricultural Economics* 106, no. 2 (2024): 595-619.

Ayerst, Stephen, Loren Brandt, and Diego Restuccia. "Market constraints, misallocation, and productivity in Vietnam agriculture." *Food Policy* 94 (2020): 101840.

Banerjee, Abhijit V., and Benjamin Moll. "Why does misallocation persist?." *American Economic Journal: Macroeconomics* 2, no. 1 (2010): 189-206.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. "Cross-country differences in productivity: The role of allocation and selection." *American economic review* 103, no. 1 (2013): 305-334.

Bento, Pedro, and Diego Restuccia. "Misallocation, establishment size, and productivity." *American Economic Journal: Macroeconomics* 9, no. 3 (2017): 267-303.

Bogaert, P., J. Delincé, and S. Kay. 2005. "Assessing the Error of Polygonal Area Measurements: A General Formulation with Applications to Agriculture." *Measuring Science and Technology* 16, no. 5:1170–78.

Carletto, Calogero, Sara Savastano, and Alberto Zezza. "Fact or artifact: The impact of measurement errors on the farm size–productivity relationship." *Journal of Development Economics* 103 (2013): 254-261.

Carletto, Calogero, Sydney Gourlay, and Paul Winters. "From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis." *Journal of African Economies* 24, no. 5 (2015): 593-628.

Caselli, Francesco. "Accounting for cross-country income differences." *Handbook of economic growth* 1 (2005): 679-741.

Chen, Chaoran, Diego Restuccia, and Raül Santaeulàlia-Llopis. "The effects of land markets on resource allocation and agricultural productivity." *Review of Economic Dynamics* 45 (2022): 41-54.

Chen, Chaoran, Diego Restuccia, and Raül Santaeulàlia-Llopis. "Land misallocation and productivity." *American Economic Journal: Macroeconomics* 15, no. 2 (2023): 441-465.

Cohen, Alex. "Estimating farm production parameters with measurement error in land area." *Economic Development and Cultural Change* 68, no. 1 (2019): 305-334.

Desiere, Sam, and Dean Jolliffe. "Land productivity and plot size: Is measurement error driving the inverse relationship?." *Journal of Development Economics* 130 (2018): 84-98.

Gollin, Douglas, David Lagakos, and Michael E. Waugh. "The agricultural productivity gap." *The Quarterly Journal of Economics* 129, no. 2 (2014): 939-993.

Gollin, Douglas, Stephen Parente, and Richard Rogerson. "The role of agriculture in development." *American economic review* 92, no. 2 (2002): 160-164.

Gollin, Douglas, and Christopher Udry. "Heterogeneity, measurement error, and misallocation: Evidence from African agriculture." *Journal of Political Economy* 129, no. 1 (2021): 1-80.

Hopenhayn, Hugo A. "Firms, misallocation, and aggregate productivity: A review." *Annu. Rev. Econ.* 6, no. 1 (2014): 735-770.

Hsieh, Chang-Tai, and Peter J. Klenow. "Misallocation and manufacturing TFP in China and India." *The Quarterly journal of economics* 124, no. 4 (2009): 1403-1448.

Keita, Naman, and Elisabetta Carfagna. 2009. “Use of Modern Geo-Positioning Devices in Agricultural Censuses and Surveys.” Paper presented at the 57th Session of the International Statistical Institute, Durban, August 16–22.

Kilic, Talip, Alberto Zezza, Calogero Carletto, and Sara Savastano. "Missing (ness) in action: selectivity bias in GPS-based land area measurements." *World Development* 92 (2017): 143-157.

McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo. "Globalization, structural change, and productivity growth, with an update on Africa." *World development* 63 (2014): 11-32.

Restuccia, Diego. "Resource allocation and productivity in agriculture." *Plenary Panel on 'Agricultural Policy in Africa'at the Centre for the Study of African Economics (CSAE). Available online: https://www. economics. utoronto. ca/diegor/research/Restuccia_ResAlloc_Oxford. Pdf (accessed on 11 December 2017)* (2016).

Restuccia, Diego, and Richard Rogerson. "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic dynamics* 11, no. 4 (2008): 707-720.

Restuccia, Diego, and Richard Rogerson. "Misallocation and productivity." *Review of Economic dynamics* 16, no. 1 (2013): 1-10.

Restuccia, Diego, and Richard Rogerson. "The causes and costs of misallocation." *Journal of Economic Perspectives* 31, no. 3 (2017): 151-174.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. "Agriculture and aggregate productivity: A quantitative cross-country analysis." *Journal of monetary economics* 55, no. 2 (2008): 234-250.

APPENDIX

Table A1.1 First stage of all crop production

	Land	Labor	Intermediate Input
Drought x Commune's below average land quality	-0.239	0.254	0.424*
	(0.203)	(0.174)	(0.246)
Drought x Commune's average land quality	0.132	0.0685	0.276*
	(0.122)	(0.106)	(0.161)
Drought x Commune's above average land	0.520	0.694	0.772
	(0.740)	(0.708)	(0.731)
Avian flu	0.0730	0.149***	0.00643
	(0.0467)	(0.0381)	(0.0540)
Change in other commodity prices	-0.0724	0.267***	-0.000883
	(0.0795)	(0.0703)	(0.0900)
Shock from illness or death	0.0268	0.0689	-0.00166
	(0.0770)	(0.0603)	(0.0838)
Head was sick	-0.129***	-0.0792**	-0.0958**
	(0.0411)	(0.0327)	(0.0479)
year = 2014	-0.0258	-0.205***	0.120***
	(0.0174)	(0.0195)	(0.0252)
year = 2016	0.0765***	-0.154***	0.0621*
	(0.0249)	(0.0251)	(0.0326)
Head's age	-0.00727***	-0.00350***	-0.00777***
	(0.00154)	(0.00119)	(0.00185)
Female head	-0.435***	-0.370***	-0.412***
	(0.0507)	(0.0395)	(0.0613)
Head's education	-0.157***	-0.0885***	-0.0137
	(0.0199)	(0.0148)	(0.0238)
Furthest distance from land (Log)	0.173***	0.116***	0.0825***
	(0.0153)	(0.0107)	(0.0162)
Land value (Log)	0.0408***	0.0235***	0.0827***
	(0.00528)	(0.00400)	(0.00638)
Irrigated fraction	-0.390***	0.107***	1.043***
	(0.0483)	(0.0389)	(0.0646)
LURC fraction	-0.0710*	-0.0506*	0.251***
	(0.0389)	(0.0301)	(0.0474)
Crop restricted fraction	-0.743***	-0.304***	-0.778***
	(0.0388)	(0.0303)	(0.0466)

Table A1.1 (cont'd)

Below average land quality	0.0639	-0.0288	-0.103**
	(0.0430)	(0.0371)	(0.0513)
Average land quality	0.211***	0.228***	0.218***
	(0.0605)	(0.0516)	(0.0725)
Above average land quality	0.169**	0.208***	0.300***
	(0.0673)	(0.0558)	(0.0807)
Missing average land quality	-0.140**	-0.0577	-0.178
	(0.0619)	(0.153)	(0.269)
Drought x Household's below average land quality	0.0307	0.175***	0.236***
	(0.0683)	(0.0598)	(0.0872)
Drought x Household's average land quality	0.190***	0.234***	0.282***
	(0.0346)	(0.0299)	(0.0417)
Drought x Household's above average land	-0.306**	-0.0454	0.0232
	(0.136)	(0.113)	(0.175)
Drought x Household's missing average land quality	0.283***	0.338	-0.213
	(0.0940)	(0.230)	(0.373)
Constant	7.896***	4.085***	7.110***
	(0.176)	(0.130)	(0.198)
Observations	6,474	6,474	6,474

Note: Standard errors are clustered at the commune level. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A1.2 Variances and Covariances of observed crop outputs and inputs *within HH-year across crops*

Var(Output)	0.8459
Var(Land)	0.7727
Var(Labor)	0.6304
Var(Interm. Input)	1.3592
Cov(Output, Land)	0.5645
Cov(Output, Labor)	0.5527
Cov(Output, Interm. Input)	0.826
Cov(Land, Labor)	0.3927
Cov(Land, Interm. Input)	0.6607
Cov(Labor, Interm. Input)	0.6583

Table A1.3 Estimates of variances and covariances of measurement error and productivity *within HH-year across crops*

Output ME	0.2
Land ME	0.14
Labor ME	0.03
Interm. Input ME	0.19
True output	0.65
Land productivity	0.16
Labor productivity	0.15
Interm. Input productivity	0.16
Total and land productivities	-0.09
Total and labor productivities	-0.1
Total and Interm. Input productivities	0.18
Land and labor productivities	-0.07
Land and Interm. Input productivities	-0.08
Labor and Interm. Input productivities	-0.07

Table A1.4 Log TFP adjustment and allocative gain estimates at crop level based on land use efficiency

C. Within HH-year (Crop level)	Min (1.6)[11]	50th percentile (2)	Mean (2.07)	90th percentile (3.6)
Unadjusted log TFP		0.39		
Adjusted log TFP		0.14		
ME		0.25		
Share of ME in log TFP		63.2%		
Unadjusted gains	548.1%	614.6%	625.1%	814.8%
Adjusted gains	0.6%	11.0%	12.6%	42.0%
Ratio of unadj. and adj. gains	913.50	56.03	49.69	19.38

[11] Numbers in parentheses represent the ratio of land use to landholdings.

ESSAY 2

THE EFFECTS OF IMPROVED LAND RIGHTS ON LAND MARKETS, LAND USE EFFICIENCY, EMPLOYMENT AND HOUSEHOLD WELFARE: EVIDENCE FROM THE 2013 VIETNAM LAND LAW

2.1. Introduction

Land is a unique and special commodity for it is a finite resource that continues to be a critical input in agricultural production. As such, the functionality of land markets is crucial in driving agricultural productivity and fostering economic growth. An efficient land rental market serves as a valuable mechanism for adjusting misallocation and improving agricultural productivity (Ayerst et al. 2020; Deininger 2003; Deininger and Jin 2005; Jin and Deininger 2009; Chen, Restuccia, and Santaeulàlia-Llopis 2023; Chamberlain and Ricker-Gilbert 2016). Additionally, the rental market is effective in improving household income (Jin and Jayne 2013; Zhang et al. 2018; Xu and Du 2022), combating poverty (Ghebru and Holden 2019; Seewald, Baerthel, and Nguyen 2023; de Janvry 2001), and addressing food insecurity (Muraoka, Jin, and Jayne 2018). Furthermore, the effects of the land rental market extend to labor choices and structural transformation, facilitating youth employment and migration (Ricker-Gilbert and Chamberlain 2018; Kosec et al. 2018), thus fostering overall economic growth and transformation (Deininger 2003; Jin and Deininger 2009).

However, the land rental markets in the developing world exhibit significant imperfections. Drawing from survey data on actual rental and willingness to rent in rural China, Deininger and Jin (2005) found substantial disparities between real and potential participation in land rental, as well as between the actual and desired land rental quantities. Other studies have revealed that farmers are unable to fully adjust their desired amount of land for cultivation through the existing land rental market mechanisms (Deininger, Ali, and Alemu, 2008; Rie, Jin, and Jayne 2018; Ricker-Gilbert and Chamberlain 2018). Market constraints are a ubiquitous reality across most developing countries. Legal restrictions directly impact the market in certain instances. For example, in Ethiopia, regulations stipulate that no more than half of a farm can be rented out (Ghebru and Holden 2019). Similarly, in China, until recently, rural land could not be rented out to non-villagers without permission from village leaders in many villages (Brandt et al. 2002). Moreover, in India, Deininger, Jin, and Nagarajan (2008) found that tenancy reform had adverse effects on land rental.

Tenure insecurity is a prevalent hindrance to the market. In some tenure systems, land rights are not guaranteed. In China and Vietnam, where land is subject to state ownership and only usufruct rights are granted to holders, the land is at risk of being appropriated for reallocation without sufficient compensation. Well-defined land rights are vital to the ability and willingness of farmers to invest and use the valued asset efficiently (Holden et al. 2010; Abdulai et al. 2011; Lawry et al. 2017; Bellemare et al. 2020). In the absence of secure land rights, confidence in renting land is often hampered (Deininger and Jin 2008). As a result, ill-defined tenure perpetuates land misallocation and truncates productivity with inefficient use of land and other resources.

Given the significant impacts of land rentals on economic efficiency and farmers' livelihoods, considerable research effort has been directed toward studying effective ways to improve tenure security. A large number of studies examining impacts of land titling programs in Asia, Africa and the former Soviet Bloc have shown positive effects of land titling on long-term investments, land values, and market participation (Ali and Deininger 2022; Brandt et al. 2017; Crewett and Korf, 2008; Do and Iyer, 2003 and 2008; Deininger, Ali, and Alemu 2011; Deininger and Jin 2005; Gao, Shi, and Fang 2021; Galiani and Ernesto 2010; Barajas 2023; Sitko, Chamberlin, and Hichaambwa 2014; Jacoby, Hanan and Minten 2017; Zhou, Cheng, and Zhang 2022).

However, compared to the growing literature on the effects of land titling programs, there have been relatively few studies on the effects of land certification without titling or the passage of land laws, with few exceptions (Deininger and Jin 2009; Holden, Deininger, and Ghebru 2010; Deininger, Ali, and Yamano 2008; Bellemare et al. 2020). The limited attention given to the economic impacts of the passage of laws and regulations is surprising and unjustified, especially considering the potential cost-effectiveness of implementing land laws and policies relative to the expenses associated with land titling programs.

Ever since the privatization and marketization of Doi Moi in the late 1980s, the land market in Vietnam has been the topic of much research interest due to its history and evolution. Many papers have been published on the land market in Vietnam and a few on the impacts of its different versions of land law. Do and Iyer (2008) find that the additional rights granted to landholders by the 1993 land law increase land share for long-term crops and lead to more labor supply in nonfarm activities. Hansen (2013) discusses the implications of the 2013 land law on land appropriation. Deininger and Jin (2008) identified the existence of significant transaction costs associated with the land rental and land sales market despite the improved land rights in the early stage of the reform.

This essay builds on the rich existing literature in the fields of tenure security and land conditions in Vietnam and contributes to filling a gap at their intersection with the focus on land law. It aims to estimate the causal effect of tenure security in response to the passage of the 2013 land law in Vietnam on land transfers and other economic outcomes. The essay employs a difference-in-differences model to evaluate the impacts of the land law. Particularly, the law extends the land lease term for annual land from 20 years to 50 years, immediately affecting leases that were going to expire in October 2013 under the older version of the law. By comparing the treatment group (annual land) to the control group (perennial land) across time periods before and after the law's enactment, the study adopts a cleaner identification strategy than previous studies for assessing the effects of the law.

This essay adopts a similar approach to Bellemare et al. (2020), utilizing a difference-in-differences framework and employing the same policy instrument. In their study, Bellemare et al. (2020) find that improvement in tenure security on annual plots increases the likelihood of investment in irrigation technology for soil and water conservation. In this essay, I aim to gain insight into the multifaceted implications of tenure security improvements. More specifically, the impact measures in this essay include farmers' participation in the land rental and sales markets, efficiency of the land markets, labor employment, and household welfare.

In summary, my findings indicate that the law indeed increases households' confidence in renting out and selling their land, while having little effect on the demand side of the rental and sales markets. This reinforces the tenure security effect of the passage of the land law on land transfers. The study also utilizes data on annual crop production to examine the law's facilitating power on the land markets' efficiency in allocating land. I find that while the land markets are distributing land from households with lower farming ability to those with higher farming ability, the law also contributes to additional efficiency improvement of the land markets. Additional analysis at the household level reveals that the passage of the land law for annual land tends to shift labor away from own farm work to wage work, especially in local agricultural sectors. My analysis also reveals evidence supporting the association between the passage of the land law and higher household expenditure.

The essay makes two contributions. First, it expands the literature on tenure security by focusing on the impacts of the passage of a land law, an aspect that has been rarely studied despite the obvious cost advantage of passing a land law than implementing a land titling program. The findings of this study are likely to have far-reaching implications for future land policy in Vietnam and other developing countries planning to strengthen land tenure security through alternative options of enacting a land law versus implementing land titling programs. This essay also makes a methodological contribution by employing the difference-in-differences (DID) estimation model to identify the effect of the land law on farmers' rental decisions. A review of the literature reveals that ordinary least squares (OLS) and probit/logit models are the most common methods used in the studies of land rental market participation. The unique nature of the land law in Vietnam that only affects annual land but not the perennial land provides the opportunity to apply rigorous DID methods. Hence, the findings from this rigorous identification strategy are more informative in guiding future policies.

The next section introduces the context of land rights and land markets in Vietnam and provides an explanation of why it is a suitable place to study tenure security. Section 2.3 formulates a conceptual framework for hypotheses that can be tested in the following sections. Section 2.4 presents the data and how it is used in this study. Section 2.5 presents the empirical strategies for the main identification of the impact of the law on land transactions, labor outcomes, and household expenditures on food. The results are discussed in section 2.6. Finally, section 2.7 offers the conclusion to the essay.

2.2. Contextual background

In Vietnam, land technically belongs to the entire people and is managed by the state, which allocates and leases land use rights to individuals. Since the enactment of the first land law in 1988, multiple updates have been introduced, aiming to clarify definitions, and enhance existing land rights. Each revision is intended to fill in the gaps and improve upon previous laws. Notably, between 1988 and 1993, when the second land law was enacted, there was no established market for land use rights.

In 1993, land use rights were defined as the rights to exchange, transfer, inheritance, lease, and mortgage. By the time the next version of the land law came out in 2003, these rights were expanded to include five additional rights, namely subleasing, granting, securing as collateral, employing as a form of capital, and claiming compensation. Under the 1993 law, plots of land allocated to renters were assigned a designated purpose of use, falling into categories such as residual land, water surface land, forestry land, grassland/pasture, and various types of agricultural land. Each agricultural type has its own terms of lease. Agricultural land designated for growing perennial crops had a lease term of 50 years, while annual crop land was only leased for 20 years. Nearing the end of a lease term, the law allowed the usufruct holder to apply to have the lease extended if they wish. To switch to a different purpose of use, the usufruct holders must have filed a legal case to ask for permission from various levels of the People’s Committee, which was a complicated and time-consuming process.

This essay specifically focuses on investigating changes introduced in the 2013 land law. The 2013 land law revised the lease term for annual crop land to 50 years and took automatic effect. This would not only reestablish the rights to use these annual land plots *en masse* at no extra bureaucratic costs but also reaffirm the overall confidence of land use right owners in renting out their land, and furthermore, recreate the market for land title exchange for annual plots whose certificate would have become void without applying for an extension (which, thanks to this new law, no one had to go through). Overall, the 2013 law enhanced tenure security for annual land use rights. Additionally, switching the purpose of use between annual and perennial plots no longer required a legal declaration. Given the huge initial net benefit of the law, the land law has the potential to be a good cost-effective tool in the policy maker's arsenal.

A key focus of this essay is to study how this security manifests in the transactional markets with (selling-buying) and without (renting in-renting out) the exchange of land titles. The fact that the law only affects one type of land (annual arable land) but not the other types of arable land (e.g., perennial arable land) provides a unique setting to study the impact of improved tenure security due to the law on households' decisions regarding land transactions using a difference-in-differences approach.

2.3. Conceptual framework

In this section, I construct a simple household model of land rental to predict the potential effects of the passage of the 2013 Vietnamese Land Law on farmers' decisions regarding land rental. Assume a household is endowed with fixed amounts of land ($\bar{L}$), labor ($\bar{N}$), and an exogenously given level of agricultural ability (α), and consider an agricultural production function, denoted as $f(\alpha;\, L, N)$, with standard properties such as $f_\alpha > 0, f' > 0, f'' < 0$ with respect to all arguments, along with positive second-order cross-partial derivatives.

Given land price (r) and labor wage (w), households are further faced with transaction costs for both renting out land (TC^{Out}) and renting in land (TC^{In}). These transaction costs are associated with the level of land tenure security. Consequently, households make decisions regarding the amount of land to be rented out (L^{Out}), rented in (L^{In}), and labor supply outside of their own farm (N^{Out}, which could be negative if households are net hirers of labor) to maximize profit. Mathematically, a household's profit maximization problem is expressed as follows:

$$\underset{L^{Out},L^{In},N^{In}}{\text{Max}}\, pf\left(\alpha;\ \overline{L}-L^{Out}+L^{In},\overline{N}-N^{Out}\ \right)+\left(r-TC^{Out}\right)L^{Out}-\left(r+TC^{In}\right)L^{In}+ wN^{Out}$$

$$\text{FOC:}\quad pf_L = r - TC^{Out} \quad \text{or} \quad pf_L = r + TC^{In} \tag{1}$$

$$\text{and}\quad pf_N = w$$

The first-order condition equations allow me to derive the comparative statics and propositions on a farmer's decision to participate in land rental in response to a change of transaction costs (in response to the passage of the 2013 Land Law) and a household's level of agricultural ability. Given that the passage of the 2013 Land Law aims to improve tenure security and transferability of annual arable land, the passage of the 2013 Land Law should be associated with the reduction of the transaction costs for renting out and in. The comparative statics analysis based on the first-order conditions derived in the appendix allows me to derive the following propositions which can be tested empirically in the main body of the analysis.

Proposition 1: A reduction in the transaction cost to renting out (TC^{Out}) due to the passage of the 2013 land law reduces net land rental and therefore increases the probability of leasing out.

Proposition 2: Holding everything else constant, the probability of renting in is unaffected by a reduction in renting out transaction cost (TC^{Out}). However, demand for land rental is also expected to increase due to the equilibrium effect as a result of a transaction cost reduction from the 2013 land law.

Proposition 3: The probability of supplying labor out from their own farm is increasing in the transaction cost to leasing out (TC^{Out}) and therefore the passage of the 2013 land law leads to a higher probability of hiring out labor from the farm.

Proposition 4: Household income is decreasing in the transaction cost of leasing out (TC^{Out}), therefore a reduction in the transaction cost of leasing out TC^{Out} due to the passage of 2013 land law improves income.

Proposition 5: The probability of renting in (out) land is strictly increasing (decreasing) in households' agricultural ability α.

In addition to farmers' decisions regarding land rental, I am also interested in investigating the impacts of the passage of the 2013 Land Law on farmers' participation in the land sales market. Although farmers' decisions regarding land sale/purchase are more complex due to considerations of maximizing land profit over its lifespan, it is intuitive to expect that the passage of the 2013 Land Law will also reduce transaction costs associated with land purchase and sale, thus increasing farmers' participation in these activities. Similarly, I anticipate that farmers with higher (lower) farming ability are more likely to lease in (out) land.

2.4. Data

This essay utilizes the Vietnam Access to Resources Household Survey (VARHS) data collected by the United Nations University World Institute for Development Economics Research (UNU WIDER). For this study, I employ data from three rounds (2008, 2010, and 2012) conducted before the land law was enacted in 2013, and two rounds (2014 and 2016) conducted after the law came into effect. The sample of households in the survey was first selected randomly in 2016 to represent the rural population in 12 provinces across all regions of Vietnam, providing an ideal context for comparing between annual and perennial land types. During this period from 2008-2016, the survey revisits the same households to collect data on household characteristics, land activities, members' employment, and livelihood and food expenditure with additional households added in 2008 and 2016.

The study draws from 13,000 household observations over five rounds of data from 2,600 households. The highly balanced household panel data was selected by the author to include only households that appear in all five waves of the survey and furthermore excludes households that are not active operators on either type of land in all five rounds. Household characteristics including head's gender, education, age, household size, and the logarithm of landholdings in 2008 are regressed on the dummy variable indicating whether household appears in all five rounds of data to determine whether there are significant differences between households surviving in all five rounds and those they are missing. Standard errors are clustered at the provincial level due to the sampling design being representative at the provincial level (Abadie et al. 2023). Table 2.2 reports that gender, education, age, and landholdings do not correlate with the survival rate. However, there is more annual land and household size is larger among those included in the balanced panel. The choice to exclude households that are non-operators in all five periods is to keep the focus of the study on active operating households.

The survey provides land transactional data at the plot level, comprising 54,428 plot observations, with 10% categorized as perennial land, and 90% as annual land. The imbalance between the treatment and control group sizes is compensated by a large total sample size, mitigating the concerns over losing statistical power. Over the years, 1,601 households reported owning and/or renting only annual land, 108 households owning and/or renting only perennial land, and 891 households working with both types of land.

The household characteristics associated with annual and perennial land exhibit statistically significant differences. Table 2.2 shows that, between households that work with or own annual land at some point and those that work with or own perennial land, annual plots are associated with smaller-sized households with lower values in durable goods. Additionally, annual plots are also more likely to be headed by female members than perennial plots, and the household heads tend to be younger with lower levels of education. Regarding land and crop management, annual plots are smaller than perennial plots and tend to be closer to home. However, the quality of both types of plots appears to be similar relative to the surrounding land in their communities, with the mean reported at “about average” by households for both types of land.

Table 2.3 provides a summary of the probability of land transfers over the years in both the rental and sales markets, for both the control perennial group and the treatment annual group. The sample shows statistically significant differences between transactional activities in the markets for the two groups of land. The difference is large in the rental market. The probability of annual plots being leased out and in is 5.4 and 2 percentage points higher than perennial plots. The rental market serves an important function in redistributing annual land. On the other hand, it is less clear which type of land is more active in the sales market. Nevertheless, the rate of participation is low with the percentage capped at one digit.

The survey reports employment by individual members in the households, as summarized in table 2.4, providing insights into the type of work households are engaged in. Managing their own farms remains the predominant type of work, as expected, but has shown a rapid decline. Starting from 98% of households in 2008, it steadily declined to 91% of households in 2016. Conversely, wage jobs have become another major source of income, and are increasing in importance, rising from accounting for barely over half of the households in 2008 to 67% in the final round of data. Interestingly, in the period before the new land law, households engaged in off-farm jobs increased by 16 percentage points between 2008 and 2012, only to slightly reduce in the following years. In contrast, farm-related jobs were seen in less than 20% of households in the pre-law period but skyrocketed upwards, with 2016 showing a doubling of the fraction of households compared to eight years before. This raises questions about whether the law may not only affect the land market but also have an impact on labor distribution.

Not only are wage jobs more prominent; but specifically, households are more likely engaged in local work as opposed to working in another commune or migrating outside of the province entirely. By 2016, local jobs accounted for 75% of the households participating in the job market, almost doubling the number of households with members working farther away. Furthermore, while wage labor and wage labor both within the commune and within the province generally trend upwards, migrating outside of the province displays an interesting pattern, seeing a drastic decline between 2008 and 2012 before bouncing back up.

In table 2.4, I also present the frequency of households owning businesses as another source of income, which fluctuates within the range of 21-26% of the sampled households. The number of businesses owned by households is also examined but not reported. No noticeable trend is detected in either statistic. Additionally, food expenditures per capita in the last column are shown to steadily increase with the exception of 2014.

2.5. Econometric approach

2.5.1. Estimation of impacts on land transfer outcomes

An important feature of the 2013 Land Law is that the law only applies to annual crop land but not perennial crop land. This unique feature allows me to employ the difference-in-differences (DID) method to identify the impacts of the new land law on land transfer outcomes. Specifically, the first difference is the difference in transfer outcomes between the annual and perennial types of plots belonging to the same households, and the second difference is the difference between after and before the law is implemented. Mathematically, the DID specification can be expressed as follows:

$$y_{iht} = \beta_0 + \beta_1 TREAT_{iht} + \beta_2 T_t + \beta_3 (TREAT_{iht} \cdot POST_t) + \boldsymbol{\beta_4 Z_{iht}} + \sigma_h + \epsilon_{it} \quad (2)$$

where y_{iht} are transfer outcomes $\{\text{rented out}, \text{rented in}, \text{sold}, \text{purchased}\}$ of plot i in household h in year t, which is valued 1 if yes and 0 otherwise. $TREAT_{iht}$ is a dummy variable for annual-type plots. T_t is the year t fixed effect to control the overall trend in rental activities. $POST_t = 1$ for observations from years 2014 and 2016 after the law is implemented and 0 otherwise. $\boldsymbol{Z_{iht}}$ are plot and household characteristics including distance between plot and home, plot area, plot land quality, household head's gender, age, squared age, education, and household size. σ_h is household h fixed effect, ϵ_{iht} is the random error term with mean zero. The coefficient β_3 on the interaction term $TREAT_i \cdot POST_t$ provides the estimated effect of law treatment of tenure extension on land transfer outcomes. Estimation is done with ordinary least squares while clustering standard errors by communes.

2.5.2. Parallel-trend assumption

The key underlying assumption for the DID regression in equation (1) to be valid is that the trend of land rental outcomes between the annual arable land and the perennial arable land would have been the same in the absence of the land law, therefore any difference in land transfer outcomes post-the law is due to the implementation of the law. Directly testing for this assumption is difficult due to the commonly known missing data problem. Instead, people tend to rely on historical data prior to the policy intervention to test whether the pre-trends between the treatment and control group are parallel. Here I follow this approach to perform the pre-trend test using data from the three pre-law rounds: 2008, 2010, and 2012.

Figure 2.1 provides a visual representation of the trends of land transactions at the means over the years between the control and treatment groups. The trends of leasing annual and perennial appear to be much more similar and parallel in the 2008-2012 periods, even as there was a slight diversion in 2012 right before the land rights would expire and the law was announced. Regarding the trend for selling and purchasing land, while the trends between the annual crop land and the perennial crop land are similar between the period from 2008 and 210, the annual crop land experienced a noticeable drop in both sale and purchase despite the increased activity in perennial land transfers. As a result, the likelihood of land sale or purchase for annual land is 3-4% smaller than the perennial land and the difference is statistically significant. I posit that this sudden drop in land sale and purchase for annual crop land is related to the impending expiration of the land certificates for annual land; furthermore, the anticipated expiration effect is likely to have a more noticeable effect on farmers' sale and purchase decisions than on their leasing decisions because land lease does not involve the exchange of land certificates.

As a diagnostic test for the parallel trend assumption, I estimate equation (1) using the restrictive sample to pre-treatment periods 2008-2012 with 2010 as the base, and then further restrict it to only 2008-2010. The parallel trends assumption holds if the null hypothesis for $\beta_3 = 0$ cannot be rejected. If, however, β_3 is statistically different from 0, then the assumption is invalid. Particularly, if 2010 is statistically indifferent from 2008 but significantly differs from 2012, this would suggest that households did respond to an expectation of land certificate.

2.5.3. <u>Parallel-trend assumption testing</u>

The results of these pre-trend tests are reported in table 2.5 using the three pre-law rounds of data with 2010 as the base for comparison. No statistical effect is detected for 2008 regarding all four transfer activities and no effect in 2012 for renting out and renting in. However, the results show approximately a 2.5 percentage point reduction in both the purchase and sale of annual crop land (relative to perennial land) in 2012 with a 95 percent level of confidence. The negative trend may be explained by the uncertainty of land certificates because they will soon expire. The stark reduction in the sales market sharply points towards the effects of uncertainty and lack of tenure security. There are reasons that this declining trend is not detected in the leasing market. First, it does not involve an exchange of land certificates. The expiration very well could induce a loss of confidence in renting out land and limit the pool of potential renters. However, another reason to circumvent this issue is that leasing can be contracted in short terms and can be terminated by the owners and rentees before their land certificates expire.

Considering there is no reason for renting, selling, and purchasing activities of annual land to be affected between 2008 and 2010, it can be concluded that the parallel trends assumption holds between these two periods. In order to identify the law's impact, I exclude 2012 from the sample. An analysis with the full sample from 2008-2016 is performed and reported in the appendix, tables A2.2-A2.5.

2.5.4. Household ability and allocative effect estimations

The second empirical investigation is to explore whether and to what extent this transfer effect of the new land law facilitates the allocative efficiency of the rental market. This analysis first involves the recovering household's farming ability from the household's annual crop production. The estimation procedure of household fixed effects as the efficiency parameter was modeled and implemented by a number of papers such as Chamberlin and Ricker-Gilbert (2016), Deininger and Jin (2005), and Jin and Jayne (2013). Following this method, I estimate the households' crop production on annual land with fixed effects.

The production function is written as follows:

$$y_{ht} = \alpha_0 + \alpha_1 l_{ht} + \alpha_2 n_{ht} + \alpha_3 m_{ht} + \boldsymbol{\alpha_4 X_{ht}} + \alpha_h + \alpha_t + \epsilon_{ht} \tag{3}$$

y_{ht} is logarithm of annual crop value of household h in year t. Ideally, the estimation applies for all crops that are grown on annual land. Lacking data on crop-specific inputs, the estimation relies on rice and maize crops to recover households' annual-crop farming ability. As two major crops in Vietnam, the production of both crops is separately reported including land, labor and other intermediate inputs. Two specifications are considered to represent annual crop production. One specification estimates only rice, the other includes both rice and maize.

Production is composed of input factors land (l), labor (n), and intermediate input (m) such as expenses on seeds, saplings, fertilizers, pesticides/herbicides, energy, irrigation, maintenance and other costs. Covariate vector $\boldsymbol{X_{ht}}$ represents household demographics, land quality and weather variables. Household demographics include the usual variables (household size, head's gender, educational level, age, squared age). Land controls are aggregated from plot level data, including mean distance of plots to households (log), land value, fraction of irrigation, fraction of land use restricted to certain purposes, fraction of households with usufructs of land. Land quality is self-reported relative to commune at three levels: below, average, or above local average. Weather shocks (flood or drought) are also controlled. Land quality interacting with weather shocks generates further time-varying controls for crop outcomes. α_h and α_t are household and year fixed effects. Coefficients are estimated with fixed effects models. Households' time-invariant farming ability estimates are the household fixed effects $\widehat{\alpha_h}$ in the regression.

The relationship between the enaction of the 2013 land law (i.e., the lease extension) and the allocative efficiency of the land market is estimated with the following equation:

$$y_{ht} = \beta_0 + \beta_1 HA_h + \beta_2 T_t + \beta_3 (HA_h \cdot POST_t) + \boldsymbol{\beta_4 Z_{ht}} + \epsilon_{ht} \quad (4)$$

Outcome y_{ht} land transfer outcomes (lease out, lease in, sold, bought) of household h in year t. HA_h is the time-invariant household fixed component $\widehat{\alpha_h}$ from equation (3). The coefficient β_1 represents the extent to which land transaction is driven by total factor productivity pre-law, β_2 is the time effect. Most relevantly, the coefficient β_3 represents the difference in rental likelihood post-law specifically driven by the household's technical ability. $\beta_3 = 0$ would indicate that the policy has little effect on redistributing land in a manner that would improve market efficiency. On the other hand, $\beta_3 \neq 0$ would suggest that the lease extension has an impact on reallocating land and, in turn, an impact on the annual crop outputs. Specifically, if β_3 and β_1 are of the same sign, then the policy facilitates land to be transferred and used in a more efficient way.

2.5.5. Estimation of impacts on labor and food expenditures

Unlike plot transfers, employment and food expenditure are household outcomes. A modification to the difference-in-differences identification is required.

$$y_{ht} = \beta_0 + \beta_1 TREAT_{ht} + \beta_2 T_t + \beta_3 (TREAT_{ht} \cdot POST_t) + \boldsymbol{\beta_4 Z_{ht}} + \sigma_h + \epsilon_{ht} \qquad (5)$$

where y_{ht} denotes outcomes for households h in year t. $y_{ht} = 1$ is a dummy variable for the household's labor choices. For this study, the employment outcomes of interest are the types of work (wage work, off-farm wage work, on farm wage work), the location of work (whether the household has any member working inside the same commune, inside the same province in another commune, and outside of the province), and whether households own a business. The welfare outcome of choice in this study is food expenditure per capita in logarithm[12].

As before, T_t is the year t fixed effect, $\boldsymbol{Z_{ht}}$ are the same household controls (household size, head's gender, educational level, age, squared age) as well as the household's lagged total annual and perennial landholding, σ_h is household fixed effect, and ϵ_{ht} is the error term for household h in year t with zero mean. The variable $POST_t$ remains the same where $POST_t = 1$ for observations after the law is implemented and 0 otherwise.

[12] While the hypothesis generated by the model predicts increased income, in practice, households' income comes from many sources, it is volatile and subject to numerous factors. Meanwhile food expenditure is more stable, and its variation better reflects changes in households' welfare.

At the household level, the treatment variable is constructed to reflect households' relative annual landholdings to perennial landholdings. In the first (and preferred) set of regressions, the independent variable is defined as the ratio between annual land and total annual and perennial land. The variation in this specification comes from the proportion of land under households' ownership affected by the law. Holding total landholdings constant, the more perennial land is owned, the less affected households are while owning more annual land places households more under the impact of the law. In order to avoid the possibility of reverse causality, the treatment variable is lagged using data reported from the previous round in year $t-2$.

The treatment in the second specification is defined as households' lagged annual landholdings in absolute terms. Since land ownership is indicative of wealth among other things, putting the identification at higher risk of confoundment. The first treatment is the better instrument for the impact of the law while the second specification should be interpreted as a reference and a robustness check only.

2.6. Results

2.6.1. Impacts of the lease extension on land transfers

Table 2.6 reports the main findings of the law's impact. The first panel reports the average effects of the land law, placing observation in the pre-law group (2008-2010) and the post-law group (2014-2016). The second panel includes interactions of treatment with all years in the sample. Examining the law's impact yearly helps gain insights into the differentiation between the immediate effects as opposed to the later more gradual effects of the land law.

I find that overall, the extension of land lease by the law results in a 3 percentage point increase in the probability of a household renting out a plot. The effect is robust and increasing from 2014 to 2016. Confirming proposition 1 in section 2.3, the law likely allows households greater confidence and thus reduces transaction costs to leasing out, resulting in a higher likelihood of leasing out annual land.

On the other hand, the law is found to have no effect on renting in. While proposition 2 in the theoretical model predicts the likelihood of renting in land to be unaffected by a change in the transaction cost, it is expected that in equilibrium, the law would lead to an increase in both rental supply and demand. A plausible explanation for why no effect is observed on rental demand is the sample bias. While the sample is designed to be representative of the rural population, land renters include big farms and commercial entities that are not captured by the sample. If that is the case, the finding is suggestive of additional effects of the law on another party besides the households.

Turning to the sales market, which involves the exchange of land titles, I find a strong average of 5.6 percentage point increase in the post-treatment period in the probability of annual plots being sold. Column 3 in the second panel further reports that most of the increase occurred immediately after the law was passed. There is an 8.4 percentage point increase in the probability of plots being sold in 2014, which tapers down by 2016 to a 2.6 percentage point increase. This suggests that a portion of the 2014 sales came from 2012's backlog of transactions when the market was arrested (table 2.5). Three years after the law was passed, in 2016, annual plot sales remain at a level higher than the pre-law period, suggesting a persistent positive effect of the law on land sales, much similar to the effect on the land rental supply.

The finding reveals no significant change in the probability of annual plots being purchased between the 2008-2010 pre-law period and after the enactment of the law. Given the drastic plunge in buying in 2012, the result may imply that buying confidence has returned to the initial level. However, the law seems to provide no additional effect on land purchasing decisions.

2.6.2. Results on allocative efficiency

The results of the production function regressions are reported in table A2.1 in the appendix for rice only and for an aggregate of rice and maize combined. Table 2.7 reports the marginal effects on the allocative efficiency of the law. There are two specifications, rice, and rice with maize, each using the household ability obtained from the corresponding production functions. I find that the estimated $\widehat{\beta_1}$ is negative for columns 1 and 3, meaning households with relatively lower technical ability are more likely to lease out or sell their land. Conversely, a positive $\widehat{\beta_1}$ in column 2 means higher ability households are more likely to rent land. As hypothesized in proposition 5, this finding indicates that the rental market improves allocative efficiency by transferring land management from lower-ability to higher-ability farming households. However, technical ability does not drive the purchasing decision as seen in column 4. The significant capital investment requirement to purchase land and its ability to be used as an investment most likely contributes to the lack of explanatory power of households' ability on land purchasing decisions.

Having said that, I find little evidence that the land law marginally improves land transactions proportionally with households' farming ability. Estimates of coefficients on the interaction terms between household ability and post-law periods are consistently statistically insignificant. This implies that the effect of the law on land transactions does not depend on households' farming ability. Mathematically, returning to the model, it has been shown in the appendix that the change in land rented out due to a change in leasing out transaction cost $\frac{dL^{Out}}{dTC^{Out}} = \frac{f_{NN}}{pf_{LL}f_{NN} - pf_{LN}f_{NL}}$.

Assuming a Cobb-Doughlas production function of the form $f(\alpha; L, N) = \alpha L^{\gamma} N^{\delta}$, the said effect is not a function of farming ability. Intuitively, the law has a positive effect on the land market participation rate, but the size of its effects does not depend on the household's efficiency parameter.

2.6.3. Impacts of the lease extension on labor and food expenditures

I report household outcomes in tables 2.8 and 2.9. As observed in columns 1-6, households' labor choices are strongly affected by the law depending on their relative holding of annual land. With more tenure security and increased confidence to rent out land, households can afford to free up labor from time spent working on farms to supply it elsewhere. Particularly, I find that the law facilitates households with relatively more annual land to participate in the wage labor market. I also find a large statistically significant increase in work that is agriculturally related and that is located within their commune close to home. Moreover, the law is found to have no impact on the probability that households engage in non-farm jobs, and similarly no effect on households working outside of their commune or farther away, most likely since those types of work involve non-farm labor. The law seems to strictly appeal to the agricultural skill set where farming households seek jobs that they have a comparative advantage in. Column 7 also reports no effects on the likelihood of households having their own business. Nevertheless, households with a relatively higher proportion of annual land are made better off by the law, as it is associated with increased household expenditure per capita in column 8, although the effect does not have a lasting impact. When the treatment variable is replaced with household absolute annual holding in table 2.9, the direction of effects stays consistent throughout although some results become statistically insignificant.

2.7. Conclusion

Land tenure security is vital to improving the efficiency of resource usage, investment, as well as allocation (Holden et al. 2010; Abdulai et al. 2011; Lawry et al. 2017; Bellemare et al. 2020). Yet it is hard to measure the effects of tenure security due to issues of endogeneity. Most of the literature focus has been on evaluating the impacts of land titling programs as an instrument for improving tenure security while neglecting less drastic policies. This essay contributes to the few studies that investigate land law impacts. Methodologically, the DID estimation contributes to the literature where endogeneity abounds, and instruments are hard to come by. It firmly establishes the causality between improved tenure security through the 2013 version of land law in Vietnam and outcomes of land, labor, and welfare.

The essay finds that the 2013 land law increases land supply through both the rental and sales markets while detecting little effect on the demand side. By increasing households' confidence in renting and selling land, the land law lowers the market participation cost. Additionally, the study confirms that the rental market reallocates land from low-ability to high-ability households, improving overall productivity. However, the effects of the law are found to be independent of households' farming ability.

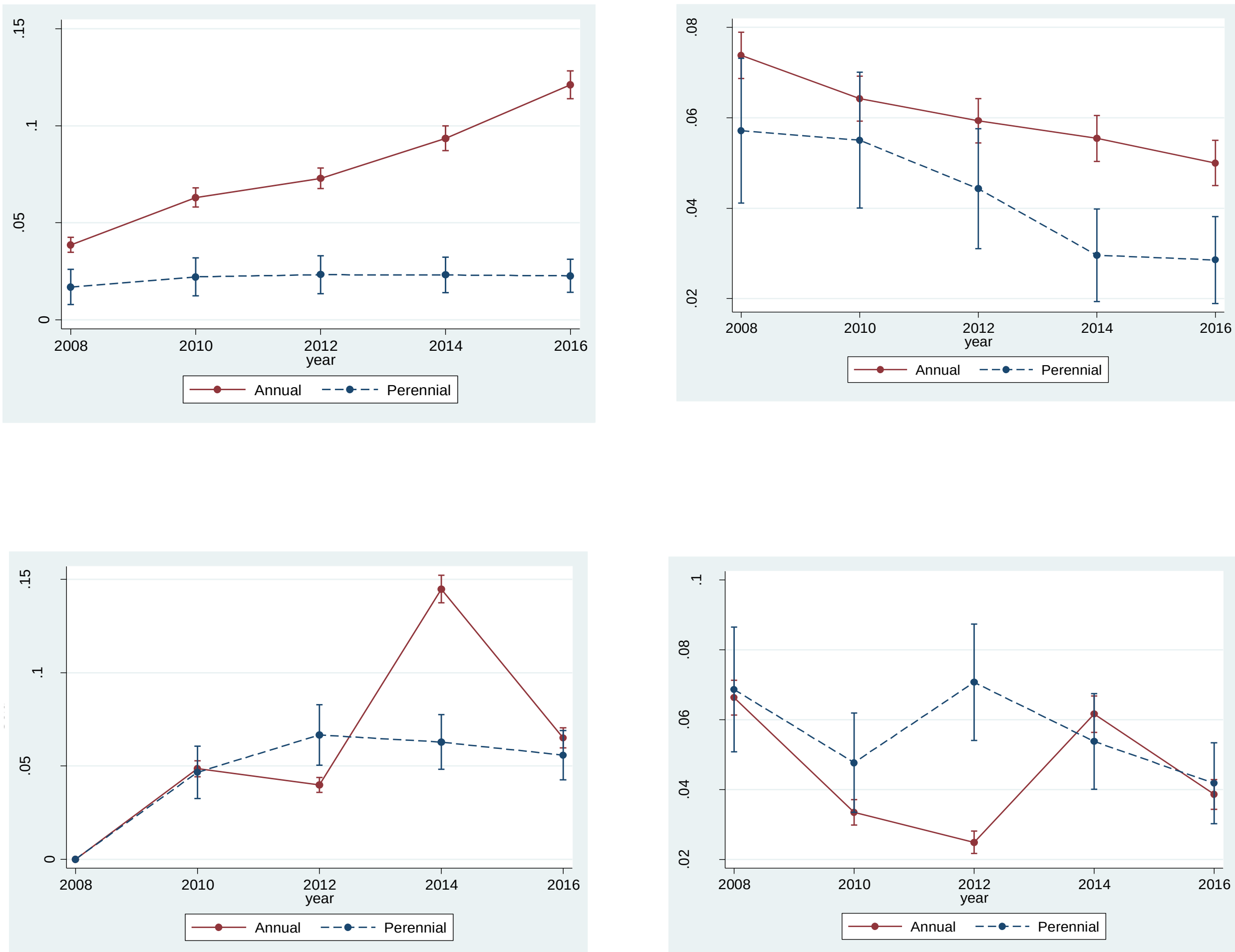
Furthermore, the effects of improved tenure security also extend to the labor market and reshape farming households' supply of labor, affecting their welfare evidenced by the increase in food expenditures. Higher security relaxes constraints on both land and labor markets to free up labor supply on households' own farms in search of other working opportunities for a wage. It found no effects on non-farm employment as well as other income opportunities through businesses owned by households. In short, these findings show that the land law is effective in improving land market participation and household expenditure. However, its effect on increasing labor supply is limited to the agricultural skillset.

Given the findings in this essay, land law could be a cost-effective way of improving tenure security, especially in countries constrained by financial resources. However, to maximize the full potential of the law's impacts, other complementary policies might be needed. Policies that improve opportunities for non-farm jobs and borrowing capital can generate a greater impact on rural economies and should be considered.

There are caveats to this study. One limitation of the analysis is that methodologically, the study, unfortunately, cannot perform true DID analysis on the household outcomes. The study cannot provide a more complete impact evaluation of the law on the demand for land, given the lack of information on the other agricultural producers besides households. Another limitation is that the identified effects of the land law are relatively short-term. Future research should utilize more appropriate data to examine the longer-term impacts of the land law.

FIGURES

Figure 2.1: Land transaction trends by year



TABLES

Table 2.1 Balanced vs. Unbalanced

	Balanced
Female head	-0.0152
	(0.0158)
Head's education	0.00826
	(0.0120)
Head's age	-0.000470
	(0.000858)
Household size	0.00665**
	(0.00248)
Annual landholding size	0.00888**
	(0.00338)
Annual and perennial landholding size	0.000203
	(0.00356)
Constant	0.784***
	(0.0542)
Observations	3,267
R-squared	0.012

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.2 Household characteristics by land types

	Perennial	Annual	Difference
Female head	0.104	0.149	-0.044
	(0.0043)	(0.0016)	(0.0051)
Head's education	1.284	1.198	0.086
	(0.0139)	(0.0045)	(0.0145)
Head's age	47.877	51.247	-3.370
	(0.1621)	(0.0566)	(0.1825)
Household size	5.036	4.845	0.191
	(0.0255)	(0.0088)	(0.0283)
Household values of durable goods	2,000.478	1,092.652	907.8255
	(168.1769)	(89.1414)	(291.7778)
Plot area	7,349.429	1,695.186	5,654.238
	(127.5354)	(29.0624)	(98.8452)
Distance from plots	3,797.306	1,468.576	2,328.730
	(411.4919)	(23.7024)	(151.9553)
Plot quality	1.932	1.941	-0.009
	(0.0054)	(0.0016)	(0.0053)
Observations	5,158	49,269	

Source: VARHS by UNU WIDER. Statistics are calculated by the author. Standard errors are in parentheses.

Table 2.3 Land transactions by land types

	Perennial	Annual	Difference
Rented Out	0.022	0.076	0.054
	(0.0021)	(0.0013)	(0.0039)
Rented In	0.041	0.061	0.020
	(0.0029)	(0.0012)	(0.0036)
Sold	0.049	0.059	0.010
	(0.0032)	(0.0011)	(0.0036)
Purchased	0.055	0.045	-0.011
	(0.0033)	(0.0010)	(0.0032)

Source: VARHS by UNU WIDER. Statistics are calculated by the author. Standard errors are in parentheses.

Table 2.4 Sample's household outcomes in fraction by year

Year	Wage labor	Nonfarm wage labor	Farm wage labor	Wage labor in commune	Wage labor in province	Wage labor outside of province	Own Business	Food Expenditure per capita
2008	0.50827	0.39846	0.15269	0.31962	0.16269	0.10769	0.2077	$134.21
2010	0.56851	0.44303	0.17937	0.38645	0.18476	0.0689	0.2623	$140.23
2012	0.60039	0.46346	0.18923	0.42538	0.21385	0.045	0.2508	$175.21
2014	0.67077	0.43846	0.30923	0.51115	0.20308	0.06769	0.1996	$167.13
2016	0.67423	0.45654	0.30808	0.50231	0.22192	0.08885	0.2462	$181.39

Source: VARHS by UNU WIDER. Statistics are calculated by the author.

Table 2.5 Pre-trend tests

	(1) Rented Out	(2) Rented In	(3) Sold	(4) Purchased
Annual x Year 2008	-0.0103	0.00572	-0.00374	0.00840
	(0.00822)	(0.0075)	(0.0105)	(0.0183)
Annual x Year 2012	0.00738	0.000840	-0.0262**	-0.0258**
	(0.00772)	(0.0075)	(0.0119)	(0.0114)
Year 2008	-0.0130*	0.00145	-0.0420***	0.0253
	(0.00693)	(0.0068)	(0.0103)	(0.0161)
Year 2012	0.00680	-0.00431	0.0274**	0.0181*
	(0.00682)	(0.0069)	(0.0115)	(0.0108)
Annual	0.0327***	0.0334**	0.00316	0.00537
	(0.00900)	(0.0144)	(0.0105)	(0.0134)
Observations	30,544	30,870	30,145	30,544
R-squared	0.528	0.357	0.237	0.237

Note: Control is included throughout but coefficients that are not reported include, plot land quality, household head’s gender, age, squared age, education, and household size, distance between plot and home, log of plot area, dummy for average quality plot, and dummy for above average quality plot. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.6 Impacts of law on land transactions

	Post-law vs Pre-law				By individual year			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Rented Out	Rented In	Sold	Purchased	Rented Out	Rented In	Sold	Purchased
Annual x Post-law	0.0338***	0.00388	0.0558***	0.0233				
	(0.00922)	(0.00912)	(0.0172)	(0.0158)				
Annual x Year 2008					-0.0125	0.00453	-0.00355	0.00808
					(0.00888)	(0.00783)	(0.0106)	(0.0183)
Annual x Year 2014					0.0203**	0.00501	0.0841***	0.0303
					(0.00987)	(0.00945)	(0.0265)	(0.0186)
Annual x Year 2016					0.0351***	0.00695	0.0255*	0.0239
					(0.0109)	(0.00996)	(0.0154)	(0.0151)
Annual	0.0193*	0.035***	-0.0297***	-0.00446	0.0250**	0.0329**	-0.0274**	-0.00821
	(0.0100)	(0.0124)	(0.0110)	(0.0107)	(0.0103)	(0.0131)	(0.0132)	(0.0109)
Year 2008	-0.025***	0.00580*	-0.045***	0.0327***	-0.0133*	0.00163	-0.042***	0.0252
	(0.00507)	(0.00317)	(0.00821)	(0.0112)	(0.00763)	(0.00727)	(0.0104)	(0.0160)
Year 2014	0.00136	-0.00956	0.0438***	0.0118	0.0134*	-0.0106	0.0184	0.00547
	(0.00789)	(0.00860)	(0.0131)	(0.0113)	(0.00775)	(0.00908)	(0.0117)	(0.0130)
Year 2016	0.0198**	-0.0125	-0.00861	-0.0107	0.0189**	-0.0152	0.0178	-0.0114
	(0.00802)	(0.00867)	(0.0134)	(0.0117)	(0.00846)	(0.00955)	(0.0118)	(0.0120)
Mean	0.0708	0.0598	0.0625	0.0504	0.0708	0.0598	0.0625	0.0504
	(0.0013)	(0.0012)	(0.0012)	(0.0011)	(0.0013)	(0.0012)	(0.0012)	(0.0011)
Observations	38,573	38,122	39,072	38,573	38,573	38,122	39,072	38,573
R-squared	0.480	0.328	0.257	0.185	0.480	0.328	0.258	0.185

Note: Control is included throughout but coefficients that are not reported include, plot land quality, household head's gender, age, squared age, education, and household size, distance between plot and home, log of plot area, dummy for average quality plot, and

Table 2.6 (cont'd)

dummy for above average quality plot. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.7 Marginal effects on allocative efficiency of the law

	Post-law vs Pre-law							
	Rice				Rice and Maize			
	(1) Rented Out	(2) Rented In	(3) Sold	(4) Purchased	(1) Rented Out	(2) Rented In	(3) Sold	(4) Purchased
HA x Post-law	-0.0152	-0.0281	0.00702	-0.000491	-0.0122	-0.0327	0.0199	0.0550
	(0.0172)	(0.0224)	(0.0202)	(0.0303)	(0.0163)	(0.0211)	(0.0156)	(0.0350)
HA	-0.043***	0.0619*	-0.0236**	0.0152	-0.050***	0.0706**	-0.0279***	-0.0454
	(0.0157)	(0.0333)	(0.0109)	(0.0264)	(0.0166)	(0.0331)	(0.00987)	(0.0288)
Year 2008	-0.023***	0.0270***	-0.0895***	0.0501***	-0.023***	0.0270***	-0.0895***	0.0505***
	(0.00600)	(0.00810)	(0.00982)	(0.0178)	(0.00602)	(0.00810)	(0.00983)	(0.0179)
Year 2014	0.0120	-0.027***	0.0697***	0.0118	0.0116	-0.026***	0.0692***	0.00992
	(0.00925)	(0.00790)	(0.0183)	(0.0217)	(0.00929)	(0.00791)	(0.0183)	(0.0218)
Year 2016	0.0302***	-0.051***	-0.0130	-0.0381**	0.0295***	-0.0497***	-0.0136	-0.0402**
	(0.0107)	(0.00927)	(0.0132)	(0.0158)	(0.0108)	(0.00930)	(0.0132)	(0.0159)
Mean	0.1150	0.1347	0.0926	0.1497	0.1150	0.1347	0.0926	0.1497
	(0.0033)	(0.0035)	(0.0030)	(0.0037)	(0.0033)	(0.0035)	(0.0030)	(0.0037)
Observations	9,291	9,291	9,367	9,367	9,291	9,291	9,367	9,367
R-squared	0.107	0.032	0.065	0.023	0.109	0.034	0.065	0.024

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, and household size. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.7 (cont'd)

	By individual year							
	Rice				Rice and Maize			
	(1) Rented Out	(2) Rented In	(3) Sold	(4) Purchased	(1) Rented Out	(2) Rented In	(3) Sold	(4) Purchased
HA x Year 2008	0.0102	0.0420**	0.0243	0.0199	0.00621	0.0269	0.0185	-0.0373
	(0.0175)	(0.0201)	(0.0192)	(0.0372)	(0.0155)	(0.0174)	(0.0173)	(0.0354)
HA x Year 2014	-0.0190	-0.00436	0.0289	-2.87e-05	-0.0209	-0.0150	0.0327	0.0359
	(0.0193)	(0.0211)	(0.0343)	(0.0440)	(0.0184)	(0.0181)	(0.0272)	(0.0444)
HA x Year 2016	-0.00124	-0.0104	0.00895	0.0187	0.00292	-0.0236	0.0254	0.0370
	(0.0270)	(0.0250)	(0.0244)	(0.0375)	(0.0256)	(0.0230)	(0.0216)	(0.0344)
HA	-0.0481**	0.0412	-0.0355*	0.00544	-0.0532**	0.0572*	-0.0371**	-0.0268
	(0.0212)	(0.0340)	(0.0198)	(0.0332)	(0.0222)	(0.0323)	(0.0178)	(0.0312)
Year 2008	-0.0237***	0.0262***	-0.0900***	0.0497***	-0.0235***	0.0262***	-0.0900***	0.0515***
	(0.00604)	(0.00812)	(0.00993)	(0.0179)	(0.00606)	(0.00804)	(0.0101)	(0.0178)
Year 2014	0.0121	-0.027***	0.0693***	0.0119	0.0120	-0.0265***	0.0688***	0.0104
	(0.00927)	(0.00789)	(0.0182)	(0.0215)	(0.00929)	(0.00790)	(0.0183)	(0.0217)
Year 2016	0.0299***	-0.051***	-0.0129	-0.0385**	0.0290***	-0.0499***	-0.0137	-0.0398**
	(0.0106)	(0.00932)	(0.0132)	(0.0159)	(0.0106)	(0.00937)	(0.0133)	(0.0159)
Mean	0.1150	0.1347	0.0926	0.1497	0.1150	0.1347	0.0926	0.1497
	(0.0033)	(0.0035)	(0.0030)	(0.0037)	(0.0033)	(0.0035)	(0.0030)	(0.0037)
Observations	9,291	9,291	9,367	9,367	9,291	9,291	9,367	9,367
R-squared	0.107	0.032	0.065	0.023	0.109	0.034	0.065	0.024

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, and household size. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.8 Impacts of law on household outcomes using Lagged Annual Ratio

	Post-law vs Pre-law							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment is Lagged Annual Ratio	Wage Labor	Nonfarm Wage Labor	Farm Wage Labor	Wage Labor in Commune	Wage Labor in Province	Wage Labor Outside of Province	Owned Business	Food Expenditure Per Capita
Lagged Annual Ratio x Post Law	0.111***	0.0142	0.141***	0.136***	-0.00139	0.0304	0.0133	0.128*
	(0.0360)	(0.0375)	(0.0484)	(0.0458)	(0.0274)	(0.0235)	(0.0460)	(0.0718)
Lagged Annual Ratio	-0.0798	-0.0193	-0.107*	-0.102*	0.0103	-0.00699	-0.0805*	-0.147*
	(0.0486)	(0.0463)	(0.0588)	(0.0610)	(0.0433)	(0.0238)	(0.0421)	(0.0773)
Year 2014	0.0321	-0.00523	0.0164	0.0212	0.0184	-0.0250	-0.0836**	0.0138
	(0.0330)	(0.0326)	(0.0407)	(0.0416)	(0.0235)	(0.0211)	(0.0329)	(0.0679)
Year 2016	0.0520	0.0999*	0.0967	0.144**	0.0453	0.0201	-0.00441	0.387***
	(0.0337)	(0.0515)	(0.0634)	(0.0686)	(0.0390)	(0.0314)	(0.0621)	(0.0926)
Mean	0.6055	0.4341	0.2374	0.4299	0.1931	0.0833	0.229	4.71
	(0.0048)	(0.0049)	(0.0042)	(0.0049)	(0.0039)	(0.0027)	(0.0041)	(0.0089)
Observations	7,795	7,795	7,795	7,795	7,795	7,795	7,797	7,797
R-squared	0.539	0.608	0.524	0.505	0.525	0.437	0.567	0.667

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, household size, and lagged annual and perennial landholdings (log). Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.8 (cont'd)

Treatment is Lagged Annual Ratio	By individual years							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Wage Labor	Nonfarm Wage Labor	Farm Wage Labor	Wage Labor in Commune	Wage Labor in Province	Wage Labor Outside of Province	Owned Business	Food Expenditure Per Capita
Lagged Annual Ratio x Year 2014	0.133***	0.0345	0.129**	0.154***	0.00498	0.0181	-0.00564	0.189**
	(0.0510)	(0.0385)	(0.0527)	(0.0512)	(0.0318)	(0.0270)	(0.0398)	(0.0891)
Lagged Annual Ratio x Year 2016	0.0907*	0.00517	0.162***	0.136***	-0.00662	0.0447	0.0358	0.116
	(0.0521)	(0.0434)	(0.0527)	(0.0492)	(0.0315)	(0.0276)	(0.0631)	(0.0774)
Lagged Annual Ratio	-0.0778	-0.0205	-0.113*	-0.106*	0.0112	-0.00892	-0.085**	-0.162**
	(0.0605)	(0.0460)	(0.0595)	(0.0607)	(0.0438)	(0.0238)	(0.0431)	(0.0792)
Year 2014	0.0145	-0.0128	0.0331	0.0198	0.0143	-0.0128	-0.065**	0.000223
	(0.0433)	(0.0329)	(0.0418)	(0.0422)	(0.0261)	(0.0234)	(0.0276)	(0.0840)
Year 2016	0.0682	0.0403	0.0146	0.0409	0.0446*	-0.00958	-0.0549*	0.0835
	(0.0469)	(0.0371)	(0.0422)	(0.0453)	(0.0268)	(0.0233)	(0.0317)	(0.0652)
Mean	0.6055	0.4341	0.2374	0.4299	0.1931	0.0833	0.229	4.71
	(0.0048)	(0.0049)	(0.0042)	(0.0049)	(0.0039)	(0.0027)	(0.0041)	(0.0089)
Observations	7,795	7,795	7,795	7,795	7,795	7,795	7,797	7,797
R-squared	0.539	0.607	0.524	0.504	0.525	0.437	0.567	0.662

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, household size, and lagged annual and perennial landholdings (log). Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.9 Impacts of law on household outcomes using Log Lagged Annual Area

Treatment is Lagged Annual Area (Log)	Post-law vs Pre-law							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Wage Labor	Nonfarm Wage Labor	Farm Wage Labor	Wage Labor in Commune	Wage Labor in Province	Wage Labor Outside of Province	Owned Business	Food Expenditure Per Capita
Lagged Annual Area x Post Law	0.0233***	1.89e-05	0.0290***	0.0243***	-0.00132	0.00608**	-0.00315	0.00424
	(0.00629)	(0.00446)	(0.00744)	(0.00572)	(0.00365)	(0.00288)	(0.00744)	(0.0107)
Lagged Annual Area	-0.0138*	0.00696	-0.0253***	-0.0200***	0.00740	-0.00382	-0.00350	-0.00240
	(0.00785)	(0.00623)	(0.00720)	(0.00692)	(0.00455)	(0.00274)	(0.00575)	(0.0111)
Year 2014	-0.0461	0.0159	-0.0729*	-0.0313	0.0282	-0.0423*	-0.0449	0.126
	(0.0382)	(0.0348)	(0.0405)	(0.0357)	(0.0291)	(0.0240)	(0.0470)	(0.0778)
Year 2016	-0.0255	0.0468	-0.0648	-0.0245	0.0499*	-0.0182	-0.00130	0.152**
	(0.0393)	(0.0347)	(0.0411)	(0.0382)	(0.0298)	(0.0207)	(0.0353)	(0.0709)
Mean	0.6055	0.4341	0.2374	0.4299	0.1931	0.0833	0.229	4.71
	(0.0048)	(0.0049)	(0.0042)	(0.0049)	(0.0039)	(0.0027)	(0.0041)	(0.0089)
Observations	7,795	7,795	7,795	7,795	7,795	7,795	7,797	7,797
R-squared	0.541	0.607	0.527	0.505	0.525	0.437	0.566	0.661

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, household size, and lagged annual and perennial landholdings (log). Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 2.9 (cont'd)

Treatment is Lagged Annual Area (Log)	By individual years							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Wage Labor	Nonfarm Wage Labor	Farm Wage Labor	Wage Labor in Commune	Wage Labor in Province	Wage Labor Outside of Province	Owned Business	Food Expenditure Per Capita
Lagged Annual Area x Year 2014	0.0245***	0.000249	0.0249***	0.0248***	-0.000502	0.00189	-0.00907	0.0130
	(0.00698)	(0.00477)	(0.00809)	(0.00710)	(0.00443)	(0.00327)	(0.00582)	(0.0120)
Lagged Annual Area x Year 2016	0.0221***	-0.000195	0.0328***	0.0238***	-0.00208	0.00997***	0.00235	-0.00388
	(0.00690)	(0.00529)	(0.00780)	(0.00591)	(0.00378)	(0.00358)	(0.0101)	(0.0112)
Lagged Annual Area	-0.0137*	0.00698	-0.0256***	-0.0199***	0.00747	-0.00419	-0.00403	-0.00161
	(0.00785)	(0.00624)	(0.00716)	(0.00691)	(0.00455)	(0.00279)	(0.00583)	(0.0111)
Year 2014	-0.0552	0.0142	-0.0436	-0.0350	0.0223	-0.0121	-0.00215	0.0627
	(0.0411)	(0.0371)	(0.0432)	(0.0424)	(0.0349)	(0.0265)	(0.0373)	(0.0882)
Year 2016	-0.0173	0.0483	-0.0915**	-0.0212	0.0552*	-0.0456**	-0.0400	0.209***
	(0.0440)	(0.0379)	(0.0425)	(0.0412)	(0.0292)	(0.0224)	(0.0478)	(0.0735)
Mean	0.6055	0.4341	0.2374	0.4299	0.1931	0.0833	0.229	4.71
	(0.0048)	(0.0049)	(0.0042)	(0.0049)	(0.0039)	(0.0027)	(0.0041)	(0.0089)
Observations	7,795	7,795	7,795	7,795	7,795	7,795	7,797	7,797
R-squared	0.541	0.607	0.528	0.505	0.525	0.438	0.567	0.662

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, household size, and lagged annual and perennial landholdings (log). Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

BIBLIOGRAPHY

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. "When should you adjust standard errors for clustering?." *The Quarterly Journal of Economics* 138, no. 1 (2023): 1-35.

Abdulai, Awudu, Victor Owusu, and Renan Goetz. "Land tenure differences and investment in land improvement measures: Theoretical and empirical analyses." *Journal of Development Economics* 96, no. 1 (2011): 66-78.

Ali, Daniel Ayalew, and Klaus Deininger. "Institutional determinants of large land-based investments' performance in Zambia: Does title enhance productivity and structural transformation?." *World Development* 157 (2022): 105932.

Ayerst, Stephen, Loren Brandt, and Diego Restuccia. "Market constraints, misallocation, and productivity in Vietnam agriculture." *Food Policy* 94 (2020): 101840.

Barajas, Jorge A. Rincón. "Productivity dynamics and state support after a land titling program: Evidence from Colombia." *Land Use Policy* 131 (2023): 106690.

Bellemare, Marc F., Kenn Chua, Julieth Santamaria, and Khoa Vu. "Tenurial security and agricultural investment: Evidence from Vietnam." *Food Policy* 94 (2020): 101839.

Brandt, Loren, Jikun Huang, Guo Li, and Scott Rozelle. "Land rights in rural China: Facts, fictions and issues." *The China Journal* 47 (2002): 67-97.

Brandt, Loren, Susan H. Whiting, Linxiu Zhang, and Tonglong Zhang. "Changing property-rights regimes: a study of rural land tenure in China." *The China Quarterly* 232 (2017): 1026-1049.

Chamberlin, Jordan, and Jacob Ricker-Gilbert. "Participation in rural land rental markets in Sub-Saharan Africa: Who benefits and by how much? Evidence from Malawi and Zambia." *American Journal of Agricultural Economics* 98, no. 5 (2016): 1507-1528.

Chen, Chaoran, Diego Restuccia and Raül Santaeulàlia-Llopis. 2023. "Land Misallocation and Productivity." *American Economic Journal: Macroeconomics*, 15 (2): 441-65.

Crewett, Wibke, and Benedikt Korf. "Ethiopia: Reforming land tenure." *Review of African Political Economy* 35, no. 116 (2008): 203-220.

De Janvry, Alain, Elisabeth Sadoulet, and Wendy Wolford. *Access to land and land policy reforms*. Vol. 3. Helsinki: UNU World Institute for Development Economics Research, 2001.

Deininger, Klaus W. *Land policies for growth and poverty reduction*. World Bank Publications, 2003.

Deininger, Klaus, Daniel Ayalew Ali, and Tekie Alemu. "Assessing the functioning of land rental markets in Ethiopia." *Economic Development and cultural change* 57, no. 1 (2008): 67-100.

Deininger, Klaus, Daniel Ayalew Ali, and Takashi Yamano. "Legal knowledge and economic development: The case of land rights in Uganda." *Land Economics* 84, no. 4 (2008): 593-619.

Deininger, Klaus, Daniel Ayalew Ali, and Tekie Alemu. "Impacts of land certification on tenure security, investment, and land market participation: evidence from Ethiopia." *Land Economics* 87, no. 2 (2011): 312-334.

Deininger, Klaus, and Songqing Jin. "The potential of land rental markets in the process of economic development: Evidence from China." *Journal of Development Economics* 78, no. 1 (2005): 241-270.

Deininger, Klaus, and Songqing Jin. "Land sales and rental markets in transition: Evidence from rural Vietnam." *Oxford bulletin of Economics and Statistics* 70, no. 1 (2008): 67-101.

Deininger, Klaus, and Songqing Jin. "Securing property rights in transition: Lessons from implementation of China's rural land contracting law." *Journal of Economic Behavior & Organization* 70, no. 1-2 (2009): 22-38.

Deininger, Klaus, Songqing Jin, and Hari K. Nagarajan. "Efficiency and equity impacts of rural land rental restrictions: Evidence from India." *European Economic Review* 52, no. 5 (2008b): 892-918.

Do, Q.-T., & Iyer, L. (2003). Land Rights and Economic Development: Evidence from Vietnam. *Policy Research Working Paper*, (August), 36.

Do, Q., & Iyer, L. (2008). Land Titling and Rural Transition in Vietnam. *Economic Development and Cultural Change*, *56*(3), 531–579.

Galiani, Sebastian, and Ernesto Schargrodsky. "Property rights for the poor: Effects of land titling." *Journal of Public Economics* 94, no. 9-10 (2010): 700-729.

Gao, Xuwen, Xinjie Shi, and Shile Fang. "Property rights and misallocation: Evidence from land certification in China." *World Development* 147 (2021): 105632.

Ghebru, Hosaena, and Stein Holden. "Land rental markets and rural poverty dynamics in Northern Ethiopia: Panel data evidence using survival models." *Review of Development Economics* 23, no. 1 (2019): 131-154.

Hansen, Kaitlin. "Land law, land rights, and land reform in Vietnam: A deeper look into “land grabbing” for public and private development." (2013).

Holden, Stein T., Klaus Deininger, and Hosaena Ghebru. "Impacts of low-cost land certification on investment and productivity." *American journal of agricultural economics* 91, no. 2 (2009): 359-373.

Holden, S.T., Deininger, K. and Ghebru, H., 2010, April. Impact of land registration and certification on land border conflicts in Ethiopia. In *World Bank Annual Conference on Land Policy and Administration in Washington DC April* (pp. 26-27).

Jacoby, Hanan G., and Bart Minten. "Is land titling in Sub-Saharan Africa cost-effective? Evidence from Madagascar." *The World Bank Economic Review* 21, no. 3 (2007): 461-485.

Jin, Songqing, and Klaus Deininger. "Land rental markets in the process of rural structural transformation: Productivity and equity impacts from China." *Journal of comparative economics* 37, no. 4 (2009): 629-646.

Jin, Songqing, and Thomas S. Jayne. "Land rental markets in Kenya: implications for efficiency, equity, household income, and poverty." *Land Economics* 89, no. 2 (2013): 246-271.

Kosec, Katrina, Hosaena Ghebru, Brian Holtemeyer, Valerie Mueller, and Emily Schmidt. "The effect of land access on youth employment and migration decisions: Evidence from rural Ethiopia." *American Journal of Agricultural Economics* 100, no. 3 (2018): 931-954.

Lawry, Steven, Cyrus Samii, Ruth Hall, Aaron Leopold, Donna Hornby, and Farai Mtero. "The impact of land property rights interventions on investment and agricultural productivity in developing countries: A systematic review." *Journal of Development Effectiveness* 9, no. 1 (2017): 61-81.

Muraoka, Rie, Songqing Jin, and Thomas S. Jayne. "Land access, land rental and food security: Evidence from Kenya." *Land use policy* 70 (2018): 611-622.

Ricker-Gilbert, Jacob, and Jordan Chamberlin. "Transaction costs, land rental markets, and their impact on youth access to agriculture in Tanzania." *Land Economics* 94, no. 4 (2018): 541-555.

Seewald, Eva, Samantha Baerthel, and Trung Thanh Nguyen. "Land rental markets as a poverty reduction strategy: evidence from Southeast Asia." *Journal of Economics and Development* 25, no. 2 (2023): 102-119.

Sitko, Nicholas J., Jordan Chamberlin, and Munguzwe Hichaambwa. "Does smallholder land titling facilitate agricultural growth?: An analysis of the determinants and effects of smallholder land titling in Zambia." *World Development* 64 (2014): 791-802.

Xu, Licheng, and Xiaodong Du. "Land certification, rental market participation, and household welfare in rural China." *Agricultural economics* 53, no. 1 (2022): 52-71.

Zhang, Lan, Shuyi Feng, Nico Heerink, Futian Qu, and Arie Kuyvenhoven. "How do land rental markets affect household income? Evidence from rural Jiangsu, PR China." *Land Use Policy* 74 (2018): 151-165.

Zhou, Nan, Wenli Cheng, and Longyao Zhang. "Land rights and investment incentives: evidence from China’s latest rural land titling program." *Land Use Policy* 117 (2022): 106126.

APPENDIX

Comparative statics for the household problem

$$V = \max_{L^{Out}, L^{In}, N^{In}} pf\left(\alpha;\ \overline{L} - L^{Out} + L^{In}, \overline{N} - N^{Out}\ \right) + (r - TC^{Out})L^{Out} - (r + TC^{In})L^{In}$$
$$+ wN^{Out}$$

FOC: $pf_L = r - TC^{Out}$ or $pf_L = r + TC^{In}$ (FOC1)

and $pf_N = w$ (FOC2)

L^{Out} and L^{In} cannot be simultaneously positive if a nonzero transaction cost exists since then $pf_L = r - TC^{Out} = r + TC^{In}$, which can only be true if $TC^{Out} = TC^{In} = 0$.

If $L^{Out} > 0$, then $L^{In} = 0$. When $L^{In} = 0, \frac{dL^{In}}{dTC^{Out}} = 0$ (Proposition 2).

Total derivative of the FOC with respect to TC^{Out} yields

FOC2: $-pf_{NL}\frac{dL^{Out}}{dTC^{Out}} - p\, f_{NN}\frac{dN^{Out}}{dTC^{Out}} = 0 \Longrightarrow \frac{dL^{Out}}{dTC^{Out}}\frac{dN^{Out}}{dTC^{Out}} > 0$ given $f_{NL} > 0$ and $f_{NN} < 0$.
Also, $\frac{dN^{Out}}{dTC^{Out}} = \frac{-f_{NL}}{f_{NN}}\frac{dL^{Out}}{dTC^{Out}}$

FOC1: $-pf_{LL}\frac{dL^{Out}}{dTC^{Out}} - p\, f_{LN}\frac{dN^{Out}}{dTC^{Out}} = -1 \Longrightarrow \frac{dL^{Out}}{dTC^{Out}}\left(\frac{-pf_{LL}f_{NN}+pf_{LN}f_{NL}}{f_{NN}}\right) = -1$
$\Longrightarrow \frac{dL^{Out}}{dTC^{Out}} < 0$ (Proposition 1) and $\frac{dN^{Out}}{dTC^{Out}} < 0$ (Proposition 3).

In equilibrium, however, an increase in supply of land being rented out leads to a lower rental price and leads to a higher amount of land rented in so that in equilibrium demand is equal to supply (Proposition 3).

Furthermore, the total derivative of the household income to transaction cost TC^{Out} yields:

$\frac{dV}{dTC^{Out}} = -L^{Out} - TC^{Out}\frac{dL^{Out}}{dTC^{Out}} + w\frac{dN^{Out}}{dTC^{Out}} < 0$ (Proposition 4).

Table A2.1 Production function of household major annual crops

	(1) Rice	(2) Rice & Maize
Area (Log)	0.419***	0.609***
	(0.0648)	(0.0243)
Labor Days (Log)	0.0791***	0.110***
	(0.0171)	(0.0207)
Intermediate Input (Log)	0.251***	-0.000999
	(0.0545)	(0.00168)
Year 2010	-0.189***	-0.139***
	(0.0276)	(0.0253)
Year 2012	-0.261***	-0.156***
	(0.0340)	(0.0241)
Year 2014	-0.290***	-0.126***
	(0.0477)	(0.0238)
Year 2016	-0.292***	-0.160***
	(0.0486)	(0.0337)
Female Head	0.0574	-0.00636
	(0.0389)	(0.0324)
Head's Education	0.00323	-0.00648
	(0.0135)	(0.0128)
Head's Age	-0.00671	-0.00202
	(0.00531)	(0.00515)
Head's Age Squared	5.12e-05	1.09e-05
	(4.80e-05)	(4.64e-05)
Household Size	0.0184***	0.0186***
	(0.00456)	(0.00488)
Land Distance from Household (Log)	0.0149***	0.0312***
	(0.00522)	(0.00690)
Land Value (Log)	-0.00369*	-0.00402**
	(0.00194)	(0.00187)
Irrigation Fraction	0.0638**	0.133***
	(0.0273)	(0.0276)
Restricted Purpose Land Fraction	0.00304	0.00527
	(0.0143)	(0.0147)
LURC Fraction	-0.0430*	-0.0372
	(0.0244)	(0.0283)
Below average land quality	-0.00913	0.00608
	(0.0179)	(0.0220)
Average land quality	0.0347	0.0602
	(0.0318)	(0.0396)

Table A2.1 (cont'd)

Above average land quality	0.0496**	0.0468**
	(0.0243)	(0.0232)
Missing land quality	0.172	0.151
	(0.144)	(0.138)
Flood/Drought x Below average land quality	-0.0199	-0.0264
	(0.0356)	(0.0358)
Flood/Drought x Average land quality	-0.0774***	-0.0591**
	(0.0202)	(0.0235)
Flood/Drought x Above average land	-0.0690	-0.0330
	(0.0511)	(0.0552)
Flood/Drought x Missing Land Quality	-0.441	-0.291
	(0.297)	(0.339)
Constant	0.397***	0.230
	(0.150)	(0.152)
Observations	12,996	12,996
R-squared	0.901	0.891
Number of hhid	2,600	2,600

Note: Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A2.2 Land transactions including all years

	(1) Rented Out	(2) Rented In	(3) Sold	(4) Purchased
Annual x Year 2008	-0.0118	0.00604	-0.00518	0.00780
	(0.00871)	(0.00747)	(0.0103)	(0.0180)
Annual x Year 2012	0.00621	0.000157	-0.0268**	-0.0265**
	(0.00775)	(0.00779)	(0.0119)	(0.0107)
Annual x Year 2014	0.0219**	0.00535	0.0809***	0.0296
	(0.00985)	(0.00955)	(0.0261)	(0.0180)
Annual x Year 2016	0.0367***	0.00696	0.0171	0.0238
	(0.0109)	(0.0100)	(0.0150)	(0.0148)
Annual	0.0266***	0.0328**	-0.0144	-0.00234
	(0.00976)	(0.0132)	(0.0118)	(0.0101)
Year 2008	-0.0139*	0.000985	-0.0397***	0.0255
	(0.00752)	(0.00682)	(0.00997)	(0.0159)
Year 2012	0.00878	-0.00493	0.0250**	0.0186*
	(0.00714)	(0.00725)	(0.0114)	(0.0101)
Year 2014	0.0134*	-0.0111	0.0246**	0.00366
	(0.00771)	(0.00912)	(0.0115)	(0.0126)
Year 2016	0.0192**	-0.0157	0.0221*	-0.0135
	(0.00821)	(0.00962)	(0.0116)	(0.0116)
Observations	48,642	48,079	49,280	48,642
R-squared	0.483	0.325	0.205	0.152

Note: Control is included throughout but coefficients that are not reported include, plot land quality, household head's gender, age, squared age, education, and household size, distance between plot and home, log of plot area, dummy for average quality plot, and dummy for above average quality plot. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A2.3 Household outcomes using Log Lagged Annual Ratio including all years

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment is Lagged Annual Ratio	Wage Labor	Nonfarm Wage Labor	Farm Wage Labor	Wage Labor in Commune	Wage Labor in Province	Wage Labor Outside of Province	Owned Business	Food Expenditure Per Capita
Lagged Annual	0.0366	0.00992	0.0525	0.0641	-0.00756	-0.00270	0.0110	0.0267
Ratio x Y2012	(0.0449)	(0.0380)	(0.0461)	(0.0479)	(0.0283)	(0.0201)	(0.0334)	(0.0562)
Lagged Annual	0.131***	0.0212	0.138***	0.163***	-0.00549	0.0108	-0.00647	0.188**
Ratio x Y2014	(0.0474)	(0.0359)	(0.0477)	(0.0463)	(0.0303)	(0.0247)	(0.0370)	(0.0792)
Lagged Annual	0.0763	-0.00682	0.156***	0.136***	-0.0230	0.0368	0.0372	0.109
Ratio x Y2016	(0.0467)	(0.0400)	(0.0465)	(0.0439)	(0.0292)	(0.0255)	(0.0587)	(0.0695)
Lagged Annual	-0.0438	0.000768	-0.0963**	-0.0841*	0.0182	0.00534	-0.0635*	-0.150**
Ratio	(0.0441)	(0.0347)	(0.0410)	(0.0437)	(0.0322)	(0.0191)	(0.0332)	(0.0586)
Year 2012	0.0105	0.0220	-0.0313	-0.00822	0.0373	-0.0200	-0.0217	0.213***
	(0.0413)	(0.0318)	(0.0465)	(0.0442)	(0.0236)	(0.0176)	(0.0275)	(0.0463)
Year 2014	0.0187	-0.00161	0.0273	0.00915	0.0265	-0.00552	-0.0604**	0.00155
	(0.0402)	(0.0308)	(0.0363)	(0.0379)	(0.0254)	(0.0212)	(0.0256)	(0.0747)
Year 2016	0.0831**	0.0510	0.0221	0.0358	0.0629**	-0.00114	-0.0497*	0.0904
	(0.0422)	(0.0344)	(0.0365)	(0.0403)	(0.0251)	(0.0211)	(0.0295)	(0.0582)
Observations	10,395	10,395	10,395	10,395	10,395	10,395	10,397	10,397
R-squared	0.493	0.564	0.462	0.438	0.468	0.375	0.532	0.636

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, and household size. Robust standard errors in parentheses. Standard errors are clustered at the commune level. *** p<0.01, ** p<0.05, * p<0.10.

Table A2.4 Household outcomes using Log Lagged Annual Area including all years

Treatment is Log Lagged Annual Area	(1) Wage Labor	(2) Nonfarm Wage Labor	(3) Farm Wage Labor	(4) Wage Labor in Commune	(5) Wage Labor in Province	(6) Wage Labor Outside of Province	(7) Owned Business	(8) Food Expenditure Per Capita
Lagged Annual	0.00717	-0.00142	0.0111**	0.00856*	-0.00191	0.00239	-0.00155	0.000204
Area (log)x2012	(0.00485)	(0.00451)	(0.00513)	(0.00484)	(0.00356)	(0.00284)	(0.00450)	(0.00786)
Lagged Annual	0.0242***	-0.000604	0.0256***	0.0256***	-0.00130	0.00135	-0.00865	0.0140
Area (log)x2014	(0.00650)	(0.00450)	(0.00732)	(0.00636)	(0.00405)	(0.00307)	(0.00525)	(0.0105)
Lagged Annual	0.0211***	-0.000819	0.0320***	0.0239***	-0.00341	0.00938***	0.00271	-0.00378
Area (log)x2016	(0.00634)	(0.00499)	(0.00697)	(0.00532)	(0.00346)	(0.00331)	(0.00921)	(0.0101)
Lagged Annual	-0.0104**	0.00407	-0.019***	-0.0149***	0.00477	-0.000854	-0.00270	-0.00441
Area (log)	(0.00526)	(0.00427)	(0.00492)	(0.00476)	(0.00359)	(0.00229)	(0.00463)	(0.00783)
Year 2012	-0.0119	0.0408	-0.0690*	-0.0179	0.0452	-0.0397*	-0.000390	0.237***
	(0.0364)	(0.0338)	(0.0383)	(0.0376)	(0.0280)	(0.0227)	(0.0316)	(0.0572)
Year 2014	-0.0509	0.0203	-0.0462	-0.0446	0.0315	-0.00684	-0.00134	0.0559
	(0.0393)	(0.0353)	(0.0376)	(0.0379)	(0.0322)	(0.0248)	(0.0336)	(0.0769)
Year 2016	-0.00691	0.0528	-0.0825**	-0.0269	0.0695***	-0.0391*	-0.0369	0.210***
	(0.0409)	(0.0362)	(0.0357)	(0.0364)	(0.0268)	(0.0207)	(0.0436)	(0.0650)
Observations	10,395	10,395	10,395	10,395	10,395	10,395	10,397	10,397
R-squared	0.495	0.564	0.465	0.440	0.468	0.377	0.533	0.635

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, and household size. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A2.5 Marginal effects on allocative efficiency including all years

	Rice				Rice and Maize			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Rented Out	Rented In	Sold	Purchased	Rented Out	Rented In	Sold	Purchased
HA x Year 2008	0.0101	0.0420**	0.0244	0.0203	0.00623	0.0269	0.0185	-0.0373
	(0.0175)	(0.0201)	(0.0192)	(0.0373)	(0.0156)	(0.0174)	(0.0173)	(0.0355)
HA x Year 2012	0.00239	0.0104	-0.0193	0.00364	0.00315	-0.00290	-0.00868	0.0562*
	(0.0139)	(0.0194)	(0.0235)	(0.0306)	(0.0121)	(0.0165)	(0.0237)	(0.0302)
HA x Year 2014	-0.0190	-0.00436	0.0286	-5.69e-05	-0.0209	-0.0149	0.0325	0.0359
	(0.0193)	(0.0211)	(0.0343)	(0.0440)	(0.0185)	(0.0181)	(0.0272)	(0.0445)
HA x Year 2016	-0.00111	-0.0104	0.00857	0.0188	0.00304	-0.0235	0.0251	0.0371
	(0.0270)	(0.0250)	(0.0244)	(0.0375)	(0.0256)	(0.0229)	(0.0216)	(0.0344)
Year 2008	-0.023***	0.0262***	-0.0902***	0.0491***	-0.023***	0.0263***	-0.0902***	0.0508***
	(0.00602)	(0.00812)	(0.00995)	(0.0180)	(0.00604)	(0.00804)	(0.0101)	(0.0179)
Year 2012	0.00390	-0.0111*	-0.00662	-0.0251*	0.00370	-0.0106*	-0.00682	-0.0267**
	(0.00693)	(0.00589)	(0.0110)	(0.0128)	(0.00698)	(0.00586)	(0.0111)	(0.0128)
Year 2014	0.0119	-0.027***	0.0692***	0.0129	0.0117	-0.026***	0.0687***	0.0116
	(0.00927)	(0.00793)	(0.0183)	(0.0217)	(0.00928)	(0.00794)	(0.0183)	(0.0218)
Year 2016	0.0294***	-0.0508***	-0.0130	-0.0369**	0.0285***	-0.0496***	-0.0138	-0.0378**
	(0.0107)	(0.00931)	(0.0133)	(0.0161)	(0.0107)	(0.00936)	(0.0133)	(0.0162)
HA	-0.0489**	0.0406	-0.0352*	0.00642	-0.0541**	0.0570*	-0.0365**	-0.0248
	(0.0212)	(0.0339)	(0.0197)	(0.0332)	(0.0224)	(0.0323)	(0.0176)	(0.0312)
Observations	11,622	11,622	11,720	11,720	11,622	11,622	11,720	11,720
R-squared	0.106	0.031	0.056	0.023	0.107	0.032	0.056	0.024

Table A2.5 (cont'd)

Note: Control is included throughout but coefficients that are not reported include, household head's gender, age, squared age, education, and household size. Standard errors are clustered at the commune level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.