Transportation Costs and Economic Development

Nicole Stephanie Hoseini

A Dissertation
submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Graduate Program in Economics
York University
Toronto, Ontario

August 2024

Abstract

This dissertation contributes to a growing literature using microeconomic data to explore questions in macroeconomic development, with particular focus on the importance of transportation costs. Specifically, I study the importance of idiosyncratic transportation costs and economic development via three different angles.

In **Chapter 1**, I study how idiosyncratic transportation frictions as labor mobility barriers affect the sectoral sorting of workers between agriculture and non-agriculture and quantify their impact on aggregate and sectoral productivity, the pattern of occupational choices and selection. I add idiosyncratic transport costs by sector to an otherwise canonical two sector Roy model. I combine panel data on household level transport costs and income from Honduras with a structural model to quantitatively estimate the impact observed transport costs on the reallocation of labor and aggregate development through this selection channel. When removing transport costs, share of agricultural employment drops by 8 percent, agricultural productivity increases 1.32-fold and real GDP per worker rises 1.19-fold.

In **Chapter 2**, I examine whether the geographic location of farmers, and their distance from markets can account for measured misallocation in the data. I quantitatively examine this question by leveraging a transport infrastructure development program in El Salvador. I use panel micro-level data on transport costs and agricultural production at the farm level along with a structural model in which farmers produce subject to transportation costs. The key insight of my model is that in the presence of transport costs the implied efficient allocation is different than that in the canonical model of misallocation.

In **Chapter 3**, I explore the role idiosyncratic transportation costs from farm to market play in contributing to a pronounced subsistence agricultural sector. Particularly, I study how idiosyncratic transportation costs to market affect crop commercialization among farmers in Tanzania in the long rainy season. Methodologically, I combine panel data from the Tanzania National Panel Survey with a structural model to quantitatively examine the role transport costs play in facilitating the transformation from subsistence to commercial farming. I find that reductions in transport costs for food crop farmers, sees switching to cash crop farming.

Dedication

For my beloved mother and father, Erika and Morteza Hoseini.
Thank you for your unwavering support and love throughout this journey in my life.
I couldn't have done this without you both.

Acknowledgements

This dissertation would not have been possible without the guidance, support, inspiration and encouragement of my dissertation supervisor, Professor Tasso Adamopoulos. From the start of my research, way back in 2019, up until now, he always remained there to help, with kindness and patience. I have learned so much from him, and am grateful to have had such an amazing supervisor like him. He is not only a brilliant educator, but also someone who challenges his students to strive for excellence. He truly made every step of this journey in my doctoral studies memorable and I am so thankful for all of his efforts.

To my two committee members, Professor Chaoran Chen and Professor Ahmet Akyol, I want to express my deepest gratitude for all your feedback, useful input and support during my research endeavours. Both of you have provided your time, thoughts, expertise and advice on my work which have been of the utmost importance for my development as a doctoral student. You both have served to shape me as a better researcher and I am so appreciative.

To my fellow cohort members, whom I started my PhD with, back in fall of 2017, Zifan Bi and Anfal Adawi, I will never forget the moments we shared–studying on campus together, sharing many laughs in our office, helping one another out and just being there, through great times and challenging ones. I never expected to make such dear friends during my doctoral studies, but I leave with wonderful memories thanks to you both, and lifelong friends.

To my friends at the Macro Development Group, Hyungju (Henry) Kim, Zahra Shahidi, Thinh Dao, Shadi Rahimfallah and Sreyphea Chheang, thank you so much for all your feedback and support on my work.

To the wonderful and hardworking administrative team at the Economics Department, Regina Pinto, Rex Ferati, the late Lisa Guidi, and Sara Bang, thank you for all your support and administrative assistance.

Last but not least, to my dear family and close friends, thank you for being there for me during this entire journey. From the bottom of my heart, I am forever grateful to all of you.

Contents

List of Tables

List of Figures

Chapter 1

Transport Costs, Selection and Development: A Quantitative Analysis with Panel Data from Honduras

1.1 Introduction

One of the key stylized facts of the growth and development literature is the huge disparity in aggregate productivity across countries. Agriculture plays an important role in understanding this disparity. Poor countries are characterized by: (1) relatively lower productivity in agriculture than the rest of the economy and (2) greater share of their labor force engaging in agricultural activities (Restuccia et al., 2008). Not only are there large productivity gaps between non-agriculture and agriculture, but lower income countries also see considerable wage gaps across rural and urban settings too (Young, 2013).

Given the gaps in wages and productivity, a puzzle emerges: why aren't more people moving from rural to urban areas in less developed countries, if there are gains to be made outside farming? One possibility is that productivity and wage gaps are potentially being mismeasured across the agricultural and non-agricultural sectors. However work by Gollin et al. (2014) show that these gaps, although they change, do not go away, even after making several corrections to the data (Gollin et al., 2014). Two broad views remain about the urban-rural wage gaps–(1) selection and (2) labor misallocation.

According to the selection view, pioneered by Lagakos and Waugh (2013), the rural-urban wage gaps reflect the self sorting of workers across the agricultural and non-agricultural sector on the basis of two potential factors: (1) observable characteristics (Young, 2013; Miguel and Hamory, 2009) or (2) unobservable characteristics (Alvarez, 2020). According to the misallocation view, workers are prevented from moving from rural to urban areas, despite these wage gains, due to distortions, frictions, or barriers to labor mobility (Pulido et al., 2018).

I examine the role transport costs play in obstructing not only the overall allocation of labor across sectors but also the type of workers that sorts across occupations. In particular, I quantify the importance of costs of transportation as a specific barrier to labor mobility across sectors in Honduras. I contribute to the literature by measuring the impact of transport costs on the sectoral re-allocation of labor, the structural transformation and development, through the selection channel, in Honduras. Particularly, I study how separate idiosyncratic transport frictions to each sector affects sectoral sorting of workers between agriculture and non-agriculture and their impact on aggregate productivity.

I add to the canonical two sector Roy model a transport cost to participating in either sector. This participation cost is a labor mobility distortion represented as a commute time from one's residence to their main job. I address the following question: Could it be that those who are sorting into agriculture are doing so due to the mere close proximity to the job (shorter/easier commute)? How do these costs correlate with agricultural ability or productivity? Perhaps those better at agricultural jobs are unfortunately running up against more expensive or arduous commute times? In turn, could it be that those we want to be sorting into the non-agricultural sector (based off the fact that they are better at these jobs) are running up against more expensive transport or longer/difficult commute times to these non-agriculture jobs?

To explore this, I conduct various counterfactual experiments where I ask the following, "What would have the pattern of occupational choice, selection and aggregate productivity had been in 2008, if the transport distortions were reduced?" In order to explore this, I do the following: (1) use panel data from the 2008-2011 MCC/Honduras Transport and Farms to Market Roads Project which includes individual occupational choices, incomes with high resolution geographic data on transport costs, (2) develop and estimate a tractable two-sector general equilibrium (GE) model with heterogeneous abilities and idiosyncratic transport costs to both sectors and (3) assess the quantitative impact of removing observed transport costs on the pattern of occupational choices, selection and aggregate productivity.

The results from my main counterfactual experiment where heterogeneous transportation costs are eliminated, indicate that removing transport costs to both sectors, has substantial effects in Honduras. I find that eliminating transport costs leads to a decrease in the share of employment in the agricultural sector by 8 percentage points. In addition, there is a 1.32-fold increase in agricultural productivity primarily coming from the improved selection of workers into agriculture. Labor productivity and average ability in the agricultural sector and GDP per worker sees an increase.

It should be noted that while the magnitude of the effects are specific to the country and data set I study, the role of transport costs as a labor mobility barrier have more general applicability. Honduras is not the poorest country in the world, nor does it have the worst infrastructure in the world. According to the World Bank, Honduras saw a GDP per capita at 1739.40 US dollars in 2008 (Schwab and Porter, 2008). On top of that, according to World Economic Forum, Executive Opinion Survey 2007, 2008, the country's overall rank on the measure of its "Quality of Overall Infrastructure," saw a score of 3.5 and a score of 3.7 on the measure of its "Quality of Roads," where a score of 1 is denoted as "underdeveloped," and a score of 7 is denoted as "extensive and efficient by international standards," (WB, 2008). Given this, my work is consistent with the literature on the topic of transportation infrastructure, particularly work from Adamopoulos (2011). He finds nonlinear effects on improving transportation, such that the impact is greater the poorer the infrastructure a country starts from. The idea is that the said mechanism is stronger among poorer countries that require better infrastructure networks.

The structure of this chapter is as follows. In Section 2, I discuss the literature on the role that transport costs play for development. Section 3 outlines the Transportation and Farms to Market Roads Project in Honduras. Section 4 provides details on the data I use. Section 5 details the set-up of my model. Section 6 displays some analytical results. Section 7 provides experiments on the effect of counterfactual transport cost changes for the patterns of selection in Honduras. The calibration approach is discussed in Section 8, including my calibrated moments, calibrated parameters and my simulated benchmark economy. Section 9 details the quantitative results. Section 10 concludes.

1.2 Literature Review

On the topic of the rural-urban wage gaps, work by Miguel and Hamory (2009) in Kenya, test the selection argument and its validity in explaining the rural-urban wage gap seen in their country of study. Although that they do find that those who have higher abilities (measured in terms of one's skill set cognitive-wise and wellness level), do see a higher likelihood of moving into urban areas in Kenya, from rural areas, their work does not find support for selection alone, as a considerable wage gap is still left unexplained (Miguel and Hamory, 2009). There is also the argument that if the agricultural sector sees a positive correlation between comparative and absolute advantage then this would too lend support to the worker selection reasoning (Alvarez-Cuadrado et al., 2019). However, this hypothesis was investigated by Alvarez-Cuadrado et al. (2019) and found to not hold true in an empirical setting for three out of their four Sub-Saharan African countries. Work by Pulido et al. (2018) test the misallocation argument and find that when labor mobility frictions are removed in Indonesia, output can see a 21 percentage point hike.

The literature indicates that transport costs have important implications for: (1) sectoral composition of the economy, (2) the productivity of the agriculture sector and (3) structural transformation (Gollin and Rogerson, 2014; Adamopoulos, 2011) . Having said that, what is commonly addressed surrounding transport costs is its role on its effect on the trade of goods within countries (see (Adamopoulos et al., 2019; Adamopoulos, 2011; Gollin and Rogerson, 2014)), whereas my paper is about how idiosyncratic transport costs affect the mobility and sorting of people across sectors (*not goods*), and its effect on structural transformation through this channel.

Gollin and Rogerson (2014) look at the role transport costs of the ice-berg type in delivering goods across rural near, rural remote and urban, serve to explain the large share of persons sitting outside urban areas. They argue, that if the cost of transporting agricultural goods from rural to urban areas, becomes pricier, then the price of food items will be relatively more expensive to those in the urban sector compared to the rural sector. For that reason, they justify that individuals will elect to sit close to their food source (rural area), in the event that the non-agricultural sector is met with high costs of delivering the said food. People therefore will choose to reside outside of the non-agricultural sector, because they need food to survive, and it is too costly to do so outside of the rural sector.

In this sense, the point of my paper is closer to that of Asher and Novosad (2020). They found that the introduction of rural roads in India saw large movements of workers out of the agricultural sector and into the non-agricultural sector. Using a fuzzy regression discontinuity approach, Asher and Novosad found that the introduction of paved roads in rural India saw the share of workers in the agricultural sector drop by a measure of 9 percent (Asher and Novosad, 2020). However, my paper does differ from Asher and Novosad in two ways. First, I examine the effect of improvements to roads on selection and the self-sorting of households across sectors (i.e., not just exiting out of agriculture). Second, methodologically, I combine the micro and infrastructure data with a structural model to quantitatively estimate the impact of improvements to highways and roads, those that were specifically targeted in the transport project in Honduras.

It is important to note that, work by Adamopoulos (2024) does assess how changes to transport costs affect the allocation of labor across the two sectors in question, similar to my work. And, the author does find positive responses in the sectoral allocation of labor (questions of re-allocation) as a result to reductions in costs of transport. However, it is important to highlight that although we study the same research question, what separates my work to what Adamopoulos did, is I have micro data, individual heterogeneity, selection and how I introduce the respective transport cost in my model. Unlike my model, Adamopoulos sees his model include iceberg costs for the transport of *goods*, measured in terms of physical units, whereas my model includes iceberg commuting costs for the transport of *people*, measured in terms of time in minutes.

1.3 Honduras Transportation Project

From 2005-2010, a compact was made between the Millennium Challenge Corporation (MCC) and Honduras whose goal was two fold– (1) to see economic growth realized and (2) the reduction of poverty. The compact consisted of $215 million for two respective projects of interest: (1) *the Rural Development Project* and (2) the *Transport Project*. The intention behind the compact was to alleviate what was argued to be the main barriers to Honduras' growth. These main barriers identified to be addressed are twofold: (1) poor levels of agricultural productivity and (2) expensive costs of transport. The Transport Project and Farm to Market Roads Activity (part of the Rural Development Project) sought to do the following: reduce transport costs and repair of two main areas of Highway CA-5. It also aimed to see 70 km of secondary roads improved and paved. *Under the Rural Development Project*, it aimed to repair and pave of farm-to-market roads to a stretch of 600 km in distance (NORC, 2017).

1.4 Data

The micro data I use (Caldwell et al., 2013) is from the household survey from the transportation project in Honduras, where the household is denoted as the unit of analysis. Round one (before the program intervention) is August 2008 and round two (after program intervention) is April 2011. As Shana (2011) reports, the household survey[1] data contain important information on the topic of incomes and occupational information, and also includes information included but not limited to, information related to socio-economic and demographic measures, land values, and information related to travel, among others[2].

It is also important to note that the panel survey was designed as follows: stratified two-stage sample where *caserios*[3] are the first stage sample units. The justification to use caserios as the first stage sample unit, in the stratified two-stage sample is twofold: (1) sampling size is adequate and (2) there is considerable geographic information at hand. 2000 households situated in 100 caserios, was the initial sample size intent, but a final sample size of 1600 households was selected, given that some caserios failed to have 20 households. Lastly, the survey itself was to be completed by the head of the household (Shana, 2011).

Here, I focus on the set of households with data from the survey in baseline round one (2008). I created a distinction between who is an agricultural versus non-agricultural worker as their main job through code. Everyone who reported a main job in skilled agricultural jobs and alike were set as agricultural workers and everyone else was set as a skilled-non agricultural worker. Households who reported a main job in unskilled (including agriculture) elementary occupations were dropped. Travel time is reported in minutes, for the time spent commuting from home to main job which sits in either of the two sectors in question[4]. For the correlation across incomes, I look at the contemporaneous correlation or households who concurrently report monthly incomes in both sectors. Table 1.1, presents the empirical

[1] This specific survey is analytic in make up (Shana, 2011).

[2] For more information on the data, please see Caldwell et al. (2013).

[3] Caserios, in Spanish, are defined as groups of houses of poor individuals located in rural settings.

[4] Baseline data does not provide information on mode of transport for main occupation commute.

observed population sample moments[5] from the household survey data in Honduras.

Correlations	**Baseline Round 1**
Correlation across incomes	0.15
Corr across income in sector n and transport cost among those in sector n	0.02
Corr across income in sector a and transport cost among those in sector a	-0.02
Summary of Variables–STD	**Baseline Round 1**
Transport Cost in Agriculture	0.44
Transport Cost in Non-Agriculture	0.48
Income in Agriculture	0.66
Income in Non-Agriculture	0.58

Table 1.1: Empirical Moments From Micro Data in Honduras

1.5 Model of Transport Costs and Selection

My framework for sorting builds on the work of: Roy (1951), Lagakos and Waugh (2013) and Adamopoulos et al. (2022), in their work in China. However, differently from previous work, I introduce the friction of idiosyncratic transportation costs to both sectors and I quantify it.

1.5.1 Environment

I develop a tractable two-sector (agricultural and non-agricultural) general equilibrium (GE) model with heterogeneous abilities and idiosyncratic transport costs[6] to both sectors, in order to evaluate the quantitative impact of observed transport costs on the pattern of occupational choices, selection and aggregate productivity in Honduras. The economy is inhabited by a continuum of agents of a measure of 1, where these individuals face a sectoral occupational choice (Roy, 1951). The sectoral occupational choice involves an individual becoming a worker in the agricultural sector (sector a) or a worker in the non-agricultural sector (sector n). In addition, agents bear a cost to participating in their main job captured by a commute time which is different for everyone. Finally, two goods are produced at every

[5] Data is not weighted.

[6] Baseline data does not provide information on distance to main job and as a result cannot distinguish in data between km and transport cost per km variation across individuals.

date— an agricultural good denoted by a and a non-agricultural good denoted by n. The n good is the numeraire, with p_a being the relative price of the agricultural good.

1.5.2 Agent Abilities and Idiosyncratic Transport Distortions

Each agent i is heterogeneous in relation to their abilities in sectors a and n and agricultural transportation distortion φ_i and non-agricultural transportation distortion η_i[7]. As a result each individual is endowed with abilities in both sectors, and sector specific transport distortions. Specifically, the vector $(s_{ai}, s_{ni}, \varphi_i, \eta_i)$ is drawn from a known population joint distribution of skills and distortions, with density $f(s_{ai}, s_{ni}, \varphi_i, \eta_i)$ and cdf $F(s_{ai}, s_{ni}, \varphi_i, \eta_i)$. I allow for correlation of skills over sector n and sector a and with idiosyncratic transport costs φ_i and η_i. A log-normal distribution is assumed for $(s_{ai}, s_{ni}, \varphi_i, \eta_i)$ with mean $(\mu_a, \mu_n, \mu_\varphi, \mu_\eta)$ and variance-covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_a^2 & \sigma_{an} & 0 & \sigma_{a\varphi} \\ \sigma_{an} & \sigma_n^2 & \sigma_{n\eta} & 0 \\ 0 & \sigma_{n\eta} & \sigma_\varphi^2 & 0 \\ \sigma_{a\varphi} & 0 & 0 & \sigma_\eta^2 \end{bmatrix}$$

Finally, the three correlation coefficients of interest are detailed next[8].

The correlation coefficient for abilities ρ_{an} is:

$$\rho_{an} = \frac{\sigma_{an}}{\sigma_a \sigma_n}$$

[7]Details on the makeup of the labor mobility distortion will be outlined on next page.

[8]Commute times were only reported for main occupations not secondary occupations and for that reason author was unable to quantify the correlation between commute times for those who worked two jobs given the missing data. Therefore, the respective correlation across commute times was set to zero.

The correlation coefficient for n and η is:

$$\rho_{n\eta} = \frac{\sigma_{n\eta}}{\sigma_n \sigma_\eta}$$

The correlation coefficient for a and φ is:

$$\rho_{a\varphi} = \frac{\sigma_{a\varphi}}{\sigma_a \sigma_\varphi}$$

I assume the two labour mobility barriers specific to each sector φ_i and η_i are inversely related to travel times (in minutes) tt_{ihj} from home to main workplace, where i is the agent index i from their home h to their main workplace in sector j. The labor mobility barriers will be thought of as *iceberg commuting costs*,

$$\varphi_i = \frac{1}{tt_{ih}} \quad and \quad \eta_i = \frac{1}{tt_{ih}}$$

where φ_i is the iceberg commute cost for those whose main job is in the agricultural sector and η_i is the iceberg commute cost for those whose main job is in the non-agricultural sector. Iceberg commuting costs represent the cost individual i's commute is from his/her home to his/her occupation[9] (Redding and Rossi-Hansberg, 2017).

Briefly, how is one to understand iceberg transport costs with respect to the *delivery of people and not goods* here? In the standard trade case where iceberg transport costs refer to the delivery of goods, these are the part of the good that *melts in transit*. What about with respect to the delivery of people? The idea here is that labor is endowed with total labor time. The amount of labor time that is spent on the job actually working is less than what it could have been, because part of the labor time was spent actually getting to the job. Using the language of iceberg transport costs, it is the labor time that *melts* in transit commuting to the job. Therefore, the part of that labor time that *melts* in transit or was *lost* is the proportional payment to the transport sector. To summarize payments to the

[9] In this case, this is in the context of a non-traded good.

transport sector, I introduce τ_a and τ_n as the respective sectoral taxes:

$$1 - \tau_a = \varphi_a \quad and \quad 1 - \tau_n = \eta_n$$

where

$$(1 - \tau_a) \cdot s_a = \varphi_a \cdot s_a = \hat{s_a} \quad and \quad (1 - \tau_n) \cdot s_n = \eta_n \cdot s_n = \hat{s_n}$$

$\hat{s_a}$ and $\hat{s_n}$ are effective ability for each respective sector. Therefore, the part of labor time that *melts* in transit is equal to $\tau_a \cdot s_a$ and $\tau_n \cdot s_n$ for each respective sector.

1.5.3 Preferences and Budget Constraint of Agent i

Every individual has preferences over agricultural and non-agricultural goods shown as follows:

$$U = \phi \log(c_{ai} - \bar{a}) + (1 - \phi) \log c_{ni},$$

where c_{ai} represents the consumption of the agricultural good for each individual i, and c_{ni} represents the consumption of the non-agricultural good for each individual i. Also, $\bar{a}$ is a minimum subsistence requirement for the agricultural good. ϕ represents the preference weight on agricultural goods. Finally, the budget constraint which agent i faces is:

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

where I_i is agent i's income.

1.5.4 Consumption Allocation of Agent i

In order to see agent i's consumption decision across the two sectors in question, he or she maximizes his or her utility (U) subject to his or her budget constraint. That is,

$$maxU = \phi \log(c_{ai} - \bar{a}) + (1 - \phi) \log c_{ni}$$

subject to

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

which implies the following result:

$$c_{ai} = \bar{a} + \frac{\phi}{p_a} \cdot (I_i - p_a \cdot \bar{a})$$

$$c_{ni} = (1 - \phi) \cdot (I_i - p_a \cdot \bar{a})$$

1.5.5 Production in Non-Agriculture

The representative firm hires all workers choosing to work in non-agriculture according to a constant returns to scale (CRTS) technology linear in labor input,

$$Y_n = A_n \hat{Z}_n$$

where Y_n is the real non-agricultural output. A_n represents total factor productivity (TFP) in the non-agricultural sector. Lastly, $\hat{Z}_n$ represents the total amount of effective labour input from workers in the non-agricultural sector purged of time commuting to job where,

$$\hat{Z}_n = \int_{i \in H_n} s_{ni} \eta_i di$$

$$H_n = \{(s_{ai}, s_{ni}, \eta_i) : I_{ai} < I_{ni}\}$$

where H_n is the set of (s_{ai}, s_{ni}, η_i) that agent chooses non-agriculture.

1.5.6 Production in Agriculture

The representative firm hires all workers choosing agriculture where the good is produced with a CRTS technology,

$$Y_a = A_a \hat{Z}_a$$

where Y_a represents the real agricultural output. A_a represents total factor productivity (TFP) in the agricultural sector. Lastly, $\hat{Z}_a$ again is the total amount of effective labour input from workers in the agricultural sector purged of time commuting to job where,

$$\hat{Z}_a = \int_{i \in H_a} s_{ai} \varphi_i di$$

$$H_a = \{(s_{ai}, s_{ni}, \varphi_i) : I_{ai} > I_{ni}\}$$

where H_a is the set of $(s_{ai}, s_{ni}, \varphi_i)$ that agent chooses agriculture.

1.5.7 Incomes

Each worker in sector $j \in \{a, n\}$ receives w_j per efficiency unit. A worker of non-agricultural ability s_{ni} receives income in the non-agriculture sector of $I_{ni} = (1-\varepsilon) \cdot w_n \cdot s_{ni} \cdot \eta_i$.

The parameter ε represents a common (across individuals) barrier from sector a to sector n and is designed to account for all other aspects that prevent accessibility to prospects outside of sector a. Given transport costs, ε captures residual labor mobility. Quantitatively it matches the gap in labor productivity across sectors in aggregate terms. That is, using transport costs alone would not be capturing the whole picture. A worker of agricultural ability s_{ai} receives income in agriculture of $I_{ai} = w_a \cdot s_{ai} \cdot \varphi_i$.

1.5.8 Firm First Order Necessary Conditions

The FONC for the agriculture firm implies that the wage per effective unit of labor in sector a is:

$$w_a = p_a \cdot A_a$$

The FONC for the non-agriculture firm implies that the wage per effective unit of labor in sector n is:

$$w_n = A_n$$

1.5.9 Individual Occupational Choice

Given $(s_{ai}, s_{ni}, \varphi_i, \eta_i)$, agent i will choose to work in the sector which has the highest potential earnings. That is, agent i chooses to become an agricultural worker (that is, $i \in H_a$) in sector a if $I_{ai} \geq I_{ni}$ or $w_a s_{ai} \varphi_i \geq \eta_i s_{ni} w_n (1-\varepsilon)$. Similarly, agent$i$ chooses to work in sector n (that is, $i \in H_n$) if $I_{ni} \geq I_{ai}$ or $\eta_i s_{ni} w_n (1-\varepsilon) \geq w_a s_{ai} \varphi_i$. $\eta_i s_{ni}$ can be regarded as the effective ability in the non-agriculture sector $\hat{s}_{ni}$. Similarly, $\varphi_i s_{ai}$ can be regarded as the effective ability in the agriculture sector $\hat{s}_{ai}$.

1.5.10 Market Clearing Conditions

There are three market clearing conditions. The market clearing condition for agricultural goods is $Y_a = C_a$. The market clearing condition for non-agricultural goods is $Y_n = C_n$. That is, total output for each good is used for consumptions purposes. Lastly, the market clearing condition for labor is $N_a + N_n = N$, where N_a and N_n are the total number of individuals employed in each sector. Therefore, each includes the total labor time-labor time spent working and labor time spent commuting.

1.5.11 Definition of Competitive Equilibrium

A competitive equilibrium is a set of prices (p_a) an agricultural firm allocation (Y_a, N_a) an allocation for the non-agricultural firm (Y_n, N_n), occupational choice, and a consumption allocation (c_{ai}, c_{ni}) for each agent i, where the following 4 conditions hold: (1) for each agent i, (c_{ai},c_{ni}) solves his or her utility maximization problem given their budget constraint, prices, abilities, and transport distortions, (2) (Y_n, N_n) solves the non-agricultural firm's profit maximization problem subject to prices, (3) (Y_a, N_a) solves the agricultural firm's profit maximization problem subject to prices, and (4) the labour market, market for agricultural goods, and market for non-agricultural goods all clear.

1.6 Analytical Results

I show some analytical results exploiting, properties of the log-normal distribution over $(s_{ai}, \varphi_i, \eta_i, s_{ni})$. When $(s_{ai}, \varphi_i, \eta_i, s_{ni})$ are drawn from a distribution which is log-normal, then the share of employment in sector a is defined by the following,

$$n_a = \Phi(b_1)$$

where $\Phi(.)$ is standard normal cdf. Also, b_1, b_a and b_n is defined as follows:

$$b_1 = \tfrac{b_a - b_n}{\sigma^*}$$

$$b_a \equiv \log(w_a) + \mu_\varphi + \mu_a$$

$$b_n \equiv \log(w_n) + \mu_\eta + \mu_n$$

and σ^* is the variance of relative effective abilities between the two sectors. Appendix A provides more details on how the above expressions is derived.

1.6.1 Possible Patterns of Sorting and Average Quality

Following Heckman and Honore (1990) and Adamopoulos et al. (2022), I calculate in my model conditional moments by sector conditional on the sectoral occupation chosen.

The *average log-effective ability* in **non-agriculture** among those that choose to work in non-agriculture is,

$$\mathrm{E}[\log(\hat{s_n}) \mid \log(I_n) - \log(I_a) > 0] = \hat{\mu_n} + \tfrac{\sigma_{nn} - \hat{\sigma}_{an}}{\sigma^*} \times \underbrace{\lambda^+(b_1)}_{\text{Heckman's Lambda}}$$

The *average log-effective ability* in **agriculture** among those that choose to work in agriculture is,

$$\mathrm{E}[\log(\hat{s_a}) \mid \log(I_a) - \log(I_n) > 0] = \hat{\mu_a} + \tfrac{\sigma_{an} - \sigma_{aa}}{\sigma^*} \times \underbrace{\lambda^-(b_1)}_{\text{Heckman's Lambda}}$$

$\hat{\mu}_n = \mu_n + \mu_\eta$

$\hat{\mu}_a = \mu_a + \mu_\varphi$

$\lambda^-(b_1) = \mathrm{E}[\xi_1 \mid \xi_1 \leqslant b_1]$

$\lambda^+(b_1) = \mathrm{E}[\xi_1 \mid \xi_1 > b_1]$

and $\lambda(b_1)$ represents Heckman's Lambda, where $\lambda^+(b_1)$ and $\lambda^-(b_1)$ is the truncation of a standard normal variable ξ_1, of the upper and lower tail, respectively.

Conditional expectations of log effective sectoral abilities *can also be represented as follows*:

$$\mathrm{E}[\log(\hat{s_a}) \mid \log(I_a) - \log(I_n) > 0] = \hat{\mu_a} + \tfrac{\hat{\sigma_n}\hat{\sigma_a}}{\sigma^*}[\rho_{an} - \tfrac{\hat{\sigma_a}}{\hat{\sigma_n}}]$$
$$\mathrm{E}[\log(\hat{s_n}) \mid \log(I_n) - \log(I_a) > 0] = \hat{\mu_n} + \tfrac{\hat{\sigma}_n\hat{\sigma_a}}{\sigma^*}[\tfrac{\hat{\sigma_n}}{\hat{\sigma}_a} - \rho_{an}]$$

This was achieved by understanding that everything following $\hat{\mu_a}$ and $\hat{\mu_n}$ can be mirrored with $\frac{\hat{\sigma_n}\hat{\sigma_a}}{\sigma^*}[\rho_{an} - \frac{\hat{\sigma_a}}{\hat{\sigma_n}}]$ and $\frac{\hat{\sigma}_n\hat{\sigma_a}}{\sigma^*}[\frac{\hat{\sigma_n}}{\hat{\sigma}_a} - \rho_{an}]$, respectively.

Therefore, the average quality of those that choose to go to sector depends on two things:

1. *correlation of effective abilities across sectors* (ρ_{an})
2. *dispersion of effective abilities in non-agriculture* $\hat{\sigma}_n$ *and dispersion of effective abilities in agriculture* $\hat{\sigma}_a$

As an illustration, if $\rho_{an} > 0$ and $\hat{\sigma}_{aa} > \sigma_{an}$ and $\hat{\sigma}_{nn} < \sigma_{an}$, then:

1. mean of log-effective skill of those that choose sector a (agriculture) will surpass the conditional mean of the population of these skills
2. mean of log-effective skill of those that choose sector n (non-agriculture) will sit below the conditional mean of the population of these skills

1.7 Pattern of Selection Experiment

Here, I provide examples of what is theoretically possible along the questions of potential patterns of sorting. I conduct a number of experiments across three cases– (1) absolute advantage (AA) in the dispersion of abilities in sector a, (2) absolute advantage (AA) in the dispersion of effective abilities in sector n and (3) comparative advantage (CA) (which is where my case with transport costs falls, as the calibrated population moments below indicate). Now I conduct a quick experiment to see how a different transport cost barrier can alter the pattern of selection and have provided three illustrations below. That is, for each case, I ask, *how would the pattern of selection change relative to the initial one (before experiment), if transport costs are the only thing I change?*

The graphs below show two lines– that of the indifference income line and the correlated abilities line. The income indifference line separates the space into two parts, where if you are below the line, one goes to sector a, with the opposite holding true if sitting above the line. What happens to an individual will depend on their draw of $(s_{ai}, s_{ni}, \varphi, \eta)$ which determines their effective abilities. One can imagine that every persons' individual specific draw, will sit like a 'dot' around the correlated abilities line. That is, the correlated abilities line, will say on average, what happens to pattern of selection.

To answer the question of the effect of transport costs, involves rotating and changing the slope of the correlated abilities line—to do this I need to be specific about how transport costs change effective ability. The variance of effective ability is determined by two things: (1) variance of transport costs and (2) correlation of ability and transport costs. In each case, I work out what would happen, theoretically, if (1) the variance of transport cost is reduced and (2) the correlation of ability and transport costs is weakened.

1.7.1 Absolute Advantage with high $\hat{\sigma}_n^2$

$\hat{\sigma}_n^2 > \sigma_{an} \rightarrow \mathrm{E}[\log(\hat{s_{ni}}) \mid \log(I_{ni}) - \log(I_{ai}) > 0] > \mu_n + \mu_\eta$

So, the best non-agricultural workers go to sector n

$\hat{\sigma}_a^2 < \sigma_{an} \rightarrow \mathrm{E}[\log(\hat{s_{ai}}) \mid \log(I_{ni}) - \log(I_{ai}) < 0] < \mu_a + \mu_\varphi$

So, the best agricultural workers go to sector n

$\therefore$ only worst skilled in sector n goes to sector a

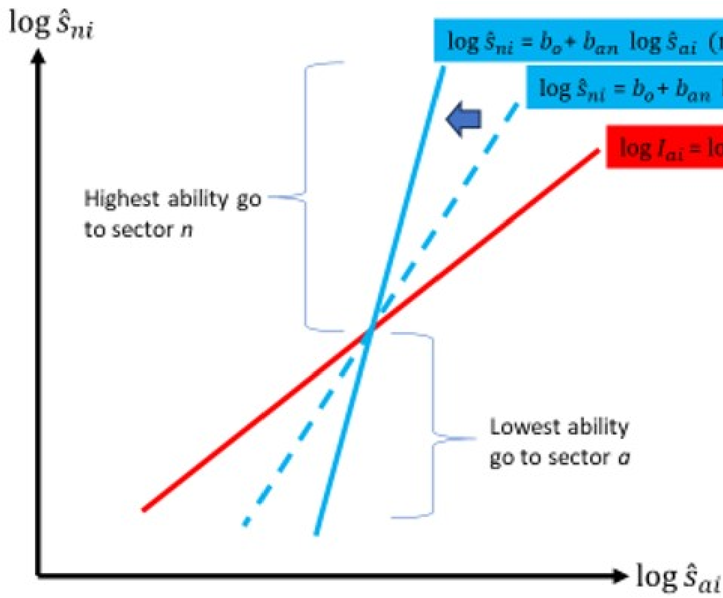

Figure 1.1: Case of Absolute Advantage with high dispersion in skills in non-agriculture

where

$\mathrm{E}[\hat{U}_{ni} \mid \hat{U}_{ai}] = \frac{COV(\hat{U}_{ni},\hat{U}_{ai})}{VAR(\hat{U}_{ai})} \cdot (\hat{U}_{ai})$

$\hat{U}_{ni} = \underbrace{\frac{COV(\hat{U}_{ni},\hat{U}_{ai})}{VAR(\hat{U}_{ai})}}_{b_{an}} \hat{U}_{ai} + \epsilon$

$log\hat{s}_{ni} = \underbrace{\mu_n + \mu_\eta - b_{an}(\mu_a + \mu_\varphi)}_{b_o} + b_{an}(log\hat{s}_{ai}) + \epsilon$

$log\hat{s}_{ni} = b_o + b_{an}(log\hat{s}_{ai}) + \epsilon \rightarrow$ Correlated Abilities Line

1.7.2 Absolute Advantage with high $\hat{\sigma}_a^2$

$\hat{\sigma}_n^2 < \sigma_{an} \rightarrow \mathrm{E}[\log{(\hat{s}_{ni})} \mid \log(I_{ni}) - \log(I_{ai}) < 0] < \mu_n + \mu_\eta$

best non-agricultural workers go to sector a

$\hat{\sigma}_a^2 > \sigma_{an} \rightarrow \mathrm{E}[\log{(\hat{s}_{ai})} \mid \log(I_{ni}) - \log(I_{ai}) < 0] > \mu_a + \mu_\varphi$

best agricultural workers go to sector a

$\therefore$ only worst skilled in sector a goes to sector n

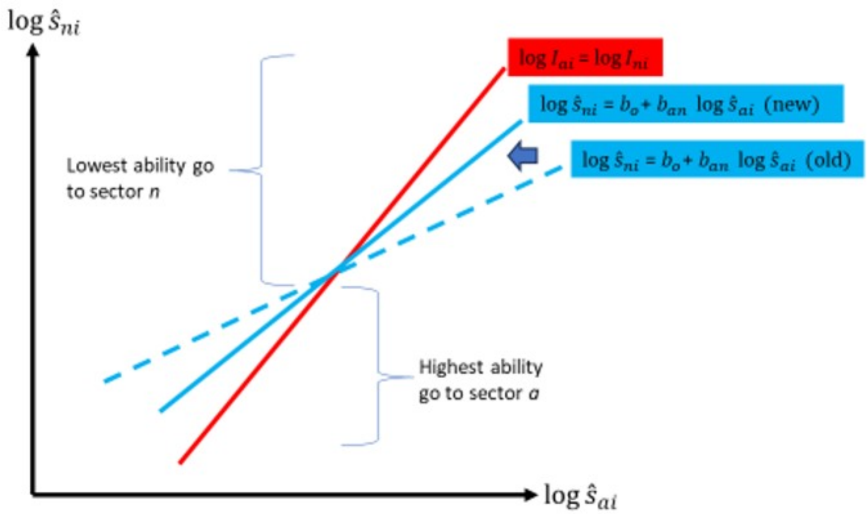

Figure 1.2: Case of Absolute Advantage with high dispersion in skills in agriculture

where

$\mathrm{E}[\hat{U}_{ni} \mid \hat{U}_{ai}] = \frac{COV(\hat{U}_{ni},\hat{U}_{ai})}{VAR(\hat{U}_{ai})} \cdot (\hat{U}_{ai})$

$$\hat{U}_{ni} = \underbrace{\frac{COV(\hat{U}_{ni},\hat{U}_{ai})}{VAR(\hat{U}_{ai})}}_{b_{an}} \hat{U}_{ai} + \epsilon$$

$$log\hat{s}_{ni} = \underbrace{\mu_n + \mu_\eta - b_{an}(\mu_a + \mu_\varphi)}_{b_o} + b_{an}(log\hat{s}_{ai}) + \epsilon$$

$log\hat{s}_{ni} = b_o + b_{an}(log\hat{s}_{ai}) + \epsilon \rightarrow$ Correlated Abilities Line

1.7.3 Comparative Advantage

$\hat{\sigma}_n^2 > \sigma_{an} \rightarrow \mathrm{E}[\log(\hat{s_{ni}}) \mid \log(I_{ni}) - \log(I_{ai}) > 0] > \mu_n + \mu_\eta$

best non-agricultural workers go to sector n

$\hat{\sigma}_a^2 > \sigma_{an} \rightarrow \mathrm{E}[\log(\hat{s_{ai}}) \mid \log(I_{ni}) - \log(I_{ai}) < 0] > \mu_a + \mu_\varphi$

best agricultural workers go to sector a

$\therefore$ best go to sector a and best go to sector n

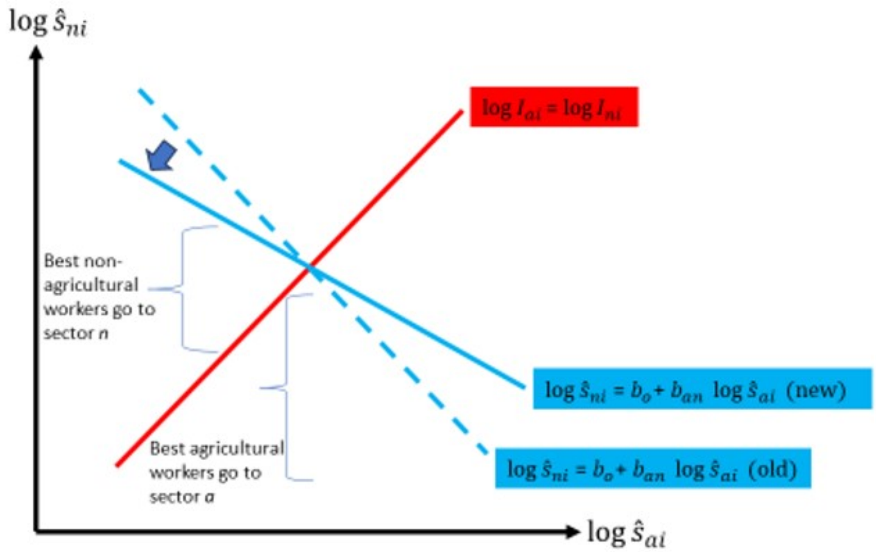

Figure 1.3: Case of Comparative Advantage

where

$\mathrm{E}[\hat{U}_{ni} \mid \hat{U}_{ai}] = \frac{COV(\hat{U}_{ni},\hat{U}_{ai})}{VAR(\hat{U}_{ai})} \cdot (\hat{U}_{ai})$

$$\hat{U}_{ni} = \underbrace{\frac{COV(\hat{U}_{ni},\hat{U}_{ai})}{VAR(\hat{U}_{ai})}}_{b_{an}} \hat{U}_{ai} + \epsilon$$

$$log\hat{s}_{ni} = \underbrace{\mu_n + \mu_\eta - b_{an}(\mu_a + \mu_\varphi)}_{b_o} + b_{an}(log\hat{s}_{ai}) + \epsilon$$

$log\hat{s}_{ni} = b_o + b_{an}(log\hat{s}_{ai}) + \epsilon \rightarrow$ Correlated Abilities Line

1.8 Computation Strategy

As in Adamopoulos et al. (2022), the strategy here too, is to calibrate labor mobility barriers, sector specific abilities, and subsequent allocation of labor in a Benchmark Economy (BE) to the household-level data from Honduras. My calibration strategy involves a two step procedure adjusted to my setup with transport costs. Firstly, I derive population moments on abilities and labor mobility barriers from observed moments on sectoral incomes and idiosyncratic transport costs. Secondly, using the inferred population moments from the aforementioned step, I solve for the remaining parameters from the GE equations of my model computed to match key targets in the data for the economy in Honduras.

1.8.1 Inferring Population Moments from Observed Ones

I assume a multivariate log-normal distribution for the joint population distribution of abilities and sector-specific transport distortions. I first back out the moments of the distribution (i.e., means, variances, and covariances), and then I will match the observed moments on sectoral incomes and the transportation distortions (φ_i, η_i). There are eleven population moments that need to be calibrated: four means $(\mu_a, \mu_n, \mu_\varphi, \mu_\eta)$; four variances $(\sigma_a^2, \sigma_n^2, \sigma_\varphi^2, \sigma_\eta^2)$; and three covariances $(\sigma_{a\varphi}, \sigma_{n\eta}, \sigma_{an})$. The means, variances and covariances are important as they inherently control sectoral selection patterns.

In order to calibrate the population moments of the variance-covariance matrix, I followed the strategy of Adamopoulos et al. (2022). First, I derive in the model the conditional moments as functions of the population moments. Second, I compute the empirical moments in my micro data corresponding to the conditional moments in the model. Third, I solve the system of equations of the conditional moments to back out the population moments. Different from Adamopoulos et al. (2022) who had moments on idiosyncratic distortions specific to the agricultural sector, I have moments on idiosyncratic transport distortions specific to the agricultural sector and the non-agricultural sector. Below I show the system of equations on conditional moments using log normality:

1. Variance of log income in agriculture conditional on choosing agricultural sector

   $VAR\{\log(I_{ai})|\log(I_{ai}) > \log(I_{ni})\} = \hat{v}_a = \hat{\sigma}_{aa}[1 - (\frac{\sigma_{an} - \hat{\sigma_a}}{\sigma\hat{\sigma}_a})^2 \lambda^-(b_1)(\lambda^-(b_1) - b_1)]$

2. Variance of log income in non-agriculture conditional on choosing non-agriculture

   $VAR\{\log(I_{ni})|\log(I_{ni}) > \log(I_{ai})\} = \hat{v}_n = \sigma_{nn}[1 - (\frac{\hat{\sigma}_{nn} - \sigma_{an}}{\sigma\hat{\sigma}_n})^2 \lambda^+(b_1)(\lambda^+(b_1) - b_1)]$

3. Covariance of log incomes in sectors a and n conditional on choosing agricultural sector

$COV\{\log(I_{ai}), \log(I_{ni})| \log(I_{ai}) > (I_{ni})\} = c_{an} = \sigma_{an} - (\frac{\sigma_{an}-\hat{\sigma}_{nn}}{\sigma^*})(\frac{\hat{\sigma}_{aa}-\sigma_{an}}{\sigma^*})\lambda^-(b_1)[\lambda^-(b_1) - b_1])$

4. Variance of log-distortions in transportation in agriculture conditional on choosing agriculture
   $VAR[\log(\varphi_i)| \log(I_{ai}) > \log(I_{ni})] = v_\varphi = (\sigma_{\varphi\varphi})[1 - (\frac{\sigma_{\varphi\varphi}+\sigma_{a\varphi}}{\sigma\sigma_\varphi})^2\lambda^-(b_1)(\lambda^-(b_1) - b_1)]$

5. Variance of log-distortions in transportation in non-agriculture conditional on choosing non-agriculture
   $VAR[\log(\eta_i)| \log(I_{ni}) > \log(I_{ai})] = v_\eta = (\sigma_{\eta\eta})[1 - (\frac{\sigma_{\eta\eta}+\sigma_{n\eta}}{\sigma\sigma_\eta})^2\lambda^+(b_1)(\lambda^+(b_1) - b_1)]$

6. Covariance of log agricultural income and log-distortions in transportation in agriculture conditional on choosing agriculture
   $COV[\log(I_{ai}), \log(\varphi_i)| \log(I_{ai}) > (I_{ni})] = c_{a\varphi} = (\sigma_{a\varphi} + \sigma_{\varphi\varphi})(\frac{\sigma_{an}-\hat{\sigma}_{aa}}{\sigma^*})(\frac{\sigma_{\varphi\varphi}+\sigma_{a\varphi}}{\sigma^*})\lambda^-(b_1)[\lambda^-(b_1) - b_1]$

7. Covariance of log non-agricultural income and log-distortions in transportation in non-agriculture conditional on choosing non-agricultural sector
   $COV[\log(I_{ni}), \log(\eta_i)| \log(I_{ni}) > (I_{ai})] = c_{n\eta} = (\sigma_{n\eta} + \sigma_{\eta\eta})(\frac{\sigma_{an}-\hat{\sigma}_{nn}}{\sigma^*})(\frac{\sigma_{\eta\eta}+\sigma_{n\eta}}{\sigma^*})\lambda^+(b_1)[\lambda^+(b_1) - b_1]$

where the variance of effective abilities are:

$$\hat{\sigma_a}^2 = \sigma_a^2 + \sigma_\varphi^2 + 2\sigma_{a\varphi}$$

$$\hat{\sigma_n}^2 = \sigma_n^2 + \sigma_\eta^2 + 2\sigma_{n\eta}$$

Targeted Moments	**Value**
N_a [10]	0.34
Correlation across both incomes	0.15
Correlation across I_n and transport costs η among those working in sector n	0.02
Correlation across I_a and transport costs φ among those working in sector a	-0.02
STD of transport distortions in sector n	0.48
STD of transport distortions in sector a	0.44
STD of I_a	0.66
STD of I_n	0.58

Table 1.2: Targeted Empirical Conditional Moments From Micro Data in Honduras Prior to Transport Project

Normalizing the means, μ_a, μ_a, μ_φ and μ_η all to 0, I will subsequently have seven population moments required to be calibrated. In my system of equations under Section 8.1, these seven moments to be calibrated are distinguished by equations 1-7– the seven conditional moments of variances and covariances. From the micro data in Honduras, the empirical conditional variances, correlations and share of labor in sector a that I target are found in Table 1.2.

1.8.2 Calibrated Population Moments - Benchmark Economy

Population Moments	**Value**
Correlation of abilities across sectors ρ_{an}	-0.31
Correlation of ability in sector n and transport distortion $\rho_{n\eta}$	-0.56
Correlation of ability in sector a and transport distortion $\rho_{a\varphi}$	-0.42
STD of ability in agriculture σ_a	1.12
STD of transport distortion in non-agriculture σ_η	0.48
STD of transport distortion in agriculture σ_φ	0.44
STD of ability in non-agriculture σ_n	0.85

Table 1.3: Calibrated Moments From Micro Data in Honduras

[10] According to ILOSTAT database, share of employment in agriculture in 2008 was 34.267 percent. Subsequently, the share of employment in non-agriculture in 2008 will be 66 percent(ILO, 2020)

Table 1.3 shows these calibrated population moments. Before interpreting the numbers, it is important to highlight what theory says. When households differ in their productivity-levels specific to each sector, then what theory says will determine if they work in one sector or both, comes down to comparative advantage (Alvarez-Cuadrado et al., 2019).

ρ_{an} of -0.31 means that households here are sorting according to comparative advantage, as $\rho_{an} < 0$. σ_a and σ_n, the dispersion of sector specific abilities are both high, and both higher than σ_{an}, suggesting comparative advantage too. Together, it means that skilled agricultural workers tend to be less good in non-agricultural jobs and skilled non-agricultural workers tend to be less good in agricultural jobs. Now looking at the correlation between abilities in sector n and transport frictions η, the correlation across the two is negative. The number practically says that the best non-agricultural workers tend to face higher transport costs. Similarly, when looking at the correlation between abilities in sector a and transport frictions φ, the correlation across the two is negative. The number also practically says that the best agricultural workers tend to face higher transport costs. Reason being is recall the set-up of the transport distortion is the inverse of the actual commute time. Therefore, if you face a lower transport distortion (φ or η), you thus see a higher travel time. It is also important to note that I am not imposing this, but am allowing the data to speak and this is coming out of the micro data.

Next, I report the remaining calibrated parameters which were found by using correlated data of 1000000 triplets of $(s_{ai}, s_{ni}, \varphi_i, \eta_i)$ drawn from a multi-variate log normal distribution, using the calibrated population moments from Table 1.3. A_n is set to 1 and ϕ is set to 0.01, which leaves $\bar{a}$, A_a and ε to be determined. These three remaining parameters are found by solving for equilibrium in my model to match the following moments: N_a is set to 0.34, p_a is normalized to 1 and the ratio of labor productivity in sector n to sector a is set at 3.76 to match 2008 aggregate stats for Honduras. Table 1.4 reports these remaining calibrated parameters and Table 1.5 reports the value of key statistics in my simulated benchmark economy.

Parameters	**Value**
$\bar{a}$	0.14
A_a	0.13
ε	0.77

Table 1.4: Calibrated Remaining Parameters

Statistics	**Value**
Y_a/N_a	0.43
N_a	0.34
Y_n/N_n	1.61
$(Y_n/N_n)/(Y_a/N_a)$	3.76
Z_a/N_a	3.80
Z_n/N_n	1.80
(Y/N)	1.21
$(Z_n/N_n)/(Z_a/N_a)$	0.47

Table 1.5: Benchmark Economy

1.9 Quantitative Counterfactual Experiment

My main counterfactual experiment looks at the aggregate effects of eliminating idiosyncratic transport distortions specific to both sectors. To do this, I set $\varphi_i = 1$ and $\eta_i = 1$. Given that the average commute times are already set to 0 in both sectors, eliminating transport costs involves eliminating variances and covariances and equalizing them. The results of my experiment are shown in Table 1.6. As is shown in the last column, seeing the removal of transport distortions to the agricultural and non-agricultural sectors, sees important effects. That is, essentially equalizing transport costs across everyone leads to a significant impact on the economy in Honduras.

Specifically, looking at the table, the share of employment in the agricultural sector decreases by 8 percentage points, from 34 percent to 26 percent. The numbers indicate that more people are leaving the agricultural sector with the removal of transport costs. Eliminating transport distortions increases labor productivity in the agricultural sector as it motivates those with higher ability in the agricultural sector to sort back into sector a. Thus, average ability in sector a, also sees an increase too. That is, agricultural productivity saw an increase with the improved selection of workers.

With respect to the non-agricultural sector, average ability and productivity of labor move in opposite directions where average ability sees a slight decrease in the non-agricultural sector, and labor productivity sees an increase with the elimination of the transport distortions. Real GDP per worker also sees an increase, given that aggregate productivity increased in both sectors and there was a movement of labor from a low to a high productivity sector (non-agriculture) compared to what my benchmark economy looked like.

Aggregate Stats	**BE**	**No Transport Distortions** ($\varphi_i = 1$, $\eta_i = 1$)
Y_a/N_a	1.00	1.32
N_a	0.34	0.26
Y_n/N_n	1.00	1.08
$(Y_n/N_n)/(Y_a/N_a)$	1.00	0.82
Z_a/N_a	1.00	1.18
Z_n/N_n	1.00	0.97
(Y/N)	1.00	1.19
$(Z_n/N_n)/(Z_a/N_a)$	1.00	0.82

Table 1.6: Impact of eliminating all transport costs

My second counterfactual experiment looks at the aggregate effects of eliminating idiosyncratic transport distortions specific to the agricultural sector. To do this, I set $\varphi_i = 1$ only. The results of my experiment are shown in Table 1.7. Specifically, looking at the table, the share of employment in the agricultural sector decreases by 9 percentage points, from 34 percent to 25 percent. The numbers indicate that more people are leaving the agricultural sector with the removal of transport costs to the agricultural sector. Eliminating transport distortions increases labor productivity in the agricultural sector as it motivates those with higher ability in the agricultural sector to sort back into sector a, given that the frictions to labor mobility to this sector are removed. Also, average ability in sector a, also sees an increase too. With respect to the non-agricultural sector, average ability and productivity of labor go down with the elimination of the transport distortions to the agricultural sector, given that there is an increase in those working in non-agriculture now, albeit whose productivity is lower then the existing pool of workers in non-agriculture. Real GDP per worker sees an increase.

Aggregate Stats	**BE**	**No Agri Transport Distortions** ($\varphi_i = 1$)
Y_a/N_a	1.00	1.40
N_a	0.34	0.25
Y_n/N_n	1.00	0.93
$(Y_n/N_n)/(Y_a/N_a)$	1.00	0.67
Z_a/N_a	1.00	1.25
Z_n/N_n	1.00	0.93
(Y/N)	1.00	1.06
$(Z_n/N_n)/(Z_a/N_a)$	1.00	0.75

Table 1.7: Impact of eliminating transport costs in agriculture

My third counterfactual experiment looks at the aggregate effects of eliminating idiosyncratic transport distortions specific to the non-agricultural sector. To do this, I set $\eta_i = 1$ only. The results of my experiment are shown in Table 1.8. Specifically, looking at the table, the share of employment in the agricultural sector increases by 2 percentage points, from 34 percent to 36 percent. The numbers indicate that more people are entering the agricultural sector with the removal of transport costs to the non-agricultural sector. Eliminating transport distortions decreases labor productivity in the agricultural sector as it motivates those with higher ability in the non-agricultural sector to sort back into sector n given the removal of labor mobility frictions to that sector. Average ability in sector a, also sees a decrease too, because those sorting in are of lower average ability. With respect to the non-agricultural sector, average ability and productivity of labor go up with the elimination of the transport distortions to the non-agricultural sector. Real GDP per worker also sees an increase.

Aggregate Stats	**BE**	**No Nagri Transport Distortions** ($\eta_i = 1$)
Y_a/N_a	1.00	0.95
N_a	0.34	0.36
Y_n/N_n	1.00	1.18
$(Y_n/N_n)/(Y_a/N_a)$	1.00	1.24
Z_a/N_a	1.00	0.95
Z_n/N_n	1.00	1.05
(Y/N)	1.00	1.13
$(Z_n/N_n)/(Z_a/N_a)$	1.00	1.10

Table 1.8: Impact of eliminating transport costs in non-agriculture

1.10 Conclusion

This paper studied how idiosyncratic transportation frictions to the agricultural and non-agricultural sectors affected the sectoral sorting of workers and their impact on aggregate productivity in Honduras. To do this, I introduced the friction of idiosyncratic transport costs to both sectors and I quantified it. Next, I calibrated my model to the 2008 makeup of Honduras' economy and my main quantitative experiment saw transport distortions for everyone eliminated. This all in an effort to evaluate the quantitative effect observed transport costs had on the pattern of occupational choices, selection and aggregate productivity in Honduras.

The quantitative results from my main counterfactual experiment of removing transport distortions to both sectors had important effects on the economy in Honduras. Removing transport distortions saw a decrease in the share of employment in the agricultural sector by 8 percentage points, and also increases were seen in labor productivity and average ability in the agricultural sector. Removing transport distortions saw slight an increase in labor productivity in the non-agricultural sector, but a decrease in average ability. Real GDP per worker also saw an increase with the elimination of the transport distortions.

My work serves to shed light quantitatively on what the policy implications are of improvements to transportation infrastructure, specifically removing labor mobility frictions. That is, if I remove frictions to labor mobility, what happens to questions of sorting and average quality? Practically this could look like improving road quality and access. The results of this paper does echo what is revealed regarding transportation infrastructure, particularly improvements to it. The standing literature finds non-linear effects on improvements to transportation, where the effect to improvements to transport infrastructure is greater to countries, who are not only poorer, but who also need better infrastructure (Adamopoulos, 2011). Future research could work with another country whose economic makeup is not only poorer, but who also requires significant improvements or introductions of transport infrastructure to gage the robustness of my model. My quantitative framework can be used to assess the quantitative importance observed idiosyncratic transportation frictions has on sectoral sorting of workers and their impact on aggregate productivity among other countries too.

Chapter 2

Transportation Costs as a Source of Resource Allocation: A Quantitative Analysis with Data from the El Salvador Connectivity Project

2.1 Introduction

A key issue in economic growth and development is understanding what factors are responsible for the wide differences in productivity and living standards across the world (Restuccia and Rogerson, 2017). Economic development is directly linked to the agricultural sector. Lower income countries (LICs) have much lower productivity in the agricultural sector and devote much more labor to producing agricultural goods than developed economies (Restuccia et al., 2008). It is thus of no surprise that seeing improvements in agriculture has historically been regarded as paramount to realizing development goals (Gollin et al., 2002).

The fundamental question is why is agricultural productivity so low in LICs? One prominent theory that has received considerable attention in the recent literature is the misallocation of resources within agricultural sectors of LICs (Adamopoulos and Restuccia, 2014). According to this theory there are policy distortions, and market frictions that introduce heterogeneity in the prices faced by individuals producers altering the allocation of factors of production among them (e.g., labour, capital, land) relative to what is efficient (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008). This misallocation of factors can artificially increase the size of unproductive producers and reduce the size of productive producers reducing aggregate productivity and income.

I examine the role that investments in transportation infrastructure play for productivity in the agricultural sector. The standing literature indicates that lower income nations have low transport network density compared to higher incomes nations (Adamopoulos, 2011). On top of that, there is evidence that there is a huge dispersion in transport costs seen in lower income countries–Adamopoulos (2024) finds that costs of transport are very different across different parts of Ethiopia. There is also evidence indicating that farmers vary in how much access they have to a transportation network and crop markets (Adamopoulos, 2024). I examine in my research how this heterogeneity in access affects how resources are allocated across spatially dispersed farmers. I quantitatively examine this question by leveraging a transport infrastructure program in El Salvador.

In particular, I will use micro-level detailed data from the El Salvador Connectivity Project which was an infrastructure venture that targeted the Northern Transnational Highway. The detailed micro data contains five rounds of survey data from 2009 to 2014 –one baseline (2009-2010) and four follow-ups. It contains information on commute times for farmers to local markets for sale of goods produced, types of crops that were harvested by farmers, inputs used in the harvesting process, pricing, among others (Social Impact Inc., 2015). To assess the effects of the El Salvador Connectivity Project on agricultural productivity, I will combine the micro data with a theoretical structural model in which idiosyncratic transport costs operate as output "taxes," for heterogeneous farmers.

Particularly, I will ask the following, *"To what extent does heterogeneity in transporta-*

tion costs in spatially dispersed farms account for measured misallocation in agriculture?" The idea is that there is a lot of evidence of misallocation in agriculture and one wants to understand what underlies it. Could part of it be driven by transport costs? Here I have farm-specific data on transport costs at the farm level and this allows me to explore this. To do this, I introduce idiosyncratic transport costs at the farm level in an otherwise standard model of misallocation. Next, I combine micro panel level data on inputs, outputs and commute times for farmers to local markets. Finally, I quantify the contribution of observed transport costs to measured misallocation in agriculture in El Salvador.

My work contributes to a better understanding of the sources of low agricultural productivity in developing countries. If the observed allocation of resources is primarily driven by a feature of the spatial economic environment such as transport infrastructure then investments that not only improve the level but also reduce the dispersion of access for farmers would raise productivity. Interestingly, in this case the observed resource allocation would be inefficient, conditional on the transport infrastructure network. What I find is that transportation costs matter for observed misallocation in El Salvador and a hypothetical social planner should be accounting for them.

The chapter is organized as follows. Section 2 will address what the literature says on the topic in question. Section 3 will discuss the data from El Salvador. Section 4 will outline the model. Section 5 will discuss the quantitative analysis followed by the results. Section 6 looks at robustness. Section 7 concludes.

2.2 Literature Review

This paper contributes to a growing literature on misallocation (see Restuccia and Rogerson (2008); Hsieh and Klenow (2009); Adamopoulos and Restuccia (2014), among others). Around the discussion of misallocation in agriculture, the literature has primarily focused on policies, institutions, and frictions in land markets that reallocate land from large productive to small unproductive farmers (Gottlieb and Grobovšek, 2019; Restuccia and Santaeulalia-Llopis, 2017; Adamopoulos and Restuccia, 2020, 2014; Adamopoulos et al., 2022; Chari et al., 2017; Chen et al., 2023). Gottlieb and Grobovšek (2019) for example look at the consequences provisions of communal land have on questions of resource allocation in Sub-Saharan Africa. Chari et al. (2017) look at the implications transformed leasing rights had on how land was allocated in rural China. Adamopoulos and Restuccia (2020) look at the consequences the Philippines' 1988 land reform had on the inappropriate allocation of resources among farmers. I take a different approach to the literature and instead study if the existing measured misallocation could be partly spurred by heterogeneity in transportation costs among farmers who are spatially scattered.

More specifically, my paper is closely related to questions of potential measurement error or mismeasurement in the discussion of misallocation in the agricultural sector (Gollin and Udry (2021); Dong and Hsieh (2021); Bils et al. (2021), just to mention some). Using manufacturing data from Korea, Dong and Hsieh (2021) document the extent and consequences of measurement error in capital measures. They find that measured misallocation is misrepresented by 10 percentage points. Using manufacturing data from the U.S. and India, Bils et al. (2021) document the extent and consequences of mismeasurement in measures of revenue and inputs. For both countries (starkly more so for the U.S.), they find potential gains from reallocation shrink substantially when measurement error is corrected for.

When discussing questions of factor misallocation in agriculture, and appropriately measuring it, it is important to note that attempts to account for features of the physical environment, like that of potential differences related to the quality of the land, have been done before (Gollin and Udry, 2021; Ayerst et al., 2020; Chen et al., 2023). Ayerst et al. (2020), for example, look at questions of misallocation in Vietnam from 2006 to 2016 and discuss the implications of it. In their work, they do address a possible complication that their measure of misallocation may be overestimating gains from reallocation, given that a feature of the physical environment may be getting picked up–that of, land quality. They circumvent this potential problem by calculating two new measures of farmer productivity and land, purged of the possible effect associated with variability in the quality of land[1]. However, to my knowledge, there is no previous paper which effectively looks at a particular feature of the spatial physical environment, like that of idiosyncratic transportation costs, as a source of resource allocation, as I do here. It is also important to note that although we both are addressing potential differences related to the physical environment, the issue of land quality matters for production, whereas the issue of transportation infrastructure matters post-production, when heading to the market. For that reason, how we handle the adjustments to our measures of misallocation are effectively not the same.

Restuccia and Rogerson (2017) have highlighted that the avenue that is classically used when it comes to not only measuring the root of misallocation, but also seeing its respective quantitative impacts evaluated is through a structural model. They also highlight that this method in measuring misallocation has seen extended use around the discussion of taxes, particularly different rates of taxation ultimately giving rise to distortions in inputs. In other words, one can take the case of heterogeneous production units who see the production of a homogeneous good, but who all face variable rates of taxation–this would give rise to misallocation (Restuccia and Rogerson, 2017). One can also interpret the tax and treat it as a wedge/price. Therefore, when it comes to understanding the discussion of misallocation rising from a varied tax, the idea is the following– these variable rates of taxation/wedges/prices that heterogeneous producers of the same good are met with give rise to an allocation, and this said allocation is inefficient from the equilibrium perspective (Restuccia, 2016). According to Restuccia and Rogerson (2008), idiosyncratic distortions are distortions that can vary across production units and subsequently see resources across plants reallocated. In this research project, I examine to what extent the heterogeneity in transportation costs for farmers in accessing crop markets can be responsible for the allocation of resources across farmers within agriculture through a structural model. Particularly, I quantify their contri-

[1]Authors do also address potential issues of labor quality.

bution to measured misallocation in agriculture in El Salvador using the classic tax scheme which will be further discussed in the quantitative section of my paper.

2.3 Data

2.3.1 El Salvador Connectivity Project

In brief, near the closing of 2006, the Millennium Challenge Corporation (MCC) and Salvadoran government (GOES) retained a five-year compact worth about $461 million whose goal was twofold: (1) encourage economic growth and (2) see the reduction of poverty through tactical investments in certain projects. This compact which ran from September 2007 up until September 2012 consisted of three projects which targeted the Northern Zone of the country, one of which was the Connectivity Project. The inherent aim of the Connectivity Project was to respond to the physical inaccessibility Northern Zone Salvadorans faced to the rest of the country given their remote position. In doing so, the hope was to see the betterment of the lives of those living in the Northern Zone via improved linkages to the rest of El Salvador, markets, schools, etc. Although the scope of the project began with two ideas in mind, the project ended up only focusing efforts on one main activity—the construction and rehabilitation of the Northern Transnational Highway (MCC, 2018).

I use micro-level detailed data from this infrastructure venture program that targeted the Northern Transnational Highway in El Salvador. There are five rounds of survey data from 2009 to 2014 which includes one baseline round (2009-2010) and four follow-ups rounds. The detailed micro-level data contains information on commute times for farmers to local markets for sale of goods produced, types of crops produced, inputs used, just to mention some[2] (Social Impact Inc., 2015).

I use first (baseline) and last available rounds of data. To compute total production of a household, I use gross real output measured in quintals. It is the basket of harvested

[2]For more information on the data, please see Social Impact Inc. (2015).

crops by the household farm i in location j, aggregated using a common price (average price per quintal of crop sold at market). Next, I focus on the inputs. I measure labor input as the the aggregated weekly total labor hours worked by everyone on all parcels of land (not including rested land)[3]. I measure land input as the aggregated size of land in manzanas[4] (this includes only land that was used in the harvest in the last 12 months). Finally, to measure intermediate inputs, I aggregated the total of both chemical and organic fertilizers applied on the parcels of land in quintals.

Following the literature aggregation is at the farm level and not plot level (Aragón et al., 2022). In addition, it is important to note, if a household farm had at least one crop that was not convertible to the equivalent quintal measure (i.e., in units, boxes, etc.), those households were dropped. Households were also dropped if they had reported the harvest of a crop labelled as "other." Households were also dropped if at least one parcel of land was set as rested (or not used in the harvest of the crops). Households were also dropped if they had neither use of chemical or organic fertilizer or if they had no labor reported. The reason for this was to try to ensure absence of measurement error or mismeasurement given the aggregation of all terms. Chemical fertilizers were reported as either pounds, quintals or litres. I converted all units to quintals, but due to absent information on what type of chemical fertilizers were used, particularly with respect to the liquid fertilizers, I used the assumption of the best possible fertilizer use in crop production of Urea ammonium nitrate (UAN) 28-0-0 nitrogen liquid fertilizer (Pattison, 2018), in order to convert litres to quintals[5].

[3]For the last round of data, labor reported was from high season.

[4]This is equivalent to about 0.7 hectares.

[5]For each variable, author recoded to missing if equal to 99, 999,9999,99999,99999.99, or 9999999.99.

2.4 Model

2.4.1 Environment

I write down a static, one sector, general equilibrium model. The economy produces one agricultural good and agricultural production occurs in the rural sector. The representative household is endowed with labor N which is inelastically supplied to the market. The representative household also owns productive land. Production of the agricultural good is done by heterogeneous farms. Finally, there are no exports or imports and no selection (no decision to either be a farm worker or manager). One can think thus of my model as the Lucas 1978 no-occupational choice version model.

2.4.2 Production of Agricultural Good

Household farm i with ability s_{ij} produces in accordance to the following production function with technology that exhibits decreasing returns to scale γ:

$$y_i = (s_i)^{1-\gamma}(l_i^{\alpha} n_i^{\beta} x_i^{\delta})^{\gamma}$$

y_i denotes real gross output, n_i is labor, l_i is land and x_i is the intermediate inputs. Finally, s_i is the farm productivity which is heterogeneous and γ denotes returns to scale[6]. α, β and δ are the input elasticities for each input respectively.

2.4.3 Estimated Transportation Cost

I now introduce the relevant transport cost piece of my paper. Here, delivery of the crop from rural locations of household farm i to local market for sale, is subject to origin specific

[6] Being bound by the data, I only had access to capital information for the baseline data and for that reason capital was excluded as a relevant input factor to production. Either way, given that the baseline and last round of data are only five years apart, inclusion of capital if available wouldn't have made a huge difference as generally household's capital holdings don't tend to change very much in the short term. I could have taken the baseline capital holdings of each household farm and used it in the endline data, but under a fixed effects approach of production function estimation, variables would have been dropped out.

transport costs of the iceberg-type, given by the following:

$$T_i = 1 + \psi \times (averagecommute^{\lambda})$$

The idea is that to sell a crop to the local market, farms located in different locations have to deliver T_i units of the crop. Different household farms i face different T_i and if $T_i > 1$, then one can think of this as a tax. The relative price of the crop p received by the farm at the market is the same regardless of origin. In other words, the crop produced in one location versus another are perfect substitutes. Therefore, the idiosyncratic iceberg transport cost T_i implies that the farm-gate price of output will look different for everyone for the same crop produced.

It is important to note that the commute times (in minutes) for household farm i to local markets for sale is set as an average. What this means is that for households who went to the local market to sell (report a non-missing commute time), I took the average travel time across all crops that were delivered to the market for sale. For example, if a household farm harvested three crops in the last twelve months, and only two were delivered to the market for sale, I took average travel time for those two crops and set it as the average measure of travel time for that respective household. This mapping of average commute times to transportation costs follows work by Adamopoulos (2024), where T_i is the estimated iceberg transport cost specific to each household, λ is an elasticity parameter that measures how sensitive transport costs are to commute times and lastly ψ is a scaling parameter designed to keep transport costs close to 1.

Here, instead of imposing a value for λ, I take advantage of the fact that in the panel, particularly in the last round of data, in addition to the commute time, I also have a cost per trip. Similar to how the commute times are handled, I too take the average cost. Next, I simply regress the log average cost per trip on the log average commute time, and the coefficient is set as λ, which here is 0.34. Finally, ψ is set at 0.0525 to match a transport cost share of about 17 percent.

2.4.4 Measuring Farm Productivity

Given that the essence of my paper is to hone in on questions of misallocation with and without including a feature of the spatial economic environment (which in my case is transport infrastructure), outlining how farm productivity is measured is important. My measure of farm productivity or TFP is constructed residually from the following equation:

$$TFP_i = (s_i)^{1-\gamma} = \frac{y_i}{(l_i^{\alpha} n_i^{\beta} x_i^{\delta})^{\gamma}}$$

and two rounds of cross-sectional data on gross labor, gross land and gross intermediate inputs (fertilizers) and gross real output aggregated at the household level. With respect to the input elasticities, standard practice is to borrow from the literature which I do here.

Recent work by Bravo-Ureta et al. (2022) see the estimation of production models for the Salvadoran farm sector. The authors make use of data from the Encuesta Nacional Agropecuaria de Propositos Multiples (ENAPM) which totals six years starting in May 2013 up until April 2019 in El Salvador. Using their preferred model of Random Parameters (RP) the estimated coefficients for a sample of 18,122 households for the inputs that match my farm-level production function. The estimated elasticities are 0.839 for land, 0.098 for labor and 0.186 for fertilizers, all of which are significant at the one percent level. In constructing my measure of farm total factor productivity, I choose $\gamma = 0.98$ (this taken from Table 2 third column), $\alpha = 0.86$, $\beta = 0.1$ and $\delta = 0.19$ to match the estimated coefficients from Bravo-Ureta et al. (2022). The returns to scale parameter was calculated residually using the Levinsohn-Petrin Production Function Estimation Approach[7]. This approach was also used in the robustness of my results which will be discussed later.

[7]This returns to scale value albeit high still exhibits decreasing returns which matters in terms of efficiency gains. As a comparison, Manysheva (2022) using dynamic panel estimation calculates a value of 0.76 for Tanzania.

Given that my measure of TFP can be subject to mismeasurement and shocks, I follow the approach of Adamopoulos et al. (2022). Particularly, I take the computed farm productivity s_{ij} from above by year and household and next estimate a household farm fixed effect purged of both time effects and individual effects. With respect to individual effects, the most disaggregated variable available to the author was municipalities and was thus employed. This household-farm productivity fixed effect measure is used in my efficiency gains experiment.

2.4.5 Definition of Equilibrium

A competitive equilibrium is a set of prices (p, w_i, q_i) and an allocation for household farm i in location j (y_i, n_i, l_i, x_i) where p is the price of the output, w_i is the labor wage and q_i is the rental price of land. The production allocation for household farm i (y_i, n_i, l_i, x_i) solves the farm's profit max problem subject to prices and transport costs (T_i). Lastly, the markets for crops, labor, land and intermediates, all clear.

2.5 Quantitative Analysis

2.5.1 Methodology

According to the literature, when it comes to assessing questions of misallocation, there are two avenues–the direct and indirect approach (Restuccia and Rogerson, 2013). In my work, I will be following the direct approach, and for purposes of this research inquest will only outline what that approach entails. According to Restuccia and Rogerson (2017), the epitome of the direct approach is basically to concentrate on a particular factor which is argued to give rise to questions of misallocation and ultimately see the results of these particular sources evaluated. That is, choose the element, see that it is directly measurable, and afterwards employ a model whose makeup includes heterogeneity in producers. This all in an effort to effectively evaluate the quantitative impact on misallocation (Restuccia and

Rogerson, 2013).

Restuccia and Rogerson (2017) have highlighted that the avenue that is typically used when it comes to not only measuring misallocation, but also assessing its quantitative consequences is through a structural model. I combine micro data with a theoretical structural model where idiosyncratic transport costs operate as "output taxes," for heterogeneous Salvadoran farms. Modelling transport costs in this way matter and prove useful in terms of trade. That is, larger farms and farms that engage more in trade will be met with a higher transport cost. Therefore, the transport cost being proportional to output captures that. One can think of transport costs here as payments to the transport sector. Given that I have data on transport costs at the farm level, I thus measure transport costs for Salvadoran farmers directly. I then use the structure of my model to examine how much dispersion in transport costs for farmers can explain observed resource allocation. My goal is to see to what extent is misallocation explained on the part of idiosyncratic transport costs.

2.5.2 Equilibrium and Identification of Distortions

T_i^Y is the output tax (transport costs) faced by farm i. Given these transport costs, the profit maximization problem for farm i is:

$$\max_{\{l_i, n_i, x_i\}} \{p(1 - T_i^Y)y_i - wn_i - ql_i - p_x x_i\}$$

where FOC with respect to land for household farm i is:

$$\frac{MRPL_i}{\gamma\alpha} = \frac{y_i}{l_i} = \frac{q}{\gamma\alpha(1 - T_i^Y)} \propto \frac{1}{(1 - T_i^Y)}$$

Therefore, given idiosyncratic transport costs, marginal product (and average product) of land is not equalized across household farms and vary in proportion to the idiosyncratic transport costs T_i. Similarly,

$$\frac{MRPN_i}{\gamma\beta} = \frac{y_i}{n_i} = \frac{w}{\gamma\beta(1-T_i^Y)} \propto \frac{1}{(1-T_i^Y)}$$

$$\frac{MRPX_i}{\gamma\delta} = \frac{y_i}{x_i} = \frac{p_x}{\gamma\delta(1-T_i^Y)} \propto \frac{1}{(1-T_i^Y)}$$

Therefore, given idiosyncratic transport costs, marginal product (and average product) of labor and fertilizers are also not equalized across household farms and again vary in proportion to the idiosyncratic transport costs T_i.

2.5.3 Household Farm Profit Max Problem with Transport Cost

Given prices, farm i maximizes profits by choosing labour, land and intermediate inputs (fertilizers) according to the first-order conditions are:

$$\underbrace{(1-T_i)^{\frac{1}{1-\gamma}} s_i}_{z_i}(\gamma p)^{\frac{1}{1-\gamma}}(\frac{\alpha}{q})^{\frac{\alpha\gamma}{1-\gamma}}(\frac{\beta}{w})^{\frac{1-\gamma(1-\beta)}{1-\gamma}}(\frac{1-\alpha-\beta}{p_x})^{\frac{(1-\alpha-\beta)\gamma}{1-\gamma}} = n_i$$

$$\underbrace{(1-T_i)^{\frac{1}{1-\gamma}} s_i}_{z_i}(\gamma p)^{\frac{1}{1-\gamma}}(\frac{\alpha}{q})^{\frac{1-\gamma(1-\alpha)}{1-\gamma}}(\frac{\beta}{w})^{\frac{\gamma\beta}{1-\gamma}}(\frac{1-\alpha-\beta}{p_x})^{\frac{(1-\alpha-\beta)\gamma}{1-\gamma}} = l_i$$

$$\underbrace{(1-T_i)^{\frac{1}{1-\gamma}} s_i}_{z_i}(\gamma p)^{\frac{1}{1-\gamma}}(\frac{\alpha}{q})^{\frac{\gamma\alpha}{1-\gamma}}(\frac{\beta}{w})^{\frac{\gamma\beta}{1-\gamma}}(\frac{1-\alpha-\beta}{p_x})^{\frac{1-\alpha\gamma-\beta\gamma}{1-\gamma}} = x_i$$

$$(1-T_i)^{\frac{\gamma}{1-\gamma}} s_i(\gamma p)^{\frac{\gamma}{1-\gamma}}(\frac{\alpha}{q})^{\frac{\gamma\alpha}{1-\gamma}}(\frac{\beta}{w})^{\frac{\gamma\beta}{1-\gamma}}(\frac{1-\alpha-\beta}{p_x})^{\frac{(1-\alpha-\beta)\gamma}{1-\gamma}} = y_i$$

where

$$\delta = (1-\alpha-\beta)$$

The first order conditions to the household farm suggest that demand for labor, land and demand for fertilizers depend on two things: (1) productivity s_i and (2) distortions specific to each farm, which here are idiosyncratic transport costs. I call this new term z_i or *adjusted* *productivity*. The above also suggests that conditional on productivity s_i, household farms taxed more (face higher transport costs to the local market) command a smaller fraction of output, labor, land, etc. It is important to note that although I do model the iceberg transport cost as an output tax, in practice adjusted productivity z_i is the following:

$$z_i = \frac{s_i}{(T_i)^{\frac{1}{1-\gamma}}}$$

2.5.4 Problem of Social Planner

In an otherwise standard model, the social planner will decide how to allocate various factors of production across heterogeneous producers. The respective allocation decision will thus go in accordance to relative productivity. But in the event that the spatial economic environment like transport costs is accounted for, the problem specifically will now look as follows:

$$\max_{\{l_i,n_i,x_i\}} \sum_{i=1}^{M} y_i$$

subject to

$$(1-T_i)^{\frac{\gamma}{1-\gamma}} s_i(\gamma p)^{\frac{\gamma}{1-\gamma}} (\frac{\alpha}{q})^{\frac{\gamma\alpha}{1-\gamma}} (\frac{\beta}{w})^{\frac{\gamma\beta}{1-\gamma}} (\frac{1-\alpha-\beta}{p_x})^{\frac{(1-\alpha-\beta)\gamma}{1-\gamma}} = y_i$$

and resource constraints

$$L = \sum_{i=1}^{M} l_i \quad N = \sum_{i=1}^{M} n_i \quad X = \sum_{i=1}^{M} x_i$$

where the efficient allocation involves allocating total labor, total land and total fertilizers across heterogeneous household farms according to *relative adjusted productivity* z_i, which here includes the idiosyncratic transport cost component.

$$h_i^e = \frac{z_i}{\sum_{i=1}^{M} z_i} H$$

where $h \in \{\ n,\ l,\ x\ \}$

To see why, take the following example with respect to the land input:

$$\frac{l_i}{L} = \frac{(1-T_i)^{\frac{1}{1-\gamma}} s_i (\gamma p)^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\gamma\beta}{1-\gamma}} \left(\frac{1-\alpha-\beta}{p_x}\right)^{\frac{(1-\alpha-\beta)\gamma}{1-\gamma}}}{\sum_{i=1}^{M} (1-T_i)^{\frac{1}{1-\gamma}} s_i (\gamma p)^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{q}\right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\gamma\beta}{1-\gamma}} \left(\frac{1-\alpha-\beta}{p_x}\right)^{\frac{(1-\alpha-\beta)\gamma}{1-\gamma}}}$$

$$l_i^e = \frac{(1-T_i)^{\frac{1}{1-\gamma}} s_i}{\sum_{i=1}^{M} (1-T_i)^{\frac{1}{1-\gamma}} s_i} L \quad \rightarrow \quad l_i^e = \frac{z_i}{\sum_{i=1}^{M} z_i} L$$

where superscript e represents the efficient allocation. The same idea will hold true for the remaining production inputs. The main takeaway to note is that the allocation discussed in the previous section is identical to what the social planner's is shown here.

As was outlined in the social planner's problem, in an undistorted standard model, total land, labor and fertilizers would have been expected to be allocated according to relative effective productivity. What that means essentially is those who were commanding the most (along production inputs) should be those who are comparatively better farmers or are closer to markets. Graphically, this can be verified in two ways: (1) production inputs and effective farm TFP would see a strong positive relation and (2) as was outlined in the previous section, the marginal products (and/or average products) would be the same across households irrespective of the farm TFP a household possesses.

2.5.5 Misallocation or Mismeasurement?

Following a typical wedges approach, I provide the relationship between my three inputs and TFP from the baseline to look for any evidence of misallocation. I graph the allocation of log land, log labor and log intermediates against the log farm TFP for the first round of cross-sectional data. For each input, I document the input against the farm productivity and the input yield against the farm productivity (again all in logs). The figures indicate evidence of what is said to be "misallocation," for each production input.

Looking at Figure 2.1, where I graph farm size against the farm productivity, it suggests that the input holding in question is virtually unconnected to farm productivity (given by

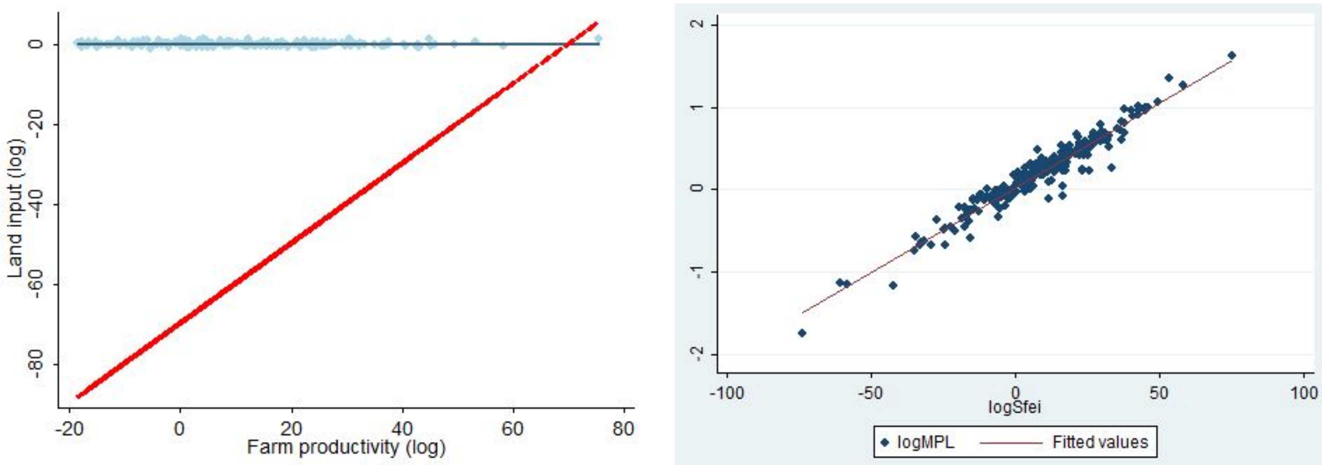

Figure 2.1: *Left*: Farm Size vs. TFP. *Right*: Land Yield vs. TFP

the blue line). What one would have liked to see is an upward sloping line (given by the red line), which would have said that the most productive farmer is commanding the biggest size of land, which is not the case here. One can also have the same takeaway looking at the figure which graphs land yield against farm productivity. One can clearly see that the actual marginal products are not equalized across every household farm. The upward sloping line suggests evidence of wedges (or distortions) giving rise to the dispersions one sees in the marginal (and/or average) products. What one would have liked to see is a flat line, indicating that marginal (and/or average) productivity of the land input is not correlated with farm productivity and inherently the same across everybody. The same sentiment of misallocation can be seen with respect to the remaining production inputs, given by Figure 2.2 and Figure 2.3.

When looking at labor input against the farm productivity in Figure 2.2, it suggests that the labor input is too virtually unconnected to farm productivity (given by the blue line). What one would have liked to see is an upward sloping line (given by the red line), which would have said that the most productive farmer is commanding has more labor, which is not the case here. One can also have the same takeaway looking at the figure which graphs labor yield against farm productivity. One can clearly see that the actual marginal products are not equalized across every household farm. The upward sloping line suggests evidence

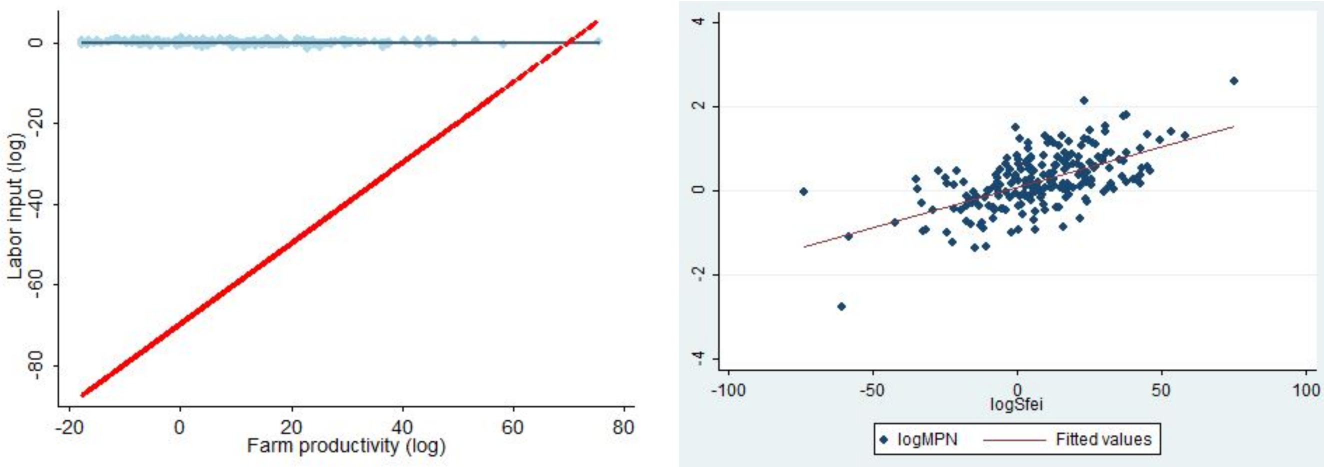

Figure 2.2: *Left*: Labor vs. TFP. *Right*: Labor Yield vs. TFP

of wedges (or distortions) giving rise to the dispersions one sees in the marginal (and/or average) products. What one would have liked to see is a flat line, indicating that marginal (and/or average) productivity of the labor input is not correlated with farm productivity and inherently the same across every household.

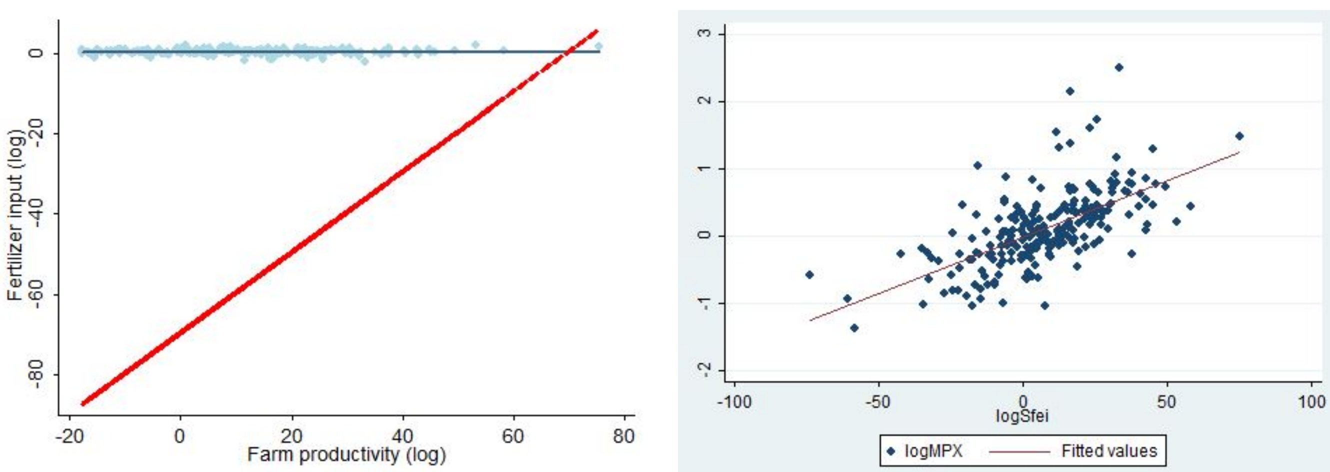

Figure 2.3: *Left*: Fertilizer vs. TFP. *Right*: Fertilizer Yield vs. TFP

When looking at intermediate input against the farm productivity in Figure 2.3, it suggests that the intermediate input is too virtually unconnected to farm productivity (given by the blue line). What one would have liked to see is an upward sloping line (given by the red line), which would have said that the most productive farmer is commanding has the most, which is not the case here. One can also have the same takeaway looking at the figure which graphs intermediate yield against farm productivity. One can clearly see that the actual marginal products are not equalized across every household farm. The upward sloping line suggests evidence of wedges (or distortions) giving rise to the dispersions one sees in the marginal (and/or average) products. What one would have liked to see is a flat line, indicating that marginal (and/or average) productivity of the intermediate input is not correlated with farm productivity and inherently the same across everyone.

Accounting for the previous figures, the "naive" typical approach would call all of this misallocation. Understanding that there is clearly evidence of misallocation being born out of what the literature has coined as wedges (or distortions), my task is to divulge the following– "how much of these wedges are giving rise to *actual* misallocation?" Could it be that what looks like misallocation could merely be picking up a feature of the physical environment, like that of transportation infrastructure? The idea is that, again, if a feature of the spatial economic environment is being picked up, what looks to be misallocation, actually isn't, it is a feature of the physical environment and is not a reason to reallocate resources. The following figure documents the relationship between the ice-berg estimated transport costs and TFP (all in logs). What Figure 2.4 suggests is that transport costs do see a slight negative correlation with farm productivity. More importantly, transportation costs do indicate some variation here and this variability should matter in practice for misallocation and encourages further examination.

2.5.6 Efficiency Gains with Transportation Costs

This brings me to my baseline efficiency gains experiment. Following Adamopoulos et al. (2022), in order to quantify the gains to reallocation, first requires me to calculate what the baseline aggregate output would have been had there been no misallocation, but in the presence of transport costs. Next, the relevant experiment essentially takes the counterfactual estimate, over the sum of all observed output (which includes the residual distortions)

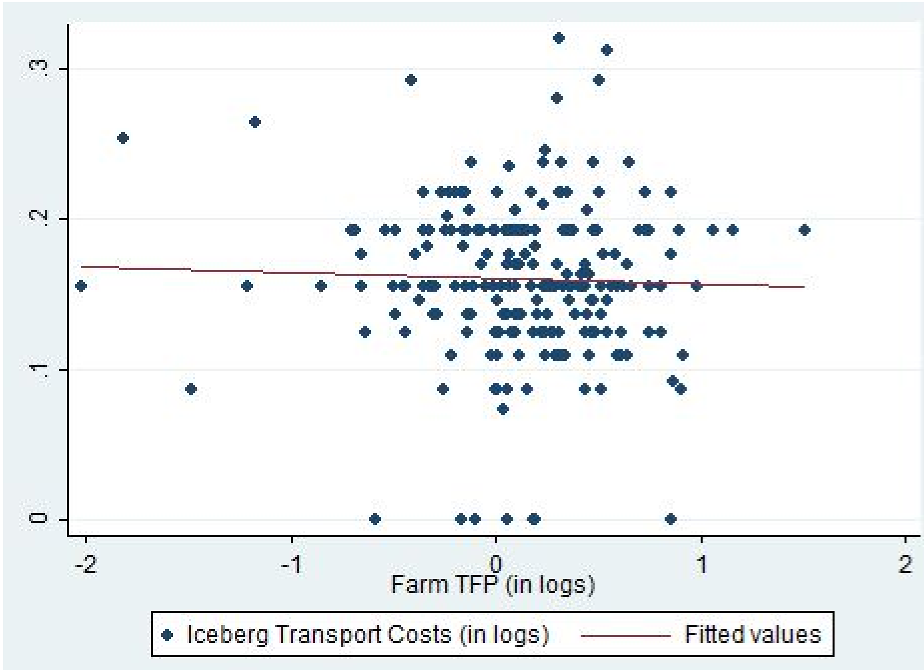

Figure 2.4: Transport Costs vs. Farm Productivity

subtracting one. Table 2.1 outlines the static efficiency gains in baseline.

Y^e/Y^a	s_i	z_i
Static Efficiency Gains	6.47	5.17
Number of Household Farms	226	226

Table 2.1: Agricultural Output (TFP) Gain from Factor Reallocation

Looking at Table 2.1, for the same set of households who reported a travel time to the local market to sell, I distinguish between reallocation based on farm-level productivity s_i as the traditional direct approach does and adjusted productivity z_i which captures the transport infrastructure. The difference between reallocation based on the two measures of productivity is the role of transport costs. Without transport costs, the efficiency gains from removing misallocation is 747 percent. With transport costs, the efficiency gains are 617 percent. Therefore, taking log(1+1.3) / log(1+6.47) shows that transport costs account for over 40 percent of observed measured misallocation. What this suggests is that transport costs matter for misallocation and should be accounted for in the efficiency benchmark, in order to avoid measurement error.

2.6 Robustness

My baseline measure of TFP used in my main experiment borrowed elasticities from Bravo-Ureta et al. (2022) and a fixed effect measure approach from Adamopoulos et al. (2022). Here, I discuss the robustness of my results, and essentially exploit the fact that I have panel data which I use to estimate an alternative production function using the Levinsohn-Petrin (LP) Approach. I then use the alternative farm productivity residually estimated using the LP approach and re-do my efficiency gains experiment.

2.6.1 Estimation Equation and Estimation Method

$$y_t = \beta_0 + \beta_l l_t + \beta_n n_t + \beta_x x_t + \omega_t + \eta_t$$

I estimate the above agricultural production function using the Levinsohn and Petrin (LP) approach (Levinsohn and Petrin, 2003). In brief, in the context of the LP approach I set the state variable as gross land, my free variable as gross labor and my proxy for unobservable productivity shocks as gross fertilizers. The two errors are ω and η, where two conditions hold: (1) ω is known to farmers only and correlated with input choice and (2) η is the error term. The demand for intermediates (fertilizers) depends on state variables (land) and ω. Lastly, the measure of predicted levels of household productivity is given by $\hat{\omega}$, which I use as my *alternative estimate* for log TFP. Table 2.2 shows the Cobb-Douglas Production Function estimates at the household farm level for two rounds of unbalanced panel data[8].

2.6.2 Efficiency Gains with Alternative TFP Measure

Here I repeat the main counterfactual experiment with an estimated measure of TFP using the LP approach. Looking at Table 2.3, for the same set of households who reported a travel time to the local market to sell, I again distinguish between efficiency gains for pro-

[8] Stata code from Rovigatti and Mollisi (2018).

	OLS	FE	LP
Log Gross Labour	0.14***	0.04	0.04***
	(0.02)	(0.02)	(0.01)
Log Gross Land	0.81***	0.60***	0.63***
	(0.02)	(0.03)	(0.00)
Log Gross Fertilizer			0.31***
			(0.00)
Cons	5.54***	5.87***	
	(0.07)	(0.10)	
N	4275	4275	4275

Table 2.2: Cobb-Douglas Production at Household Farm Level

Y^e/Y^a	s_i	z_i
Static Efficiency Gains	2.57	2.18
Number of Household Farms	373	373

Table 2.3: Agricultural Output (TFP) Gain from Factor Reallocation

ductivity s_i and for adjusted productivity z_i which captures the transport infrastructure. As before, the difference between the two metrics of productivity is the role of transport costs. Using my alternative measure of productivity found residually using the LP approach, without transport costs, measured misallocation is 357 percent and with transport costs, measured misallocation is 318 percent. Therefore, again taking log (1+0.39)/log(1+2.57) shows that transport costs account for almost 30 percent of observed measured misallocation. Although, the number in magnitude shrinks compared to my main experiment, it still suggests that transport costs matter considerably for misallocation.

2.7 Conclusion

A major step low income nations can take to improve productivity is to improve the allocation of factors of production especially among farmers. This will lead to agricultural productivity growth and substantial structural change. The literature has also noted that seeing the most affect on realizing gains to productivity in agriculture is via investing on physical infrastructure, like that of rural roads (Nadeem et al., 2011). In fact, Obare et al.

(2003) find evidence that poor transportation infrastructure in Africa can adversely affect not only productivity, but also further the subsistent makeup of the agricultural sector.

My work sought to evaluate to what extent misallocation is attributed to transport costs. In other words, I essentially asked the following question, "*Do transport costs matter, in that should a hypothetical social planner account for them when deciding how to allocate various factors of production among heterogeneous producers?*" In order to explore this, I used data from the El Salvador Connectivity Project (Social Impact Inc., 2015) together with a structural model to examine the role idiosyncratic transport costs specific to farmers can explain observed misallocation. My project suggests that transportation costs do in fact matter and a social planner should be accounting for them, because in doing so one can avoid mismeasurement or measurement error associated with difference of distance to market. The main takeaway from my work is that in the presence of transport costs, the implied efficient allocation is not the same as what it looks like in the typical model of misallocation.

Chapter 3

Transportation Costs and Commercialization: A Quantitative Analysis with Data from Tanzania

3.1 Introduction

The commercialization of farmers in poorer nations is said to be one of the major requisites to realizing development initiatives (Carletto et al., 2017). Zhou et al. (2013) note that the key policy objectives behind agricultural commercialization is effectively seeing poverty lessened and development realized via the betterment of household income. This in order for rural welfare to improve (Mekonnen and Alamirew, 2017). The literature has noted that the notion of agricultural commercialization has been defined along a diverse set of lenses (Zhou et al., 2013). Some define it as the movement to seeing all or part of their harvest go to the market, away from being wholly consumed by the household (subsistent-oriented production) (van Asselt and Useche, 2022; Jones Govereh and Nyoro, 1999). Pingali and Rosegrant (1995) however sees the idea of commercialization go beyond the aforementioned idea of the saleability of crops. They say it also involves the type of inputs employed in the production process and also what products to ultimately produce. Kennedy and Cogill (1987) look at the notion of agricultural commercialization from the perspective of seeing the cultivation of crop-types of the "cash" kind expanded. Finnis (2006), for example, studies the notion of agricultural commercialization by looking at farms in South India in the transition to growing cassava, a type of cash crop.

There are a number of recent works that have studied the impacts of rural road infrastructure with specific mention to the agricultural sector. Aggarwal (2018) is one of the earlier papers that contributes to the cognizance of the effect of rural roads and questions of realizing development initiatives in agriculture. With respect to agricultural production outcomes, the author finds that there was a greater likelihood of taking up technologies that were novel amongst farmers who were beneficiaries of more roads. Shamdasani (2021) looks at the role sub par roads in rural India hinder the adoption of more sophisticated technologies, in agricultural production. The author found that enhancements to tranportation infrastructure were met with improvements to the development of the agricultural sector via the up take of modern technologies, marketable agricultural output, among others. Similarly, Asher and Novosad (2020) look at the economic effects the introduction of rural roads has in India. Although, they do not find corroboration regarding productive outcomes in agricultural sector with the advent of rural roads, (e.g., crop commercialization or greater farm output, for example), what the authors do note is a significant movement of labor out of the agricultural sector, to the non-agricultural sector, which Shamdasani (2021) does also find, albeit among those whose villages were near the sector. Correspondingly, Sotelo (2020) studies the ramifications of transport infrastructure improvements in Peru via counterfactual analysis where he simulates two policies related to road infrastructure. The first simulation related to the paving of roads saw agricultural productivity rise and welfare for the median farmer increase too.

On the topic of rural infrastructure and transportation, these are said to be lacking in many areas globally. Carney (1999) states that it is the isolation of rural people which can serve to explain the poverty they experience. The literature also highlights that distance to market matters, indicating that if one is closer in the vicinity of markets, the more likely they are to be commercial (Tafesse et al., 2020). In addition to this, Mekonnen and Alamirew (2017) note that the quality and quantity of road transport matters too. Among the four significant variables on their effect of output market accessibility, Ahmed et al. (2016) finds both farm to market distance and cost of transport to be a deterrent for small farmer access in Pakistan. On the topic of its impact on incomes, Ahmed et al. (2016) finds that both variables see an inverse relationship to incomes for farmers.

Mukarumbwa et al. (2018) found that small farmers in Zimbabwe saw their vegetable produce sold at their farm gate. Reason being these farmers found this avenue to see transaction costs lessened. Similarly, Khapayi and Celliers (2016) find that of the many factors that emerging farmers in the Eastern Cape Province are met with, when it comes to moving to farming of the commercial agriculture type, is that of expensive costs of transport. They found that farmers in the Eastern Cape Province saw most of their produced goods put for sale in informal markets whose market value was poor. Among a number of significant factors, Abayneh and Tefera (2013) find that transport costs for farmers to the market are also important. Farmers who are farther from the respective market could likely find it harder to see the transport of their goods to the market compared to those located closer. This due to the fact that they (farmers located farther away) will likely be met with pricier costs of transport. Omiti et al. (2009) also find that transportation costs are particularly important surrounding the discussion around rural farmers and their participation in commercial agriculture in Kenya. They find that the culprit behind the continuous cultivation of cereals, like maize (low value crops) among rural farmers, is questions of distance.

This brings me to my work. I study how idiosyncratic transportation costs to the market affect crop commercialization among farmers in Tanzania in the long rainy season. This paper supports the view of crop commercialization particularly that of shifting towards cash crop production away from subsistent or food crop farming[1]. Using rich data from the *Tanzania National Panel Survey (TZNPS)* which includes data on harvests, production inputs, distance to market, and high resolution geographic data on transport costs, together with a structural model, I examine the role transport costs for farms play in the transformation from subsistence to commercial farming. I quantify the importance of transportation costs for the level of commercialization among farmers in Tanzania. My counterfactual experiments suggest that the crop technology decision is sensitive to transportation costs. When transportation frictions for food crop farmers from farm to market are alleviated along different degrees, these farmers elect to switch to more sophisticated cash crop farming, away from subsistent crop farming. When I reduce transport frictions for every food crop farmer to a minimum, for example, the share of cash crop farmers goes up by 14 percentage points.

The chapter is organized as follows. Section 2 will address what the literature says on the topic in question. Section 3 will briefly discuss Tanzania and its background. Section 4 will outline the details of the micro data. Section 5 will outline the model. Section 6 will discuss the quantitative analysis followed by the results of the counterfactual experiments in Section 7. Section 8 concludes.

3.2 Literature Review

There is a growing recent literature that has effectively put structure on transportation costs in potentially explaining questions of growth and productivity in the development arena of study. These including Aggarwal et al. (2022), Sotelo (2020), De Soyres et al. (2020), Moneke (2020), just to mention some. However, to my knowledge, there is none that explicitly quantifies the role transportation costs to market play in facilitating questions of agricultural commercialization using a structural approach outside of the typical reduced-form analysis.

Having said that, it is important to highlight where my paper sits in the standing liter-

[1] I do not consider commercialization from the point of view of scale or market participation.

ature. This work relates to papers that aim to shed light on the role transportation costs play in explaining the large fraction of people engaging in subsistence farming using general equilibrium models (Gollin and Rogerson, 2014; Adam et al., 2012). Similar to these papers, I too am interested in attempting to see what role transport costs play in explaining the large share engaging in the subsistence agriculture in Tanzania, but unlike them, my focus again is development *within* agriculture, (particularly evidence of commercialization in agriculture). Their work in contrast looks at movement out of the agricultural sector all together. In other words, I do not study how transport costs affect the sectoral composition of labor across the sectors (agricultural versus non-agricultural), but am interested in the composition of operators engaging in food crop farming versus cash crop farming within the agricultural sector. That is, departing from the literature, I consider to what extent *idiosyncratic* transportation costs for farmers are related or not, to their subsequent crop technology choice.

This paper also is similar in spirit to papers that effectively look to quantify the importance of improved market access through improved transport infrastructure initiatives using a structural approach (Brooks and Donovan, 2020; Adamopoulos, 2024). Brooks and Donovan (2020) explore the affect the creation of novel bridges in rural Nicaragua serve to connect isolated populations and necessary labor markets. Adamopoulos (2024) studies the impact of new roads in Ethiopia and the ultimate question of structural change using a spatial model. Although I do not have structural change forces in my work, similar to his work, he does touch on what role changes in transport costs, that are spatially dispersed, had on the make-up of the agricultural sector, one of which being the type of crops that are produced. He finds evidence of switching towards crop production of the "cash," kind. Seeing that I too look to assess the impact changes in transport costs have on agricultural productivity, different to Adamopoulos (2024), I infer a transport cost specific to each farmer, irrespective if they went to the market or not, to see its impact on the subsequent crop technology decision.

The last paper which my work is also quite closely related to is work by Blankespoor et al. (2018) who also study the effects of transport infrastructure, particularly improvements to it with the advent of the creation of a bridge. The authors find that those who face lower transport costs (are in closer vicinity to the bridge) dedicate more land towards the production of vegetables (high value crops). This despite the possibility that the productivity of the land could very well fare poorly compared to others. Seeing that there is an obvious thematic overlay between my paper and theirs, the major difference is our approaches–whilst they employ a reduced-form approach, my analysis makes use of a structural approach.

3.3 Background

Most Tanzanians reside in the rural areas of the country and are seen to be mainly engaging in subsistent-type farming (Coleman and Raleigh, 2020). Like other developing countries, the agricultural sector is hugely important for Tanzania, seeing it make-up 28 percent share of it's GDP (FAO, 2024). In addition to that, there is a huge share of its labor sitting in the agricultural sector, accounting for a whopping 66 percent share of employment in 2022 (ILO, 2024). Yet, along measures of productivity, Tanzania does not fare so great along this measure in agriculture (Adam et al., 2012). The literature has also noted that enduring poverty is a reality for most in Tanzania. The rural population unfortunately is said to fall the most prey to impoverishment, with a lot of blame shifting to the rural country's questionable infrastructure (Coleman and Raleigh, 2020).

Although Tanzania does not fare the worst along questions of its overall road networks compared to the rest of Africa, ease of access still remains an issue to confront for rural communities. No more than 24 percent reside within 2 km of all-season roads in rural Tanzania (Shkaratan, 2012). That is, transportation infrastructure in rural Tanzania is said to be characterized as either poor or entirely absent (Temu et al., 2005), with impediments to urban access not only be that of distance, but also related to poor maintenance (Adam et al., 2018). Of the roads that do exist in rural Tanzania, the literature has noted that throughout the country's events of wet weather, these said roads are mainly untraversable (Temu et al., 2005). Adam et al. (2018) note that the virtual inaccessibility that many Tanzanian farmers face during this season of rainy conditions can very well impact their incentives to produce certain crops like vegetables, among others.

3.4 *TZNPS* Micro Data

I use household level micro data from the two most recent waves available (wave 4 and wave 5) of the *Tanzania National Panel Survey* in the long rainy season. Wave 4 panel was conducted from 2014-2015 (NBS, 2015) and wave 5 was conducted from 2020-2021 (NBS, 2022). The data includes rich information including transportation costs to market, distance to markets, agricultural production information, harvest information, input information, just to mention some[2]. To construct the panel, I only keep households who were present in both waves. That is, if a household split-off in the next wave, I do not include them and effectively only keep "original," households. This to attempt to ensure that my estimate of farm productivities I do in the quantitative section is free of potential mismeasurement.

To compute total production of a household, I use gross real harvest in kg, where I aggregate the crops using a common price (average) and account for intermediate inputs (value of fertilizers in kg). This allows me to get a final output measure of value added for each household. Next, I take the sum of all lands to get a total size of land measured in acres and sum of all labor (both hired and family) measured in days[3]. This gives me a final balanced panel of 402 household farms. Finally, to categorize farm operators across the two crop technologies, I do the following: if a household farm has at least one cash crop in their harvest bundle, I categorize them as a "cash crop farm," and if a household farm has only food crops in their harvest bundle, I categorize them as a "food crop farm." Following Adamopoulos and Restuccia (2020), I distinguish between the two crop types by

[2]For more information on the data from wave 4 and wave 5, please see NBS (2015) and NBS (2022).

[3]I use the *z score method* to account for any potential labor outliers.

producing motive. Food crops are agricultural crops that are grown mainly to be consumed by the household and put on the market to be sold–i.e., for semi-subsistence purposes. Some examples of some food crops in the data are maize, paddy and sorghum, among others. Cash crops, in contrast, are agricultural crops produced for the purpose of sale to earn money. Some examples of some cash crops in the data are cotton, tobacco, jute, among others. Unsurprisingly, 81 percent of the household farms in the panel are food crop farmers.

3.5 Model

My model builds on the work of Lucas Jr (1978) and Adamopoulos and Restuccia (2020). Akin to Adamopoulos and Restuccia (2020) I abstract from including the non-agricultural sector, and solely have a one sector model with a crop technology choice for farmers. However what separates my work, is that, in my paper I introduce the friction of idiosyncratic transportation costs from farm to market and I quantify it. This allows me to see to what extent idiosyncratic transport costs for Tanzanian farmers can explain crop technology choice, and the subsequent share electing to sit in subsistence farming.

3.5.1 Environment

I develop a one-sector (agricultural) static general equilibrium (GE) model with a crop technology choice and idiosyncratic transportation costs from farm to market for farmers. Each farmer must decide between either choosing a food crop technology or a cash crop technology. Each farmer also faces their own specific transportation cost to the market. Finally, I focus on a closed economy, abstracting from including imports or exports.

3.5.2 Agricultural Production

There are two technologies for producing the agricultural good given by:

$$y_i(s) = (A\kappa_i s)^{1-\gamma}(l)^{\gamma}$$

where γ is the land income share $\in (0,1)$, $i \in \{$ c, f $\}$ where c = cash crop, f = food crop. y_i is the output of farm and is measured as value added. There is a single production input which is l and represents total land size measured in acres. s is the permanent farm productivity (idiosyncratic) and A is the common economy wide productivity. κ_i is the crop specific technology. It is important to note that the manner in which I model the transport cost in my work is as an output tax $(1-\tau)$. Like Chapter 2, modelling the transport cost as a heterogeneous tax on output matters with respect to how much one trades. However, given that transport costs intrinsically are not taxes (do not look like taxes), I circumvent this by making the simple transformation where $(1-\tau) = 1/T$.

3.5.3 Optimality

The profit max problem under each technology, for a farmer, given productivity s transport tax τ and for $i \in \{$c,f$\}$ is given by the following:

$$\pi_i(s,\tau) = \max_{\{l_i\}} \{(1-\tau)p_i y_i - q l_i - p_i C_i\}$$

where q is the price of land and p is the exogenous price of each crop type and C_i is the fixed cost of operating a farm for crop type i. Given the first order conditions, the input demand for land, output and profits are given by the following:

$$l_i(s,\tau) = \left(\frac{\gamma p}{q}\right)^{\frac{1}{1-\gamma}} A\kappa_i s \underbrace{(1-\tau)^{\frac{1}{1-\gamma}}}_{\varphi}$$

$$y_i(s,\tau) = \left(\frac{\gamma p}{q}\right)^{\frac{\gamma}{1-\gamma}} A\kappa_i s (1-\tau)^{\frac{\gamma}{1-\gamma}}$$

$$\pi_i(s,\tau) = (1-\gamma)(p_i)^{\frac{1}{1-\gamma}} \left(\frac{\gamma}{q}\right)^{\frac{\gamma}{1-\gamma}} A\kappa_i s \underbrace{(1-\tau)^{\frac{1}{1-\gamma}}}_{\varphi} - p_i C_i$$

Let $x \equiv s\varphi$. The pair (s, φ) is drawn from known population distribution, density $f(s, \varphi)$ and cdf $F(s, \varphi)$.

3.5.4 Farmer's Technology Choice

The decision threshold that determines crop choice is represented by occupational choice $o_i(s, \varphi) = o_i(x)$, such that: $o_i(x) = 1$ if $\pi_i \geq \pi_{-i}(s, \varphi)$ Threshold $\bar{x}$ determines split between two crop farmers, such that: $\pi_f(\bar{x}) = \pi_c(\bar{x})$. Figure 3.1 depicts the farmer's technology choice. Obviously, to operate a cash crop farm will require a higher fixed cost to the operation of a food crop farm. The choice of fixed costs matters in that, following my calibration $(C_c > C_f)$, there is a single threshold given by $\bar{x}$ which determines the crop choice decision. The space in figure 1 is split into two parts, such that if a farmer's profit is below $\bar{x}$ they choose food crop technology and if a farmer's profit is above $\bar{x}$ they choose cash crop technology. What the figure suggests is that a farmer's skills alone does not solely determine the subsequent crop technology choice, but a combination of skills and the idiosyncratic transport cost to market they face.

3.5.5 Competitive Equilibrium

The definition of a competitive equilibrium is price q_i, occupational choice $\{o_i(s, \varphi)\}_{i\in\{c,f\}}$, and $\{l_i(s, \varphi), y_i(s, \varphi), \pi_i(s, \varphi)\}_{i\in\{c,f\}}$ such that: Given price, farmers optimize and o_i is the optimal occupational choice decision. Lastly, the land market clears.

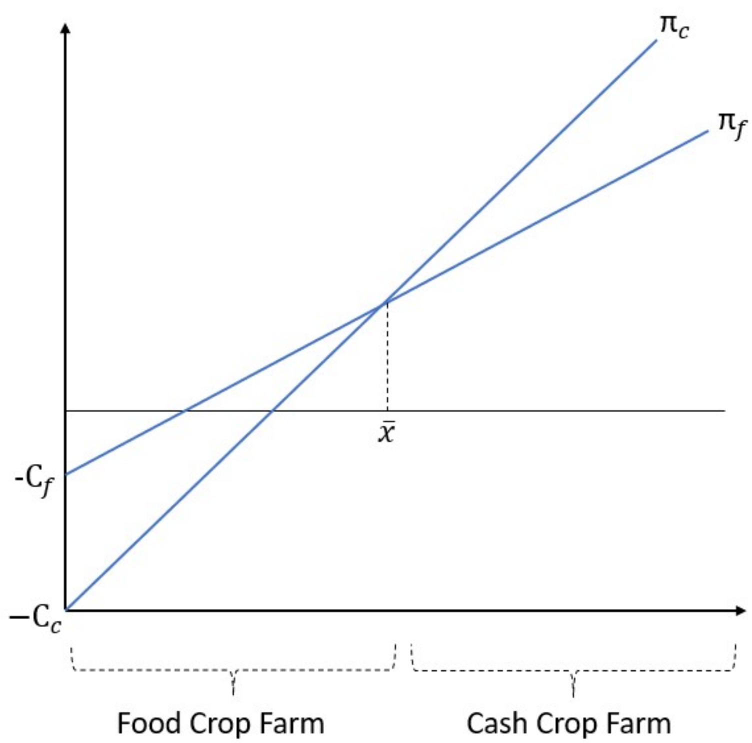

Figure 3.1: Split Between Food Crop Versus Cash Crop Farms

3.6 Quantitative Analysis

3.6.1 Farm Productivity

I construct farm productivity s_i found residually using the agricultural production function and use factor share for land borrowed from Manysheva (2022). She uses a functional approach to estimate production function estimates in Tanzania. I set $\gamma = 0.26$. Following Adamopoulos et al. (2022), I then use the computed s_i values to estimate a household-farm productivity fixed effect measure purged of time and individual effects. I have a balanced panel using the two most recent waves of the *TZNPS*, with the most disaggregated variable available to me for administrative regions.

3.6.2 Transportation Costs

I infer a transportation cost for everyone, irrespective of if a farmer actually went to the market. In the data, I have per km distance for everyone, so I *back out* the transport costs. For those who actually went to the market, I calculate the cost per km and then take average of cost per km by administrative region. The reason to find the mean transport cost per km by region, is to capture the potential different infrastructure levels across the country. Next, I estimate a transport cost T for every farmer, equal to distance (in km) multiplied by the per unit transport cost by administrative region.

3.6.3 Calibration

Using the second last wave of the *TZNPS* balanced panel, I calibrate the benchmark economy including s and transport cost term φ. I assume a bivariate log-normal distribution for (s, φ) with mean (μ_s, μ_φ) and variance-covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_s^2 & \sigma_{s\varphi} \\ \sigma_{s\varphi} & \sigma_\varphi^2 \end{bmatrix}$$

where σ_s, σ_φ and $\sigma_{s\varphi}$ is the dispersion of farm ability, dispersion of transport and covariance of farm ability and transport. Following Adamopoulos and Restuccia (2020), I set the means to 0. The three sets of parameters to be calibrated are $\{A, \kappa_f, \kappa_c, C_f, C_c, \gamma\}$, $\{\sigma_s, \sigma_\varphi, \sigma_{s\varphi}\}$, and L which is the land size.

Following Adamopoulos and Restuccia (2020), I choose the distributional population moments to match those in the data given by Table 3.1. The calibration strategy is as follows and is borrowed by Adamopoulos and Restuccia (2020). I choose 100,000 pairs of (s, φ) generated where I iterate on $\{\sigma_s, \sigma_\varphi, \sigma_{s\varphi}\}$ drawn from a log-normal distribution. Next, using the correlated data that was generated in the previous step, I then calibrate the remaining parameters. Table 3.2 provides the resulting parametrization. I normalize the economy wide productivity A, food-crop productivity κ_f , and relative prices all to 1. I also normalize the fixed cost for operating a food crop farm c_f to 0. I borrow the land income share γ from Manysheva (2022), and set κ_c to parallel within reach the ratio of average crop productivities of 1.51. The implied land L is 6.53. Finally, an equilibrium is solved for such that the fixed cost for operating a cash crop farm C_c matches the share of cash crop farmers in the data, equal to 18%. As Table 3.3 illustrates, the model's performance seems strong, as the targeted versus calibrated moments are in the same neighbourhood of one another.

Statistics	**Value**
STD of log-farm TFP	1.33
STD of log-farm transport distortions	2.21
CORR of log-farm TFP and log-farm transport distortions	-0.09

Table 3.1: Targeted Empirical Conditional Moments

Parameter	Value	Target
A	1	Normalization
κ_f	1	Normalization
p_c/p_f	1	Normalization
γ	0.26	land income share
κ_c	1.45	Ratio of average crop productivities
C_f	0	Normalization
C_c	2.24	Share of cash crop farmers
σ_s	1.77	Dispersion of abilities
σ_φ	4.88	Dispersion of transport distortions
$\sigma_{s\varphi}$	-0.26	Correlation of transport distortions-TFP
L	6.53	Average farm size

Table 3.2: Parametrization

Moments	Data	Model
STD of log-farm TFP	0.99	1.02
STD of log-farm transport distortions	2.21	2.21
CORR of log-farm TFP and log-farm transport distortions	-0.09	-0.02
Share of cash crop operators	0.18	0.18
Average farm size	6.53	6.53
Output ratio	1.98	1.96

Table 3.3: Targeted Moments Versus Calibrated Moments

3.7 Results

I conduct various counterfactual experiments where I keep everything the same as the benchmark economy, and solely manipulate transport costs only–particularly reducing transport costs to various levels for food crop farmers. This effectively to gage what happens to crop technology choice, if I remove frictions to transportation to market among food crop farmers. Table 3.4 shows my first counterfactual experiment. My first counterfactual experiment involves me reducing the transportation costs to the minimum for every food crop farmer. What I find is switching away from food crop farming towards cash crop farming, such that the share of cash crop farmers goes up by 14 percentage points. However I do see labor productivity go down for each type of farm. Labor productivity in food crops drops considerably primarily due to the drop in the land-labor ratio, as much more land is reallocated to the larger in size cash crop farms. In addition, average abilities to both farm types goes down. Table 3.5 shows my second counterfactual experiment. When I reduce the transport costs to the median for every food crop farmer, I again see evidence of switching to cash crop farming, but here the increase sits lower at 9 percentage points. Similar to the first experiment, labor productivity in both farm types does go down. Although labor productivity went down for food crop farming, average ability actually went up, whereas for cash crop farming it went down. This due to the fact that the land-labor ratio dropped considerably. The decrease in labor productivity for food crop farms was not because of the labor, since that actually went up. It was because of the drop in land, because the cash crop farms are larger in size and command much more land. Finally, the last counterfactual, given by Table 3.6, looks at reducing the transport frictions to the average for every food crop farmer. Here, I see switching to cash crop farming, but the intensity of the change is far less than the other two experiments. In this experiment, although I do see labor productivity go down for cash crop farming, I do see the labor productivity in food crop farming go up. This speaking to the idea that the farm operators that switched into cash crop farming saw abilities less than the average of the existing pool of workers (average ability went down for cash crop farming, but went up for food crop farming).

Moments	**BE**	**Counterfactual**
Share of Cash Crop farmers	0.18	0.32
Share of Food Crop farmers	0.82	0.68
Y_c/N_c	1	0.71
Y_f/N_f	1	0.11

Table 3.4: Reduce Transport Costs to Minimum

Moments	**BE**	**Counterfactual**
Share of Cash Crop farmers	0.18	0.27
Share of Food Crop farmers	0.82	0.73
Y_c/N_c	1	0.74
Y_f/N_f	1	0.91

Table 3.5: Reduce Transport Costs to Median

Moments	**BE**	**Counterfactual**
Share of Cash Crop farmers	0.18	0.19
Share of Food Crop farmers	0.82	0.81
Y_c/N_c	1	0.93
Y_f/N_f	1	1.01

Table 3.6: Reduce Transport Costs to Average

3.8 Conclusion

I explore the role idiosyncratic transportation costs play in facilitating transformation from subsistence to commercial farming in the long rainy season in Tanzania. That is, I consider to what extent idiosyncratic transportation costs for farmers are related or not to their subsequent crop technology choice. As the literature has noted, farmers lack the motivation to produce for the market, given how pricey it is to effectively participate. Tanzania like other developing countries sees a huge share of labor sitting in the agriculture sector, with majority of rural Tanzanians engaging in subsistence farming. Given the poor state of transport infrastructure in rural Tanzania, this paper aimed to effectively quantify its importance in relation to affecting crop commercialization in the long rainy season. With various reductions in transportation costs for food crop farmers, the experiments suggested that crop technology choice is sensitive to transportation costs, such that one saw switching away from food crop to cash crop farming, or evidence of crop commercialization.

Bibliography

Abayneh, Y. and Tefera, T. (2013). Factors influencing market participation decision and extent of participation of haricot bean farmers in meskan district, ethiopia. *International journal of management and development studies*, 2(8):17–26.

Adam, C., Bevan, D., and Gollin, D. (2018). Rural–urban linkages, public investment and transport costs: the case of tanzania. *World Development*, 109:497–510.

Adam, C., Bevan, D., Gollin, D., and Mkenda, B. (2012). Transportation costs, food markets and structural transformation: The case of tanzania. Working Paper]. University of Oxford, Department of Economics, Working Paper.

Adamopoulos, T. (2011). Transportation costs, agricultural productivity, and cross-country income differences. *International Economic Review*, 52(2):489–521.

Adamopoulos, T. (2024). Spatial integration and agricultural productivity: Quantifying the impact of new roads. *American Economic Journal: Macroeconomics, Forthcoming*.

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

Adamopoulos, T. et al. (2019). Spatial integration, agricultural productivity, and development: A quantitative analysis of ethiopia's road expansion program. In *2019 Meeting Papers*, number 86. Society for Economic Dynamics.

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–1697.

Adamopoulos, T. and Restuccia, D. (2020). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics*, 12(3):1–39.

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? evidence from india. *Journal of Development Economics*, 133:375–395.

Aggarwal, S., Giera, B., Jeong, D., Robinson, J., and Spearot, A. (2022). Market access, trade costs, and technology adoption: Evidence from northern tanzania. *Review of Economics and Statistics*, pages 1–45.

Ahmed, U., Ying, L., Bashir, M., Abid, M., Elahi, E., Iqbal, M., et al. (2016). Access to output market by small farmers: The case of punjab, pakistan. *Journal of Animal and Plant Sciences*, 26(3):787–793.

Alvarez, J. A. (2020). The agricultural wage gap: Evidence from brazilian micro-data. *American Economic Journal: Macroeconomics*, 12(1):153–73.

Alvarez-Cuadrado, F., Amodio, F., and Poschke, M. (2019). Selection and absolute advantage in farming and entrepreneurship: microeconomic evidence and macroeconomic implications.

Aragón, F. M., Restuccia, D., and Rud, J. P. (2022). Assessing misallocation in agriculture: plots versus farms. Technical report, National Bureau of Economic Research.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, 110(3):797–823.

Ayerst, S., Brandt, L., and Restuccia, D. (2020). Market constraints, misallocation, and productivity in vietnam agriculture. *Food Policy*, 94:101840.

Bils, M., Klenow, P. J., and Ruane, C. (2021). Misallocation or mismeasurement? *Journal of Monetary Economics*, 124:S39–S56.

Blankespoor, B., Emran, M. S., Shilpi, F., and Xu, L. (2018). Bridge to bigpush or backwash? market integration, reallocation, and productivity effects of jamuna bridge in bangladesh. *Market Integration, Reallocation, and Productivity Effects of Jamuna Bridge in Bangladesh (April 13, 2018)*.

Bravo-Ureta, B. E., Njuki, E., Palacios, A. C., and Salazar, L. (2022). Agricultural productivity in el salvador: A preliminary analysis. IADB: Inter-American Development Bank.

Brooks, W. and Donovan, K. (2020). Eliminating uncertainty in market access: The impact of new bridges in rural nicaragua. *Econometrica*, 88(5):1965–1997.

Caldwell, J., Ordonez, F., Dayaratna, V., Reynolds, M., and Felkner, J. (2013). Impact evaluation of the millennium challenge corporation transportation project in honduras: Merged baseline and endline household survey data norc at the university of chicago. data retrieved from MCC, `http://data.mcc.gov/evaluations/index.php/catalog/51`.

Carletto, C., Corral, P., and Guelfi, A. (2017). Agricultural commercialization and nutrition revisited: Empirical evidence from three african countries. *Food policy*, 67:106–118.

Carney, D. (1999). Holistic approaches to poverty reduction: Where does agricultural research fit in? *Department for International Development (DFID), London,(Unpublished)*.

Chari, A., Liu, E. M., Wang, S.-Y., and Wang, Y. (2017). Property rights, land misallocation and agricultural efficiency in china. Technical report, National Bureau of Economic Research.

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

Coleman, M. and Raleigh, D. (2020). Perpetutation of Rural Poverty in Tanzania. Ballard Brief. https://ballardbrief.byu.edu/issue-briefs/perpetuation-of-poverty-in-rural-tanzania.

De Soyres, F., Mulabdic, A., and Ruta, M. (2020). Common transport infrastructure: A quantitative model and estimates from the belt and road initiative. *Journal of Development Economics*, 143:102415.

Dong, B. and Hsieh, Y.-W. (2021). Misallocation or mismeasurement? evidence from plant-level data. *Annals of Economics and Statistics*, (142):211–234.

FAO (2024). Tanzania at a glance. https://www.fao.org/tanzania/fao-in-tanzania/tanzania-at-a-glance.

Finnis, E. (2006). Why grow cash crops? subsistence farming and crop commercialization in the kolli hills, south india. *American Anthropologist*, 108(2):363–369.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.

Gollin, D., Parente, S., and Rogerson, R. (2002). The role of agriculture in development. *American economic review*, 92(2):160–164.

Gollin, D. and Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, 107:38–48.

Gollin, D. and Udry, C. (2021). Heterogeneity, measurement error, and misallocation: Evidence from african agriculture. *Journal of Political Economy*, 129(1):1–80.

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

Heckman, J. J. and Honore, B. E. (1990). The empirical content of the roy model. *Econometrica: Journal of the Econometric Society*, pages 1121–1149.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

ILO (2020). Employment in agriculture (estimate)-honduras in 2008. `https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=HN`.

ILO (2024). Employment in agriculture (% of total employment) (modeled ILO estimate) - Tanzania. https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=TZ Accessed on February 7 2024.

Jones Govereh, T. and Nyoro, J. (1999). Smallholder commercialization, interlinked markets and food crop productivity: Cross-country evidence in eastern and southern africa.

Kennedy, E. T. and Cogill, B. (1987). *Income and nutritional effects of the commercialization of agriculture in southwestern Kenya*, volume 63. Intl Food Policy Res Inst.

Khapayi, M. and Celliers, P. (2016). Factors limiting and preventing emerging farmers to progress to commercial agricultural farming in the king william's town area of the eastern cape province, south africa. *South African Journal of Agricultural Extension*, 44(1):25–41.

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2):948–80.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341.

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

Manysheva, K. (2022). Land property rights, financial frictions, and resource allocation in developing countries. Technical report, Working Paper.

MCC (2018). El salvador compact. Millenium Challenge Corporation.

Mekonnen, N. and Alamirew, B. (2017). The effect of rural road transport on crop commercialization: Evidence from the ethiopian living standard measurement survey. *Ethiopian Journal of the Social Sciences and Humanities*, 13(2):81–108.

Miguel, E. and Hamory, J. (2009). Individual ability and selection into migration in kenya.

Moneke, N. (2020). Can big push infrastructure unlock development? evidence from ethiopia. *STEG Theme*, 3:14–15.

Mukarumbwa, P., Mushunje, A., Taruvinga, A., Akinyemi, B., and Ngarava, S. (2018). Analysis of factors that influence market channel choice of smallholder vegetable farmers in mashonaland east province of zimbabwe. *International Journal of Development and Sustainability*, 7(1):734–754.

Nadeem, N., Mushtaq, K., and Javed, M. I. (2011). Impact of social and physical infrastructure on agricultural productivity in punjab, pakistan-a production function approach. *Pakistan Journal of life and social sciences*, 9(2):153–158.

NBS (2015). National Panel Survey NPS Wave 4, 2014-2015. Ref:TZA_2014_NPS-R4_v03_M https://microdata.worldbank.org/index.php/catalog/2862/get-microdata Accessed on February 1 2023.

NBS (2022). National Panel Survey NPS Wave 5, 2020-2021. Ref:TZA_2020_NPS-R5_v02_M https://microdata.worldbank.org/index.php/catalog/5639/get-microdata Accessed on February 1 2023.

NORC (2017). Study description. `https://data.mcc.gov/evaluations/index.php/catalog/51/study-description`.

Obare, G. A., Omamo, S. W., and Williams, J. C. (2003). Smallholder production structure and rural roads in africa: the case of nakuru district, kenya. *Agricultural Economics*, 28(3):245–254.

Omiti, J. M., Otieno, D. J., Nyanamba, T. O., and McCullough, E. B. (2009). Factors influencing the intensity of market participation by smallholder farmers: A case study of rural and peri-urban areas of kenya. *African Journal of Agricultural and Resource Economics*, 3(311-2016-5509):57–82.

Pattison (2018). Maximize production using liquid fertilizer. retrieved from, `https://liquidsystems.net/wp-content/uploads/2018/02/Liquid-Fertilizer-Booklet-2018l.pdf`.

Pingali, P. L. and Rosegrant, M. W. (1995). Agricultural commercialization and diversification: processes and policies. *Food policy*, 20(3):171–185.

Pulido, J., Swiecki, T., et al. (2018). Barriers to mobility or sorting? sources and aggregate implications of income gaps across sectors and locations in indonesia. *Unpublished Working Paper, Vancouver School of Economics.*

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9:21–58.

Restuccia, D. (2016). Resource allocation and productivity in agriculture. *University of Toronto, Canada.*

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

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic dynamics*, 16(1):1–10.

Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–174.

Restuccia, D. and Santaeulalia-Llopis, R. (2017). Land misallocation and productivity. *NBER working paper*, (w23128).

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics*, 55(2):234–250.

Rovigatti, G. and Mollisi, V. (2018). Theory and practice of total-factor productivity estimation: The control function approach using stata. *The Stata Journal*, 18(3):618–662.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Schwab, K. and Porter, M. (2008). The global competitiveness report 2008-2009. http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2008-09.pdf.

Shamdasani, Y. (2021). Rural road infrastructure & agricultural production: Evidence from india. *Journal of Development Economics*, 152:102686.

Shana, B. (2011). Design and implementation of mca honduras program evaluation. https://data.mcc.gov/evaluations/index.php/catalog/51/related_materials.

Shkaratan, M. (2012). Tanzania's infrastructure: a continental perspective. *World Bank Policy Research Working Paper*, (5962).

Social Impact Inc. (2015). Northern transnational highway 2009-2014. data retrieved from MCC, http://data.mcc.gov/evaluations/index.php/catalog/135.

Sotelo, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy*, 128(7):2690–2738.

Tafesse, A., Megerssa, G. R., and Gebeyehu, B. (2020). Determinants of agricultural commercialization in offa district, ethiopia. *Cogent Food & Agriculture*, 6(1):1816253.

Temu, A., Nyange, D., Mattee, A., and Kashasha, L. (2005). Assessing rural services, infrastructure and their impact on agricultural production, marketing and food security in tanzania. *Dar es Salaam: IFPRI Eastern Africa*.

van Asselt, J. and Useche, P. (2022). Agricultural commercialization and nutrition; evidence from smallholder coffee farmers. *World Development*, 159:106021.

WB (2008). Gdp per capita. `https://data.worldbank.org/indicator/NY.GDP.PCAP.CD`.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, 128(4):1727–1785.

Zhou, S., Minde, I. J., and Mtigwe, B. (2013). Smallholder agricultural commercialization for income growth and poverty alleviation in southern africa: A review. *African journal of agricultural research*, 8(22):2599–2608.

Appendix A: Ramification of Model Under Log Normal Assumption

Difference of draws of abilities from averages defined below:

$$U_{ai} = \log(s_{ai}) - \mu_a$$

$$U_{ni} = \log(s_{ni}) - \mu_n$$

$$U_{\eta i} = \log(\eta_i) - \mu_\eta$$

$$U_{\varphi i} = \log(\varphi_i) - \mu_\varphi$$

Difference of log effective skill in sector n from average defined :

$$\hat{U}_{ni} = \log(\eta_i) + \log(s_{ni}) - \mu_\eta - \mu_n \quad or \quad \hat{U}_{ni} = U_{\eta i} + U_{ni}$$

Difference of log effective skill in sector a from average defined :

$$\hat{U}_{ai} = \log(\varphi_i) + \log(s_{ai}) - \mu_\varphi - \mu_a \quad or \quad \hat{U}_{ai} = U_{\varphi i} + U_{ai}$$

Note:

$\hat{U}_{ni}$ and $\hat{U}_{ai}$ are both normally distributed with means $\mathbb{E}(\hat{U}_{ni}) = 0$ and $\mathbb{E}(\hat{U}_{ai}) = 0$ and variances $Var(\hat{U}_{ni})$=$\hat{\sigma}_n^2$ and $Var(\hat{U}_{ai})$=$\hat{\sigma}_a^2$, respectively.

That is,

$\mathbb{E}(U_{ai}) = \mathbb{E}(\log_{\varphi i}) - \mu\varphi + \mathbb{E}(\log s_{ai}) - \mu a = \mu\varphi - \mu\varphi + \mu a - \mu a = 0$ $Var(\hat{U}_{ai}) = \mathbb{E}(U_{ai}^2) =$

$\sigma_\varphi^2 + \hat{\sigma}_a^2 + \sigma_{a\varphi} = \hat{\sigma}_a^2$

$\mathbb{E}(\hat{U}_{ni}) = \mathbb{E}(\log_{\eta i}) - \mu\eta + \mathbb{E}(\log s_{ni}) - \mu n = \mu\eta - \mu\eta + \mu n - \mu n = 0$ $Var(\hat{U}_{ni}) = \mathbb{E}(U_{ni}^2) =$

$\sigma_\eta^2 + \hat{\sigma}_n^2 + \sigma_{n\eta} = \hat{\sigma}_n^2$

Covariance of $\hat{U}_{ni}$ and $\hat{U}_{ai}$ is given by,

$cov(\hat{U}_{ai}, \hat{U}_{ni}) = \mathbb{E}[(U_{ai} + U_{\varphi i})(U_{ni} + U_{\eta i})] = \sigma_{an}$

Lastly, $(\hat{U}_n - \hat{U}_a)$ has a mean of $\mathbb{E}(\hat{U}_{ni} - \hat{U}_{ai}) = 0$ and variance given by:

$Var(\hat{U}_{ni} - \hat{U}_{ai}) = \sigma_a^2 + \sigma_n^2 - 2\sigma_{an} \equiv \sigma^2$

Sector specific log incomes of agent i are:

$\log(I_{ai}) = \log(w_a) + \log(\varphi_i) + \log(s_{ai})$

$\log(I_{ni}) = \log(w_n) + \log(\eta_i) + \log(s_{ni})$

$\log(I_{ai}) = b_a + U_{\varphi i} + \hat{U}_{ai}$

$\log(I_{ni}) = b_n + U_{\eta i} + \hat{U}_{ni}$

where

$b_a = \log(w_a) + \mu_\varphi + \mu_a$

$b_n = \log(w_n) + \mu_\eta + \mu_n$

Sectoral Employment: Probability individual chooses to work in sector a

$n_a = Pr\{\log(I_{ai}) > \log(I_{ni})\} = Pr(b_a + \hat{U}_{ai} > b_n + \hat{U}_{ni})$

$$n_a = Pr(b_a - b_n > \hat{U}_{ai} - \hat{U}_{ni}) = Pr(\tfrac{b_a - b_n}{\sigma^*} > \tfrac{\hat{U}_{ni} - \hat{U}_{ai}}{\sigma^*})$$

$$\text{let } b_1 = \tfrac{b_a - b_n}{\sigma^*} \quad \text{and } \xi_i = \tfrac{\hat{U}_{ni} - \hat{U}_{ai}}{\sigma^*}$$

where

σ^* is the variance of relative effective abilities

ξ_i is a standard normal variable

So,

$$n_a = \Phi(b_1) \text{ where } \Phi \text{ is standard normal cdf}$$

n_a is also fraction of individuals that choose sector a (agriculture) because we have a sequence of agents of measure 1 $\Rightarrow N_a = n_a = \Phi(b_1)$

Therefore,

Probability the individual chooses to work in sector n (non-agriculture) is

$$N_n = n_n = 1 - \Phi(b_1)$$