Title
Essays on Agriculture, Misallocation, and Economic Development

Author
Hang, Jing

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Agriculture, Misallocation, and Economic Development

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Jing Hang

2018

ABSTRACT OF THE DISSERTATION

Essays on
Agriculture, Misallocation, and Economic Development

by

Jing Hang
Doctor of Philosophy in Economics
University of California, Los Angeles, 2018
Professor Lee Ohanian, Chair

This dissertation consists of three chapters. This first chapter studies ability sorting between rural and urban regions induced by regional differences in prices and evaluates the importance of this mechanism in explaining the large rural-urban (agriculture-non-agriculture) income gap in developing countries. The second chapter studies how capital deepening might help explain several development facts regarding agriculture when capital-labor substitution is easier in agriculture than in non-agriculture. The third chapter explores resource misallocation in production networks and studies how ignoring them might lead to mismeasured efficiency loss from misallocation.

The dissertation of Jing Hang is approved.

Mark J. Garmaise

Pablo David Fajgelbaum

Gary D. Hansen

Lee Ohanian, Committee Chair

University of California, Los Angeles

2018

To my family . . .

for all the support during this long journey

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The completion of this dissertation benefits greatly from many people throughout the years. First, I would like to express my sincerest gratitude to my advisor Professor Lee Ohanian, for his guidance, encouragement, and support from the beginning to the end. I am also deeply grateful to my committee members, Professors Gary Hansen, Pablo Fajgelbaum, and Mark Garmaise for their comments and generous help. I would like to thank Andy Atkeson, Saki Bigio, and many others for their helpful comments at the UCLA macroeconomics proseminar. I thank Ioannis Kospentaris, Liyan Shi, Yizhi Xu, and others in my cohort for encouragement, support, and sharing the unique experience together. Finally, none of this would be possible if not for the love, understanding, and support from my parents and my fiancee Mengqi Niu, for which no words can express my gratitude.

VITA

2003–2007	B.A. (Management), School of Business, Nanjing University.
2005	First Prize, National Undergraduate Contest in Mathematical Modeling, China
2009–2012	M.A. (Economics), National School of Development, Peking University.
2009-2011	Graduate Fellowship, Peking University
2012–present	Ph.D (Economics, expected June 2018), Department of Economics, UCLA.
2012-2017	Graduate Fellowship, Department of Economics, UCLA
2017-2018	Graduate Dissertation Year Fellowship, UCLA
2013-2018	Teaching Assistant, Department of Economics, UCLA

PUBLICATIONS

The Determinants of China's Structural Change during the Reform Era (with Kaiming Guo and Se Yan), *Economic Research Journal*, No. 3, 2017. (in Chinese)

CHAPTER 1

Introduction

This dissertation consists of three essays focusing on the role of agriculture and resource misallocation in economic development.

Chapter 2 studies the "dual economy" in developing countries, namely the productivity gap between the agricultural and the non-agricultural sector. I present a new explanation for this phenomena which relies on the interaction between non-homothetic preferences and costly domestic trade. In particular, food is a necessity which has an income elasticity less than one. Food is produced in rural regions and shipped to cities incurring trade costs, while non-agricultural good is produced in cities and shipped to rural regions. Combining the preference and production structure, high ability workers sort into cities and work in non-agriculture as they spend a larger fraction of expenditure on non-agricultural goods, which are cheaper in cities. Low ability workers prefer to live in rural regions and work in agriculture for easier access to food. The sorting mechanism has larger effects in developing countries as these countries have higher trade costs and higher consumption shares of food. I formalize the idea in a two-sector multi-region general equilibrium model. Quantitative analysis of the model shows that the ability sorting mechanism can explain around 30% of the rural-urban income gap in Malawi relative U.S. I also find empirical support of model predictions using a detailed household survey from Malawi.

Chapter 3 explores the role of capital deepening on the productivity gap between agriculture and non-agriculture studied in Chapter 1, while also paying attention to two other important development facts regarding agriculture: 1. the employment share of agriculture declines in income levels; and 2. the international productivity gap is much larger in agriculture than in non-agriculture. Empirical studies show that the elasticity of substitution is larger in agriculture than in non-agriculture. This means the agricultural sector responds more strongly to the reduction in the relative price of capital. In response to this, labor will move out of agriculture, and labor productivity in agriculture rise relative to non-agriculture. I explore this capital-labor substitution mechanism quantitatively by comparing a model with CES production functions and a model with Cobb-Douglas production functions. I find that the model with CES production functions is successful in explaining the sectoral gaps in productivity while does not help in explaining the other two development facts. The reason is because agriculture is also more capital intensive than non-agriculture, which counteracts the effect coming from a higher elasticity of substitution, consistent with the findings of Herrendorf et al. (2015) in U.S. data.

The sectoral gaps in labor productivity has been deemed as a sign of severe misallocation of labor in developing countries (Gollin et al., 2014). Both the previous two chapters present arguments that this is not necessary true. Chapter 4 on the other hand discusses the measurement of the cost of resource misallocation across micro production units in more detail. The burgeoning literature on misallocation following Hsieh and Klenow (2009) has mainly used value-added production functions, which potentially could understate the cost of misallocation because these studies ignore the magnification of the effect of misallocation through intersectoral linkages (Jones, 2011, 2013). This chapter compares the efficiency loss coming for resource misallocation measured in two models conditional on observing the same data. One model has output production functions and allow for sectoral linkages and the other model uses value-added production functions. I find that when there are no distortions in intermediate input use, measured efficiency loss in the two models is identical. Both models are correct representations of the underlying data. When there are distortions in intermediate input use, the value-added model produces incorrect measures of efficiency loss. Empirical analysis using Chinese data however shows that the bias is small as the distortions in intermediate input use are substantially smaller than that in primary input use. Existing studies using the value-added model might actually have overstated the cost of misallocation due to mis-specified parameter values.

CHAPTER 2

Transportation Costs, Ability Sorting, and the Dual Economy

2.1 Introduction

Many developing countries can be characterized as a "dual economy", where the non-agricultural sector is much more productive than the agricultural sector. Despite low agricultural productivity, these countries have a large share of workers in agriculture (Gollin et al. (2002); Caselli (2005), see Figure 2.1). A mirror image of the sectoral productivity gap is the equally conspicuous rural-urban income gap (Young, 2013).[1] The "dual economy" is such a robust and salient feature of developing economies that reallocation of workers from agriculture to non-agriculture (industrialization) or from rural to urban areas (urbanization) has been viewed as synonymous to economic development. The traditional view of the sectoral productivity gap is that there are barriers preventing workers from moving to urban areas. Reallocation can bring the economy closer to the efficiency frontier. Given the large income gaps and the large share of agricultural workers in developing countries, the efficiency gain from reallocation can be huge.

This chapter proposes a new explanation for the "dual economy" which implies the gains from reallocation are not warranted. I argue that high ability workers sort into cities due to spatial price differences induced by transportation costs. As low ability (income) workers spend a larger fraction of their income on food (Engel's law), they prefer the rural areas

[1] To the extent that rural areas mostly engage in agricultural production while urban areas focus on non-agricultural production. I use agriculture-non-agriculture and rural-urban interchangeably throughout this chapter and only make clear the distinction when necessary. I do not distinguish between income and productivity either.

with cheaper food. On the other hand, high ability (income) workers prefer urban areas because they value non-food consumption more and it is cheaper in cities. The ability sorting mechanism is stronger in developing countries because 1) food consumption is more important there, and 2) the spatial price differences are larger in developing countries due to higher transportation costs. This explains why the income gap shrinks as a country develops.

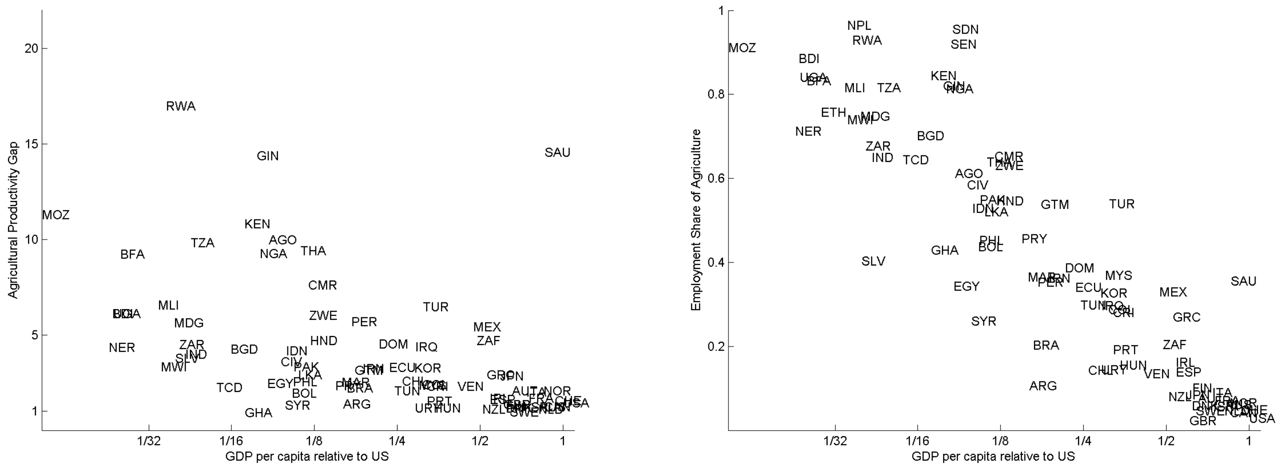
I formalize the ability sorting mechanism in a two-sector multi-region general equilibrium model. A non-agricultural sector locates in an urban center while all other regions are rural and engage only in agricultural production. There is trade between the urban center and the rural regions with varying iceberg trade costs, of which transportation costs are a main component.[2] The trade costs generate spatial price differences. Workers have Stone-Geary utility with subsistence requirement in food. Each worker draws a pair of sector-specific productivity and makes the location choice to maximize utility.

The introduction of sector-specific productivity draws follows Lagakos and Waugh (2013), in which workers make the sector choice according to their comparative advantage in different sectors, as in the classic Roy model. They find that agricultural labor productivity relative to non-agriculture is lowered when unproductive agricultural workers select into the agricultural sector as agricultural employment increases. My model reduces to that of Lagakos and Waugh (2013) when trade costs are zero, which allows me to assess the importance of the ability sorting mechanism on top of worker selection based on comparative advantage. On the other hand, the multi-region setting follows Gollin and Rogerson (2014), which allows me to zoom in on the rural regions as location choice across rural regions is only determined by the ability sorting mechanism. I add worker heterogeneity to Gollin and Rogerson (2014) and explore the role of ability sorting in explaining the spatial income differences in developing countries.

The model predicts that income declines across rural regions in the trade costs with the urban center. Using a detailed household survey from Malawi, I find that income declines in the distance to the urban center, which is taken as a proxy for trade costs. This pat-

[2] I thus do not distinguish between trade costs and transportation costs in this chapter.

Figure 2.1: The “Dual Economy”



(a) Agricultural Productivity Gap

(b) Agriculture Employment

Note: The left panel plots the agricultural productivity gap, defined as the ratio of agricultural labor productivity to non-agricultural labor productivity both measured in nominal terms, against GDP per capita relative to U.S. The right panel plots the employment share of agriculture against GDP per capita to U.S. Data come from the Food and Agriculture Organization of the United Nations (FAO,Rao (1993)) and the Penn World Table 8.1 (Feenstra et al., 2015).

tern lends support to the ability sorting mechanism as it is not readily explained by worker selection. I further test the ability sorting mechanism using detailed information on agricultural production in the data. I construct a measure of farmer's productivity by estimating a plot level production. Consistent with the ability sorting mechanism, my measure of agricultural productivity also declines in the distance to city. This finding could just reflect genuine benefits of locating near to the city, such as cheaper intermediate inputs or better access to agricultural technologies. I show that my findings are robust to these alternative explanations.

I calibrate a zero transportation cost benchmark of the model to U.S. data using data moments provided in Lagakos and Waugh (2013). I allow countries to differ in three exogenous factors: an economy wide efficiency measure that affecting productivity in both sectors, transportation costs, and land endowment. I evaluate the model by varying these factors to match aggregate labor productivity in Malawi, the spatial price differences in the micro data, and arable land per capita. The ability sorting mechanism, namely adding transportation costs in the model, significantly increases the model's explanatory power of the rural-urban income gap in Malawi. The explained share of the rural-urban income gap in Malawi relative to U.S. goes down from 41.7% to 11.7% when the ability sorting mechanism is shut down while holding aggregate labor productivity constant. Both aggregate efficiency and transportation costs affect ability sorting. To explore their quantitative importance, I perform several counterfactual experiments by varying them separately. These experiments produce very different results. In particular, raising aggregate productivity have significantly larger effects on productivity and welfare than reducing transportation costs, but similar effects on the rural-urban income differences. Transportation costs are more important in understanding regional income differences.

If the government can focus its infrastructure investment on raising aggregate efficiency or reducing transportation costs in different regions, these experiments provide useful information for evaluating different investment projects. Of particular interest is that welfare gains under these policy experiments are not equally distributed among workers. The bottom workers gain less than average workers when transportation costs are reduced. In some cases, the very poor even experience welfare loss. On the other hand, improvement in aggregate efficiency benefits the bottom workers more than others. Dollar and Kraay (2002); Dollar et al. (2016) find that the bottom workers tend to have similar income growth as the average workers in many developing countries. This analysis suggests that this might be due to different drivers of economic growth counteract each other. Governments should be careful in selecting growth-promoting policies when poverty reduction is also a target. Empirical studies such as Jacoby (2000); Jacoby and Minten (2009) find that improvement in rural transportation infrastructure benefits the poor, my findings suggest that it might not be the case if the improvement is at a large scale with general equilibrium effects taken into account.[3]

The final piece of empirical support for the ability sorting mechanism comes from international data. Lagakos and Waugh (2013) argue that as women have lower physical strength than men, they must have comparative disadvantage in strength intensive agricultural production. The selection mechanism can thus explain why the share of women in agricultural workers is increasing in agricultural employment. The same argument also implies women should always be less likely to choose agriculture than men. Using data from International Labor Organization (ILO) I find plenty of cases where women are more likely to choose agriculture than men, and the probability of observing that is increasing in agricultural employment. This can be explained by the ability sorting mechanism if women also have absolute disadvantage to men. Women's disadvantage can be explained by the model of Pitt et al. (2012) where sector-specific worker productivity takes both physical strength and human capital as inputs. If women and men have similar human capital, women will have lower productivity than men in both sectors due to their lower physical strength. The mechanism should be stronger in developing countries as strength intensive technologies may be adopted due to their lack of human capital (Caselli and Coleman, 2006), or because those countries have a smaller services sector where women have comparative advantage (Ngai and Petron-

[3]The fact that perfectly mobile workers have unequal welfare gains from reductions in transportation costs echos a recent trade literature studying the distributional effects of reduction in trade costs in the presence of non-homothetic preferences and worker heterogeneity (Fajgelbaum and Khandelwal, 2016; Nigai, 2015).

golo, 2017). The sorting mechanism is at work even when there is no difference between men and women but women receive lower wages in both sectors due to discrimination.

The rest of the chapter is structured as follows. The next section presents a literature review. Section 2.3 motivates the model in Section 2.4 using Malawi data. Section 2.5 studies the model quantitatively. Section 2.6 presents empirical support in the Malawi data for model predictions, while Section 2.7 supports the model using ILO data. Section 2.8 concludes, which is followed by an appendix containing additional results and a detailed description of data construction.

2.2 Related Literature

This chapter belongs to a recent literature on agriculture and development.[4] Caselli (2005) and Restuccia et al. (2008) single out agriculture as key to understand the huge international income differences because 1) agricultural productivity in developing countries is extremely low relative to non-agriculture, and 2) the low agricultural productivity induces large agricultural employment in developing countries due to subsistence food requirement. Many studies have since tried to explain why agricultural productivity is extremely low in poor countries.[5] This chapter relates to two sets of explanations and derives new implications from them.[6]

[4]The idea that agriculture is important for understanding development is not new. It has long been at the center stage of understanding economic development, see, e.g. Lewis (1954); Harris and Todaro (1970). The new literature builds on the old literature and brings in better data and new perspectives.

[5]Related to that,Gollin et al. (2002, 2007) emphasize the pivotal role of agricultural productivity growth in jump-starting modern economic growth.

[6]Another branch of the literature focus on the use of intermediate inputs in agriculture. Agricultural production in poor countries use very little modern intermediate inputs than developed countries. This could be due to distortions in the intermediate input markets (Restuccia et al., 2008), low productivity in producing intermediate inputs (Yang and Zhu, 2013), or farmers not willing to use intermediate inputs because of the associated risk when an insurance market is missing (Donovan, 2014). Land market misallocation represents another reason for low agricultural productivity in developing countries (Adamopoulos and Restuccia, 2014; Adamopoulos et al., 2015; Restuccia and Stantaeulàlia-Llopis, 2017). In particular, Restuccia and Stantaeulàlia-Llopis (2017) use the same Malawi data that I use to find more severe misallocation in agriculture than what Hsieh and Klenow (2009) have found in manufacturing. Other researches have studied specific factors behind land misallocation, such as untitled land (Chen, 2016), the communal land tenure arrangement in Sub-Sahara Africa (Gottlieb and Grobovšek, 2015), and a land reform in the

First, this chapter is related to a literature that emphasizes the role of worker selection in explaining productivity differences. Adopting a Roy model, Lagakos and Waugh (2013) find that an increase in agricultural employment reduces average worker productivity in agriculture but raises it in non-agriculture. This narrows the international productivity differences in non-agriculture and widens that in agriculture. Young (2013) provides the best empirical support to the self-selection argument. Using micro data from 65 countries, he shows that workers born in the urban areas have a similar probability of moving to rural areas as workers born in rural areas move to urban areas, despite a sizable rural-urban income gap observed in data. Alvarez (2017) provides further evidence using Brazil panel data, which allows him to track workers who actually move between sectors. He finds that there is no significant wage growth for workers moving from agriculture to non-agriculture when they don't change their occupation.[7] I contribute to this literature by proposing an ability sorting mechanism based on absolute advantage of workers and the differences in the consumption value of different regions. This literature relies on unobserved ability to explain the productivity gap, which is not susceptible to econometric testing. The multi-region setting in my model provides a way of testing the ability sorting mechanism.

The second literature traces the failure of agriculture in developing countries to high transportation costs (Adamopoulos, 2011; Herrendorf et al., 2012; Gollin and Rogerson, 2014).[8] High transportation costs distorts the spatial allocation of workers, which is particularly detrimental to agriculture. This is because 1) more workers are allocated to agriculture, a sector with decreasing returns due fixed land supply, and 2) agricultural production also uses less intermediate inputs because transportation costs raise the price. While these studies mainly assume homogeneous workers, I show that allowing for heterogeneous worker productivity introduces a new channel for transportation costs to generate a "dual economy" in

Philippines (Adamopoulos and Restuccia, 2015).

[7]Other studies such as Adamopoulos et al. (2015) have combined the worker selection mechanism with other distortions in the economy.

[8]It should be noted that trade literature has long been examining the effects of trade costs. Most of them do not try to explain low agricultural productivity in developing countries. An exception is Tombe (2015) who studies a multi-sector trade model with explicit reference to the problems of developing countries' agricultural sector.

developing countries.

The international comparison of sectoral productivity mirrors another finding previously emphasized in Gollin et al. (2004), that is, agricultural labor productivity is significant lower than non-agricultural labor productivity within a country when measured in nominal prices. The gap in sectoral productivity is larger in developing countries. One explanation of the productivity gap is measurement issues in the data. For example, Gollin et al. (2004) argue that agricultural workers engage in more home production than non-agricultural workers. Wingender (2015) interprets the gap as coming from sectoral differences in skill composition. By carefully correcting measurement errors in the data, Gollin et al. (2014) find the gap is still substantial after the adjustment.[9] Another explanation is that regions differ in cost-of-living or amenities such that welfare is equalized across regions. However, price differences across regions are small relative to the income differences (Ravallion and van de Walle, 1991; Brandt and Holz, 2006). Gollin et al. (2017) find that urban areas have better amenities that rural areas in almost all the amenity measures they consider. These findings indicate labor might be severely misallocated in developing countries (Gollin et al., 2014). Early studies by Lewis (1954) and Harris and Todaro (1970) trace the source of misallocation to institutional settings preventing the equalization of marginal product between the sectors.[10] Caselli and Coleman (2001) argue that reduction in migration barriers lowered the sectoral productivity gap in the U.S. The fact that many urban workers migrate to rural areas (Young, 2013) suggests worker selection might be the reason behind the sectoral productivity gap. Lagakos and Waugh (2013) though find self-selection cannot quantitatively explain the sectoral productivity gap in developing countries very well. The ability sorting mechanism on the other hand finds its biggest success in explaining the gap.

My multi-region model has features of the von Thünen model of "dual economy" (Nerlove

[9] Herrendorf and Schoellman (2015) show that the agricultural productivity gap in U.S. can be explained away by measurement errors in the data.

[10] Similar in that vein, Munshi and Rosenzweig (2016) show the gap can be explained by the rural insurance networks which raises the benefits of locating in rural areas. Young (2013) however shows that the urban workers migrating to rural areas fail to experience a reduction in the variance of consumption. Assuming the income gaps reflecting labor misallocation, Vollrath (2009, 2014) estimates the efficiency loss from that.

and Sadka, 1991). The original von Thünen model emphasizes the differential use of land in rural areas according their distance to the urban center. Fafchamps and Shilpi (2003, 2005) have examined those predictions of the von Thünen model using data from Nepal. They find that the spatial specialization of economic activities and organization of labor do vary systematically to the distance to the urban center. By adding worker heterogeneity to the original model, my model generates spatial income differences across rural areas which I find support in the Malawi data. My findings suggest that it might be fruitful to go beyond the rural-urban dichotomy and study spatial inequality within each sector, in particular that across the rural areas.

The sorting mechanism has been used to explain the spatial income gaps between cities in the urban literature (Lee, 2010; Black et al., 2009; Handbury, 2013), which often model spatial differences in housing prices as the source of ability sorting. I adopt the idea in a rural-urban setting. In developing countries, the rural-urban income gap is a more salient feature of spatial income inequality and food is a much more important consumption item for the poor which has an income elasticity less than 1.

2.3 The Spatial Profile of Income and Prices in Malawi

This section presents some motivating evidence. I examine how income (consumption) and prices vary geographically with respect to a region's distance to an urban center using data from Malawi. Malawi is a landlocked country in the Southern-Eastern part of Africa. It is one of the least developed countries in the world. In 2005 it has a GDP per capita of only 580 dollars valued at PPP prices. Over 80% of its labor force work in agriculture. Transportation infrastructure is very poorly developed in Malawi. Due to the poor conditions of its road, Malawi's domestic transport costs are much higher than neighboring countries such as Zimbabwe, which has similar income level to Malawi. Transport costs are the single largest expenditure for Malawian farmers. In sugar production, which is one of Malawi's major exports, regional and international transport costs add nearly 50 per cent to production costs for Malawian sugar.[11]

The data I use is the Malawi 2010-2011 Integrated Survey on Agriculture (the Third Integrated Household Survey, IHS3). It is part of the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) project implemented by the LSMS team of the World Bank. The ISA improves on previous LSMS surveys with a strong focus on agriculture, which allows me to study the production and consumption decisions of farmers in detail.

IHS3 has a sample size of 12,271 households and 56,397 individuals. The primary sampling unit is the census enumeration areas (EAs). There are 768 EAs in total. The EAs differ in the distance to urban markets, which is recorded in the community survey of IHS3.[12] I use two measures of distance: the physical distance to the nearest urban center and the cost of total fare by regular motola to the nearest urban center.[13] As the cost of fare doesn't include all the transportation costs to the urban center, I will focus on the physical distance.[14] The results using the cost of fare is similar but I do not report them in the text. IHS3 also asks what the nearest urban center is. There are four urban centers to choose from: Mzuzu, Lilongwe, Zomba, and Blantyre. These urban centers account for 11.9% of total population, while the fraction of urban population is only 15.3% in total. The fact that the majority of urban population live in the four urban centers allows me to treat all EAs except for those in the urban centers as the rural area without worrying too much about the small cities.[15]

Figure 2.2 presents the geographic distribution of EAs according to the distance and cost

[11]The description above is based on OECD's African Economic Outlook 2005-2006.

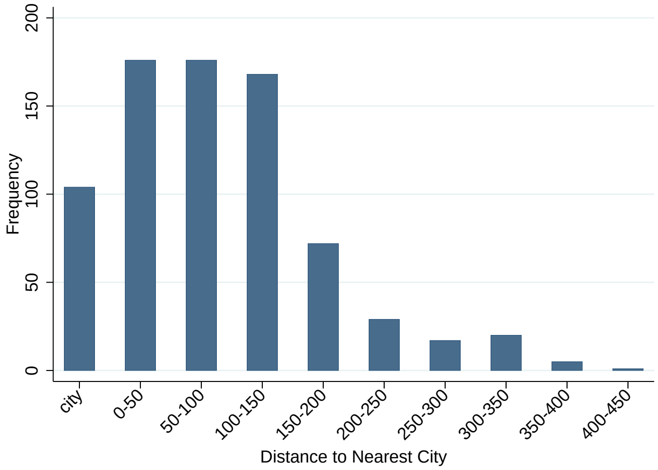
[12]A community identified in the community survey doesn't corresponds to an EA exactly. It is representative of the EA as a whole.

[13]A motola is a pick-up truck that serves as an informal public transit.

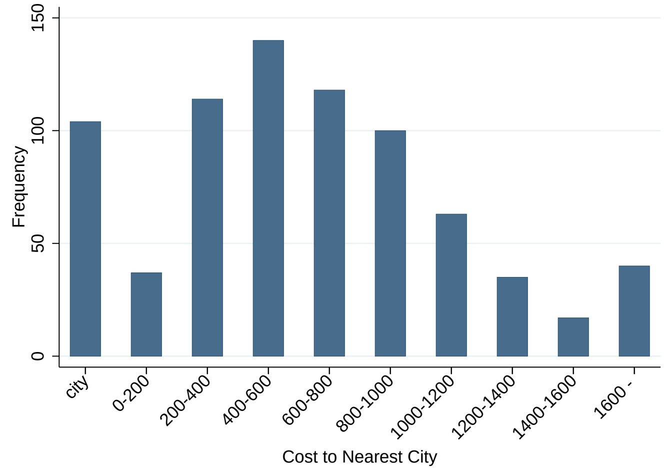
[14]The question in the survey is: What is the cost of the total fare to go by regular matola from here or the nearest matola stage to the nearest major urban centre, even if one has to change matola en route? It thus might miss costs to the matola stage for some EAs. Changing matola might also incur additional time cost that is not included in the total fare.

[15]Ideally, I would like to consider the distance to all cities, but the data is insufficient for that task. Fafchamps and Shilpi (2003, 2005) use more detailed data on the geographic characteristics of regions in Nepal. They focus on the examining the spatial specialization pattern instead of income.

Figure 2.2: Geographic Distribution of EAs



(a) Distance (Kilometer)



(b) Cost (Malawian Kwacha)

of fare to the nearest urban center. The distribution of EAs is relatively concentrated around the urban centers. Most EAs locate within a 150km radius of one of the urban centers. Most EAs have a cost of fare less than 1000 Malawian Kwacha.[16] The two measures are highly correlated with a correlation coefficient of 0.82. Given the scarcity of data for the remote areas, it is not surprising that the spatial profile presented below has much larger confidence interval for the farthest areas.

I construct measures of income and consumption at the household level from the data. These measures are aggregates of income from different sources and consumption of different categories. The procedure mainly follows De Magalhães and Stantaeulàlia-Llopis (2015) who also use the data to document the cross-sectional facts of consumption, income and wealth in Malawi. Section 2.9.2 gives a brief description of how I construct the income and consumption measures.

[16] As of September 26, 2017, the exchange rate between Malawian Kwacha and U.S. dollar is 723.6 Kwacha/dollar.

2.3.1 Income and Consumption

This subsection studies spatial inequality in Malawi. The fact that the urban-rural income gap can be very large in developing countries is well known in the literature (Gollin et al., 2014). Malawi is a good example of that. According to Malawi's national statistics, agriculture's share of GDP in 2010 is 29.6% while its share of employment is 86.4%, which implies an sectoral productivity gap over 15![17] This gap might be explained by measurement errors in employment and GDP, and the gap might also reflect differences in observable characteristics of agricultural and non-agricultural workers. Gollin et al. (2014) estimate that a half of the raw agricultural productivity gap can be explained by these factors.[18] Even with the adjustment, the sectoral differences is still staggering, implying large potential gains from labor reallocation. Looking at the rural-urban income differences, my calculation using the Malawi data shows a gap of 3.01, which is a lot lower than the agricultural productivity gap but still respectable.[19]

The rural-urban dichotomy obviously is an over-simplified description of regional inequality in developing countries. All people certainly do not have equal income within rural or urban areas. Economists often consider economic isolation thus lack of trade as a source of low income. In line of that, we might expect income to vary substantially across rural regions, as the vast rural hinterland in developing countries hides large differences in the access to the urban markets. I next examine how income varies with the distance to the nearest urban center.[20] For all the results presented below, I net out the differences across

[17]The data comes from Malawi's Statistical Year Book 2012, which can be downloaded at http://www.nsomalawi.mw/images/stories/data_on_line/general/yearbook/.

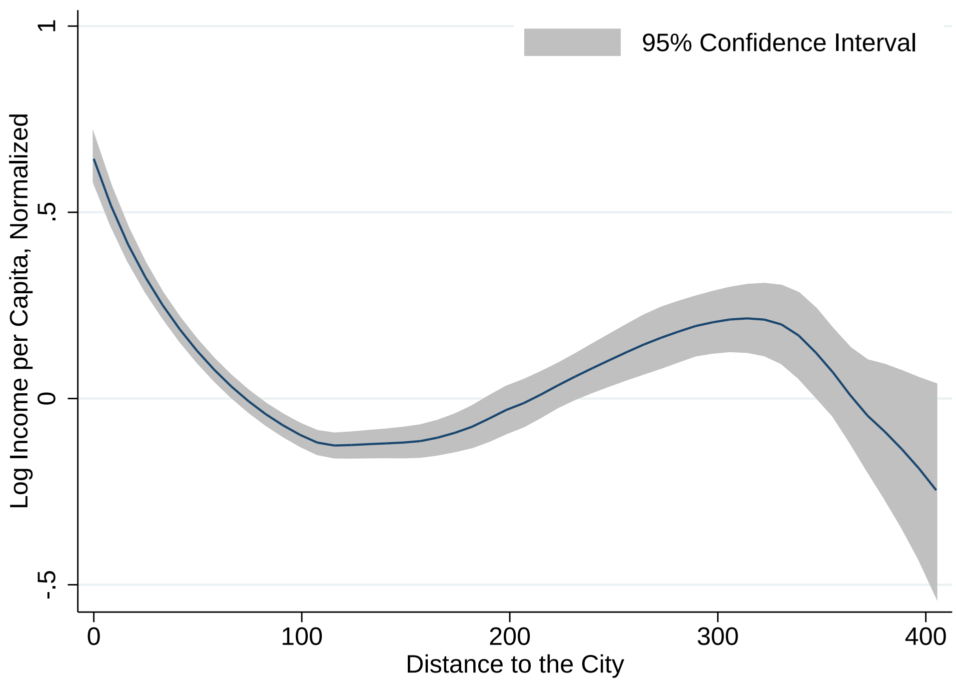
[18]Gollin et al. (2014) use data for 2005. The raw and adjusted productivity gap between agriculture and non-agriculture for Malawi are respectively 12.54 and 6.23.

[19]The difference between agricultural productivity gap and rural-urban income gap indicates that even within the same area there is an income gap between agriculture and non-agriculture. This could be due to labor market frictions such as a rationed non-agricultural labor market, or ability sorting between the two sectors, or some insurance role provided by agriculture (Munshi and Rosenzweig, 2016). It could also be due to ability sorting caused by transaction costs within an area, such as the cost of going to the market. I am not examining this aspect of the data. The model studied below have agriculture and non-agriculture take place in different areas.

[20]As mentioned above, I also use the cost of fare to the nearest urban center as a measure of access to urban market. The results are similar but I do not report them in the text.

urban centers by regressing the variables of interest on a set of city dummies. Through out this chapter, I examine the relationship between a variable of interest and the distance to the nearest urban center using the non-parametric kernel-weighted local linear smoothing method. Unless otherwise noted, I choose an Epanechnikov kernel and a bandwidth of 50 for the local regressions.

Figure 2.3: Spatial Income Profile



Note: Income comes from author's calculation based on IHS3 data. Log income per capita is smoothed using the kernel-weighted local linear regressions with Epanechnikov kernel and bandwidth equal to 50.

Figure 2.3 plots log income per capita against the distance to the nearest urban center.[21] Income initially declines in the distance to the city, then slightly increases and finally decreases in the end. This suggests that geographic isolation reduces income near the city but its effect tappers off as we go further away from the city. As discussed above, most EAs locate within 150 kilometers from a urban center, this suggests that the results for the far away areas should carry less weight given the paucity of data for those areas. This is

[21] I also use income per worker instead of income per capita, where a worker is defined as someone aged between 15 and 64. The results are similar.

reflected in the much larger confidence interval for the distant areas. Another reason to put less weight on the results for the far away areas is that these areas are less likely to be economically influenced by the urban center than some other nearby small cities. Taking these into consideration, the spatial income profile in Figure 2.3 suggests that there is not only a large income gap between urban and rural areas, but income also tends to decline in the distance to the city within rural areas.

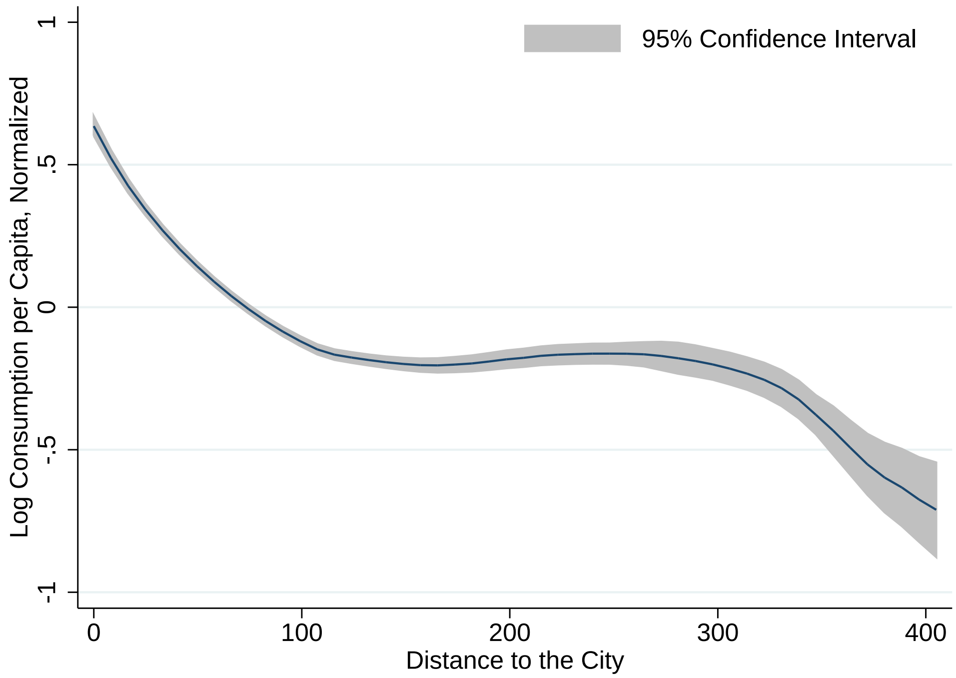
Since I only have a cross-section of data. The income measures I calculated contains a permanent income component and a random shock. We are interested in the permanent income component, which is better captured by consumption instead of income (Battistin et al., 2009). In the data, consumption shows much less variation than income. Consumption might also be more reliable if survey respondents are more reluctant to report income than consumption or if income is hard to measured, which is particularly the case when most people produce agricultural products not sold in the markets. In line of these considerations, Figure 2.4 presents the spatial profile of consumption per capita.[22] The rise in the middle range is much less significant. Overall, there is a declining trend extended to the farthest areas. This confirms that income might be systematically related to the location of a place. The spatial variation is also economically significant: both the income and consumption gaps between the poorest and richest regions can be as large as one log point.

The spatial income profile might not only reflect the productivity differences between agriculture and non-agriculture. Fafchamps and Shilpi (2003) find that regions closer to the city also engage more in non-agricultural production, which should lead to higher income in those regions if non-agriculture has higher productivity. If these differences in regional production structure can fully explain the spatial income profile, we should try to understand why the production structure differs across regions. To examine whether it is the case in data, I include the share of income coming from agriculture as a determinant of household income.[23] Even though the agricultural share of income is an economically and statistically

[22]The results are similar if I use consumption per adult equivalent. Following Meyer and Sullivan (2009), I use the equivalence scale $(A + PK)^F$, where A is the number of adults in the family, and K is the number of children, with the child proportion of an adult $P = 0.7$ and economies of scale factor $F = 0.7$.

[23]Income shares are easily available as I construct total income from different income sources. It is hard

Figure 2.4: Spatial Consumption Profile



Note: Consumption comes from author's calculation based on IHS3 data. Log consumption per capita is smoothed using the kernel-weighted local linear regressions with Epanechnikov kernel and bandwidth equal to 50.

significant determinant of household income, the spatial income profile documented above is robust to the inclusion of the share.[24]

Regional differences in nominal income don't translate into real income differences if higher nominal income is offset by higher prices or worse amenities. If real income is equalized across regions for these reasons, the rural-urban difference in nominal income might not be a sign of resource misallocation. A recent study by Gollin et al. (2017) however shows that a spatial equilibrium due to differences in amenities is very unlikely in a developing country like Malawi. Using detailed data from 20 African countries including Malawi, they find that amenities are constant or increasing in population density for almost all aspects of amenities

---

to construct employment shares because there is no detailed data on labor supply for all activities.

[24] A one percent increase of the share leads to a decline of log consumption per capita by 0.07, with a p-value less than 0.001.

they consider, such as housing quality and health, pollution and crime.[25]

If not amenities, can the differences in cost-of-living equalize real income across regions?[26] Many casual observations suggest that there is a big difference in cost of living between rural and urban areas in developing countries. Quantitative measures of that however are lacking, to some extent due to the lack of good price data. Most studies constructing spatial cost-of-living indexes do find that urban areas have larger cost-of-living, but the differences are far from explaining all the differences in nominal income (eg., Ravallion and van de Walle (1991); Brandt and Holz (2006)). I next turn to the evidence on prices from the Malawi data.

2.3.2 Food Prices

The rest of the section analyses the spatial profile of prices. The analysis of prices is to show that 1) spatial differences in nominal income can not be offset by differences in cost-of-living, and 2) rural areas have lower relative price of food. I first look at food prices in this subsection. Like many expenditure surveys, IHS3 asks both the value and quantity of purchases for over 100 food items. This allows me to calculate unit values as proxy of prices. To reduce the noise in unit value reported by individual households, I take the medium unit value in a region to be the the price prevailing in that region. Combined with average consumption shares of food items in each region, the prices are then used to construct a spatial food price index. I classify the rural EAs into four regions according their distance to the urban center and the urban EAs form a separate region. For each of the four urban centers, a separate Paassche price index is constructed for the city and the surrounding rural areas. Prices are normalized to 1 in the city. I report the average over the four cities in Table 2.1. The differences in food prices between rural and areas are broadly consistent with the findings of Deaton and Dupriez (2011) for Brazil and India. Within rural regions, there

[25]Housing quality and health are not amenity per se, but they use them as proxies for public good provision in the household's location.

[26]Gollin et al. (2017) look at real amenities. That is, they measure quantities which already take into account the effect of price differences. Since they don't consider all consumption goods, prices for other goods might still differ across regions to equalize welfare.

is also a decline in prices farther away from the city. Between the city and farthest rural area, there is an economically significant 17% price gap. Keeping nominal income constant, this means moving from the farthest area to the city could reduce real income by 8.5% given food's expenditure share of 0.5 in Malawi.[27]

Table 2.1: Spatial Price Differences

Regions	0	1	2	3	4
Distance to city	city	< 50	$[50, 100]$	$[100, 150]$	> 150
Price	1	0.91	0.87	0.86	0.83
Own-production	10.4	44.9	46.4	46.3	46.3

Note: Region 0 is the city, the rest are rural regions arranged in increasing distance from the city. The first row shows the distance to the city measured in kilometers. The second row is the Paasche price index for food. The third row is the percentage of food not purchased.

Two caveats regarding the use of unit value are worth mentioning. First, I use unit value to evaluate own-produced food. Table 2.1 reports that close to a half of food consumption in rural areas come from own production while that number is only 10% in cities. This could lead to an underestimation of the regional price differences if own-produced food is valued at lower prices by the family than purchased food. For example, Deaton and Dupriez (2011) find that the rural food price is substantially overestimated in Brazil and India when only using unit values from cash purchases. Second, unit values are known to be contaminated by quality differencesDeaton (1988). If high price goods also have high quality goods, this however leads to an over-estimation the regional price gap.

To construct the price index, I need to balance the number of item prices available and the number of regions, as smaller regions are more likely to recording zero purchase. Having a few regions probably misses the price differences within an aggregated region. I next look at the spatial profile of prices for individual food items. I make use of all unit values reported by regressing prices at household-item level on distance to the nearest urban

[27]Of course, this rough calculation doesn't consider the changes in other prices and the possibility of substitution.

center with a set of controls. The set of controls include the city dummies and the month when the household were visited by survey teams, which is supposed to capture the effect of seasonality. I experiment with two specifications, one includes the log of total expenditure and one does not. The inclusion of total expenditure is to correct for the quality differences in unit values.

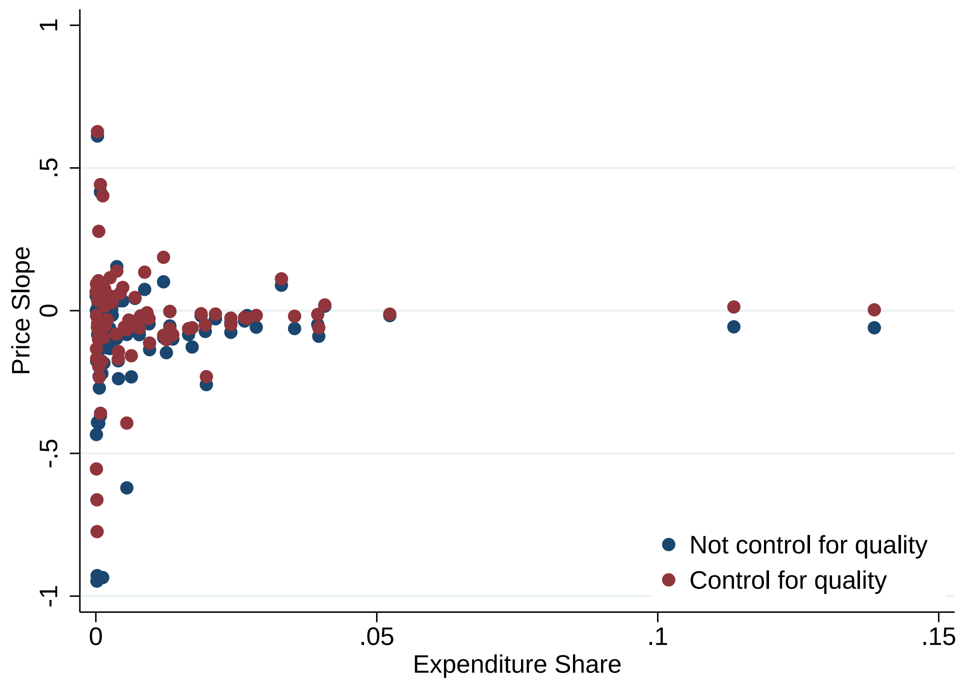
Figure 2.5 plots the distance estimates against average expenditure shares. Most food prices decline with distance from the city. It is especially so for those items with large expenditure shares. Adding total expenditure increases the estimates, indicating the existence of quality bias. However, the qualitative conclusion still holds. Looking at the individual items closely, items less likely to be locally produced such as "Tinned meat or fish", "Powered milk", "butter", "Sugar", "Cooking oil" all have positive estimates. Interestingly, while "Sugar" has a significant negative estimate, "Sugarcane" which is used to produce sugar has a significant positive estimate. This says rural areas tends to have lower prices for locally produced goods but higher prices for goods more likely to be exported from outside. As descried in the introduction, I emphasize the effect of regional price differences coming from the combination of regional specialization in production and costly trade. These findings support my presumption.

My focus is on the aggregate cost-of-living and the relative price of food. While the discussion of food prices is interesting, it is still only part of the story. As price is not recorded for non-food items in IHS3, I next turn to a more structural estimation of both the aggregate price index and relative price of food using information from Engel curves.

2.3.3 Price Estimates from Engel Curves

A recent literature pioneered by Hamilton (2001) and Costa (2001) uses information on shifts in Engel curves to estimate differences in cost-of-living over time or across space. The method relies on Engel's law, which states that food's budget share is inversely related to household real income, conditional on relative prices and other household characteristics. Movements in food's budget share then provides an estimate of real income, *ceteris paribus.*

Figure 2.5: Distance Slope of Prices and Expenditure Share



Note: I regress log of item prices on the distance to city with a set of controls. What's plotted is the estimated coefficient against the item's expenditure share. Both the results with total expenditure as a control and without are reported.

My method of estimating the aggregate price index and relative price of food builds on this literature.

I estimate two demand systems, one for food and one for individual food items. The demand system for food is assumed to be the Almost Ideal Demand System (AIDS): food's budget share is given by

$$\omega_{ij} = \phi + \beta(\ln Y_{ij} - \ln P_j) + \gamma(\ln P_j^F - \ln P_j^N) + X'_{ij}\theta + u_{ij},$$

where i and j indicates household and EA respectively, ω is food's budget share, Y is nominal income, P is the price index and subscripts F and N refer to food and non-food respectively, X is a set of household characteristics controlling for differences in preference between households, and u is a random error. Without information on prices, the model can be estimated as,

$$\omega_{ij} = \phi + \beta \ln Y_{ij} + X'_{ij}\theta + \sum_j \delta_j D_j + u_{ij}, \tag{2.1}$$

where D_j is an EA dummy and $\delta_j = -\beta \ln P_j + \gamma(\ln P_j^F - \ln P_j^N)$ contains information on both the cost of living and the relative price of food. Given the estimates for β and γ, I should be able to decompose δ_j into the two components if I have either the cost of living or the relative price.

The second demand system I estimate provides an estimate for $\ln P_j$. Following Nakamura et al. (2016), the item-level Engel curve is specified as,

$$\omega_{ij}^k = \phi^k + \beta^k(\ln Y_{ij} - \ln P_j) + \gamma^k(\ln P_j^k - \ln P_j) + X'_{ij}\theta^k + u_{ij}^k,$$

where k indexes item. For each item k, I construct at the EA level item price P_j^k as the medium unit value reported by households in that EA. Without variation in item prices within the EAs, this precludes the estimation of P_j from a single item. Nakamura et al. (2016) estimate the model non-linearly, pooling all items together. Since there are 768 EAs and many items, a non-linear estimation like theirs is computationally not feasible.[28] Another concern is the large amount of zero expenditures for each item in the data.[29] Not considering the zeros will lead to biased estimates. In line of these considerations, I propose the following three-step procedure to estimate P_j. The first two steps adopts the censored regression approach of Heien and Wesseils (1990) to correct for the zero expenditures. The demand system to be estimated is

$$\omega_{ij}^k = \phi^k + \beta^k \ln Y_{ij} + X'_{ij}\theta^k + \sum_j \delta_j^k D_j + u_{ij}^k, \tag{2.2}$$

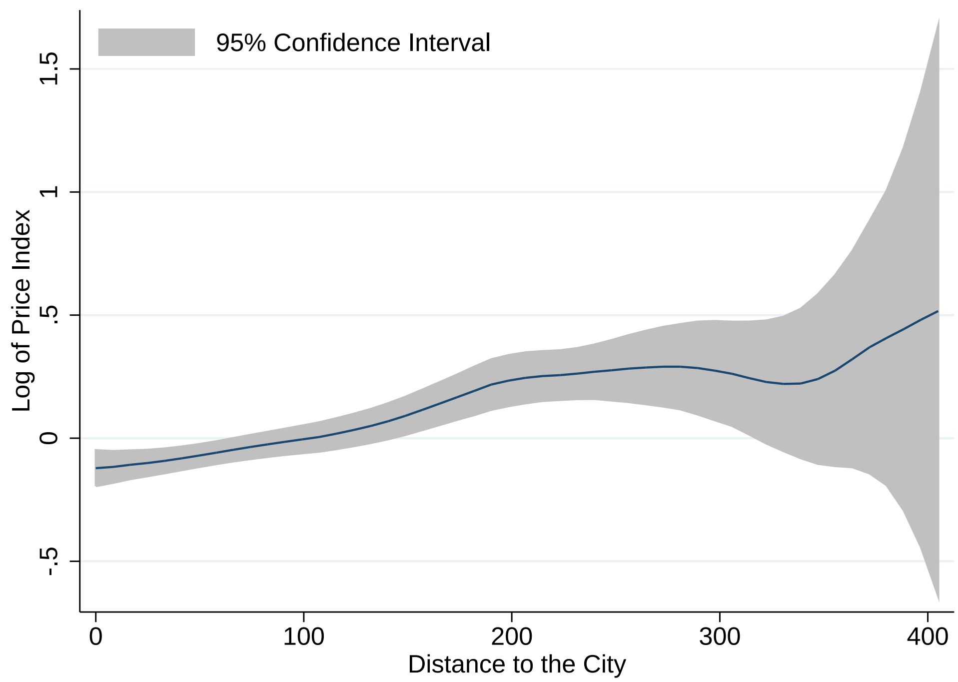
where $\delta_j^k = \gamma^k \ln P_j^k - (\beta^k + \gamma^k) \ln P_j$. In the first step, a probit model of whether or not zero expenditure is observed is estimated on the regressors of the demand equation for each

[28] Their study of Chinese inflation rate only considers 30 provinces.

[29] This could be due to household not consuming the good or misreporting. As the households only report food consumption during the last week, this could be also due to irregular purchasing.

item. The estimates are used to calculate the inverse Mills ratios.[30]. In the second stage, the item level demand equation 2.2 is estimated with the inverse Mills ratio as an additional control. In the last stage, the estimated δ_j^k are regressed on $\ln P_j^k$ without a constant term. The residuals can be viewed as estimates of $\ln P_j$. I average the residuals over item to form my estimates of aggregate price $\ln P_j$.[31] Finally, the relative price $\ln P_j^F - \ln P_j^N$ up to a constant γ is retrieved as $\delta_j + \beta \ln P_j$, where δ_j and β come from the Engel curve estimation for food.

Figure 2.6: Spatial Profile of Log of Cost-of-Living Index



Note: The price index comes from the Engel curve estimation described in the text. Log price is smoothed using the kernel-weighted local linear regressions with Epanechnikov kernel and bandwidth equal to 50.

I select ten food items according to their expenditure share and prices available.[32] The

[30]The inverse Mills ratio is $\frac{\phi(\cdot)}{\Phi(\cdot)}$ if expenditure is positive, and $\frac{\phi(\cdot)}{1-\Phi(\cdot)}$ if otherwise. $\phi(\cdot)$ and $\Phi(\cdot)$ are the predicted probability density and accumulative density for each observation

[31]This procedure leads to biased estimates of γ as $\ln P_j$ and $\ln P_j^k$ might be correlated. I note that my estimates of γ reported in the appendix are similar in magnitude compared to those of Nakamura et al. (2016). I am working on improve the estimation procedure.

[32]They are maize ufa mgaiwa (normal flour), maize ufa refined (fine flour), dried fish, Sugar, tomato, rice, nkhwani, brown bean, goat, and salt.

controls used in the estimation are household size and composition, and household head's characteristics.[33] Figure 2.6 plots the non-parametric fit of $\ln P_j$ against distance to the city. Surprisingly, price increases in distance. Cost-of-living do not offset the rural-urban income differences but makes the gap in real income wider. In view of the food price differences presented above, This could be the case if remote areas face really high prices for non-food goods. For example, most services are not traded across regions. As most studies find urban areas have slightly higher cost of living, this also points to the importance of accounting for quality differences in measuring prices, especially for the non-food items.[34,35] Another possibility is urban areas provide more varieties which lowers aggregate price. The rural-urban price difference is 0.11 log points, which is also economically significant. I freely admit that the magnitude should not be taken too seriously given the complexity in the estimation and measurement errors in data. The evidence does cast doubt on the role of cost of living in explaining the spatial income differences found above.

Figure 2.7 shows that $\gamma(\ln P_j^F - \ln P_j^N)$ declines in distance except the inaccurate estimates in the farthest areas. As most studies estimate γ to be positive (e.g., Hamilton (2001); Costa (2001)), this also implies relative price declines in distance as expected. For individual items, my estimates are also positive as shown in Section 2.9.1. If we take γ to be 0.05, which is roughly the middle point of the values found in the literature, this implies the biggest gap in relative price is around 1 log point. The measure of relative price is mechanically related to that of aggregate price with my estimation procedure. However, given the spatial profile of food price, aggregate price must be much higher in urban areas to reverse the pattern of relative price, which contradicts most estimates of rural-rural price differences in the literature.

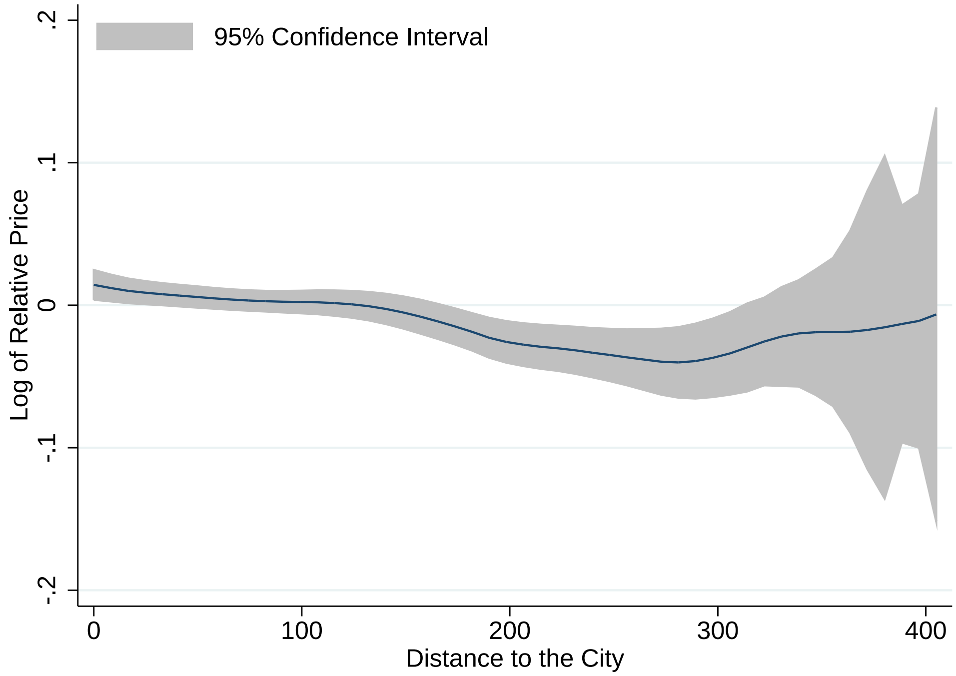
---

[33]Household composition is described as the ratio of children less than 15 years of age in the household. Household head's characteristics include sex, education, age, religion, marital status, language spoken, and a dummy indicating whether the household head is a farmer.

[34]Deaton and Dupriez (2011) find that quality differences explain some of the rural-urban food price differences in Brazil and India.

[35]A theoretical possibility is that rural areas export high quality goods to urban areas while import low quality goods given the income differences and quality is a luxury. Since transportation costs don't depends on quality, transportation costs can have a larger share in traded non-agricultural goods due to the differences in the value of goods.

Figure 2.7: Spatial Profile of Relative Price



Note: The relative price comes from the Engel curve estimation described in the text. Log relative price is smoothed using the kernel-weighted local linear regressions with Epanechnikov kernel and bandwidth equal to 50.

2.3.4 Taking Stock

Taking all the evidence together, there are two main points to take away. First, rural areas have much lower income per capita than urban areas, which also declines in the distance to the city. Spatial differences in cost of living is less likely to offset the differences in nominal income. Since the majority of Malawi's population lives in the rural areas, this spatial variation within rural areas is not a trivial issue. This is a new though not surprising finding. It might be explained by increasing migration costs for workers farther away from the city. Worker selection as in Lagakos and Waugh (2013) and Young (2013) however cannot easily explain the differences within rural areas. In the next section I show that workers might sort within rural areas due to the spatial differences in relative price, which is the second main point from this section.

My finding of higher relative price of food in urban areas echos a recent literature emphasizing the role of transportation costs in distorting the spatial allocation of workers in developing countries (Adamopoulos, 2011; Herrendorf et al., 2012; Gollin and Rogerson, 2014). For example, Adamopoulos (2011) find that transportation costs in the 5% poorest countires are 2.74 times higher than those of the 5% richest countries. From this point of view, my price estimates from Engel curve estimation might not seem so surprising. In reality, price determination depends much more than an aggregate measure of transportation costs. The measured price differences probably are more relevant in determining the spatial allocation of workers than rough measures of transportation costs.

2.4 A Model of Transportation Costs and Ability Sorting

This section presents a model to explain the spatial distribution of income discussed above. The income inequality within rural areas motivates a multi-region model, which also helps me to test the model implications using the Malawi Data in Section 2.6. The main idea of the model is that spatial prices differences due to transport costs can lead to sorting of workers across regions when food consumption is a necessity. Remote rural regions will have relatively cheaper foods, which makes them more attractive to low ability/low income workers with large expenditure share of food. Higher ability workers choose regions closer to the city because they focus more on the consumption of non-food good. The model has a similar geographic structure to Gollin and Rogerson (2014) who study the allocation of workers between an urban center and several rural regions in presence of transportation costs. It also follows a long tradition of von Thünen's model of agricultural land use (e.g., Nerlove and Sadka (1991)). In the model workers have sector-specific productivities as Lagakos and Waugh (2013), such that comparative advantage is also a determinant of the sectoral choice of workers.[36]

The economy features an urban center surrounded by a group of rural regions. I index

[36]Sector-specific productivity is not essential in inducing ability sorting due to spatial price differences. However, if individual productivity is homogeneous across ectors, there will be perfect sorting across all regions, as long as transport costs are not zero. With heterogeneous productivity, transportation costs play a role in determining the sectoral choice of workers.

the regions by $j = 0, 1, 2...J$, such that region 0 is the urban center, and the rest are rural regions. The urban center produces a non-agricultural good. All rural regions produce a homogeneous agricultural good. There will be trade between the urban center and the rural regions, but not among the rural regions. Trade is costly. For a unit of agricultural good transported from rural region j, only τ_a^j units reach the urban center. Similarly, for a unit of non-agricultural good transported to region j, τ_n^j units arrive. Goods prices in region j thus are given as

$$P_a^j = \tau_a^j P_a^0, \text{ and } \tau_n^j P_n^j = P_n^0. \tag{2.3}$$

The rural regions are arranged such that trade costs are increase in j. That is, τ_a^j and τ_n^j are decreasing in j. The relative price of agricultural good $\frac{P_a^j}{P_n^j} = \tau_a^j \tau_n^j \frac{P_a^0}{P_n^0}$ is hence decreasing in j. Remote regions provide cheaper food. Finally, each rural region is endowed with T_j units of land which is used in agricultural production.

There is a continuum of workers. Each is endowed with a pair of individual productivity $\{z_a, z_n\}$, representing the efficiency of the worker in agriculture and non-agriculture. The pair of productivity is draw from a joint distribution $G(z_a, z_n)$ with $z_a > 0$ and $z_n > 0$. The production of the non-agricultural good uses only labor according to,

$$Y_n^0 = AL_n^0, \tag{2.4}$$

where A is productivity, $L_n^0 = \int_{i \in \Omega^0} z_n^i dG(i)$ is the labor input measured in efficiency units. Profit maximization leads to

$$w^0 = P_n^0 A. \tag{2.5}$$

Agricultural production uses both labor and land. All regions have access to the same technology, such that production in region $j > 0$ is given by

$$Y_a^j = A(L_a^j)^\alpha (T_j)^{1-\alpha}, \tag{2.6}$$

where $L_a^j = \int_{i \in \Omega^j} z_a^i dG(i)$ is the labor input in region j. Profit maximization leads to,

$$w^j = \alpha P_a^j A \left(\frac{T_j}{L_a^j}\right)^{1-\alpha}, \tag{2.7}$$

$$q^j = (1-\alpha) P_a^j A \left(\frac{L_a^j}{T_j}\right)^{\alpha}, \tag{2.8}$$

where q^j is the rental rate of land in region j.

Worker's preference over the agricultural goods C_a and non-agricultural good C_n is defined as follows,

$$U = \left(\frac{C_a - \bar{a}}{\mu}\right)^{\mu} \left(\frac{C_n}{1-\mu}\right)^{1-\mu}, \tag{2.9}$$

where $\bar{a} > 0$ is the subsistence requirement, and μ is the weight on agricultural consumption. While the Stone-Geary preference or its extension (Herrendorf et al., 2013) have been wildly used in the structural transformation literature, it has the unsatisfying property of being homothetic asymptotically.[37] My analysis however is not driven by the Stone-Geary assumption. In Section 2.9.3 I show that ability sorting holds under any non-homothetic preference that treat agricultural good as a necessity.[38] I choose the Stone-Geary preference for its analytical tractability.

A worker chooses a region to maximize utility, based on her productivity and local wage and prices. Given the location choice, the budget constraint reads,

$$P_a^j C_a + P_n^j C_n = I^j, \tag{2.10}$$

where I^j is the worker's income in region j. In the urban center, workers only receive wage income. In rural regions, workers receive wage income and a share of land rents. The land rents are distributed across workers within a rural region. Each worker receives a share proportional to her wage income. This amounts to workers receive all agricultural output

[37] The income effect is important in understanding structural transformation even in rich countries (Boppart, 2014), which is not possible with the Stone-Geary preference.

[38] The proof assumes homogeneous ability, so it corresponds to the within rural area ability sorting. Sector choice under general preference is much harder to analyze when workers have sector-specific productivity.

they produce. This assumption in essence mimics that of the classical dual economy model of Lewis (1954): workers receive their average product. It is not so unrealistic. In developing countries, rental markets for land are often missing, and most of the land is owned by the family but without land titles (Chen, 2016).[39] In this model, it results in worker income proportional to their productivity, which simplifies the analysis below. But it does not drive the results.[40] Income in region j hence is summarized as,

$$I^j = \begin{cases} w^j z_n, & \text{if } j = 0, \\ \frac{1}{\alpha} w^j z_a = P_a^j y_a^j z_a, & \text{otherwise,} \end{cases} \tag{2.11}$$

where $y_a^j = \frac{Y_a^j}{L_a^j}$ is the agricultural output per efficiency unit in region j.

Given the location choice and assuming $I^j > P_a^j \bar{a}$,[41] optimal consumption decision is simply given as,

$$P_a^j C_a = \mu I_j + (1-\mu) P_a^j \bar{a} \tag{2.12}$$

$$P_n^j C_n = (1-\mu)(I^j - P_a^j \bar{a}) \tag{2.13}$$

[39]In the Malawi data, only 7% of land plots are rented. Most land plots are either inherited (73.6%) or granted by local leaders (11.6%).

[40]We can otherwise assume land rents are accrued to land owners who consume locally. Ability sorting among workers still holds. If all workers receive an equal share of land rents irrespective of their location, a sufficient condition for the results to hold is $R < P_a^j \bar{a}$ where R is the total land rents. This condition is likely to hold given I calibrate land rents to be only 18% of agricultural output while subsistence consumption is 79% of agricultural output in the United States. The ability sorting results will be reversed if only the distribution of land rents favors the near regions over remote regions and land rents are large relative to food consumption. This is because lower ability farmers will value the fixed rental income more than high ability farmers. This however is highly unrealistic.

[41]Otherwise only the agricultural good is consumed, we have $P_a^j C_a^j = I^j$ and $C_n^j = 0$. How the utility is defined under this case doesn't affect the equilibrium outcome. It only matters for welfare analysis. I will come back to this point below.

which leads to the indirect utility function,

$$V(j, z_a, z_n) = \begin{cases} \frac{w^j z_n}{\left(P_a^j\right)^{\mu}\left(P_n^j\right)^{1-\mu}} - \bar{a}\left(\frac{P_a^j}{P_n^j}\right)^{1-\mu}, & \text{if } j = 0, \\ \frac{w^j z_a}{\alpha\left(P_a^j\right)^{\mu}\left(P_n^j\right)^{1-\mu}} - \bar{a}\left(\frac{P_a^j}{P_n^j}\right)^{1-\mu}, & \text{otherwise.} \end{cases} \tag{2.14}$$

The indirect utility for a worker residing in region j is a linear function of her productivity with the intercept being determined by the relative price and the slope being a real wage measure. The simple form of the indirect utility function makes the worker's location choice easy to analyze. The problem can be broken down into two sub-problems: the location choice among rural regions and the sector choice between rural regions and the urban center. I analyze them in turn.

Location Choice between Rural Regions The linear indirect utility implies that remote regions have an advantage in attracting workers due to lower relative price of the agricultural good. It must be counteracted by a lower real wage (income) if all regions are populated.[42] We thus have real wage $\frac{w^j}{\left(P_a^j\right)^{\mu}\left(P_n^j\right)^{1-\mu}}$ decreasing in j for the rural regions.[43] It is easy to see that under this scenario, if worker i chooses a near region over a remote region, any agent with a higher z_a will also prefer the near region. This means the choice between rural regions can be described by a set of cutoffs $\{\tilde{z}_a^j\}_{j=1,...J-1}$, where the worker with $\tilde{z}_a^j$ is indifferent between region j and $j+1$.

$$\frac{w^j \tilde{z}_a^j}{\alpha\left(P_a^j\right)^{\mu}\left(P_n^j\right)^{1-\mu}} - \bar{a}\left(\frac{P_a^j}{P_n^j}\right)^{1-\mu} = \frac{w^{j+1} \tilde{z}_a^j}{\alpha\left(P_a^{j+1}\right)^{\mu}\left(P_n^{j+1}\right)^{1-\mu}} - \bar{a}\left(\frac{P_a^{j+1}}{P_n^{j+1}}\right)^{1-\mu}, \forall j > 0. \tag{2.15}$$

The presence of non-homothetic preference and regional price differences leads to perfect worker sorting: higher productivity workers sort into higher wage regions and pay higher price for the agricultural good. Figure 2.8 illustrates this for three rural regions, where the allocation of workers between rural regions is fully described by two cut-offs. I summarize

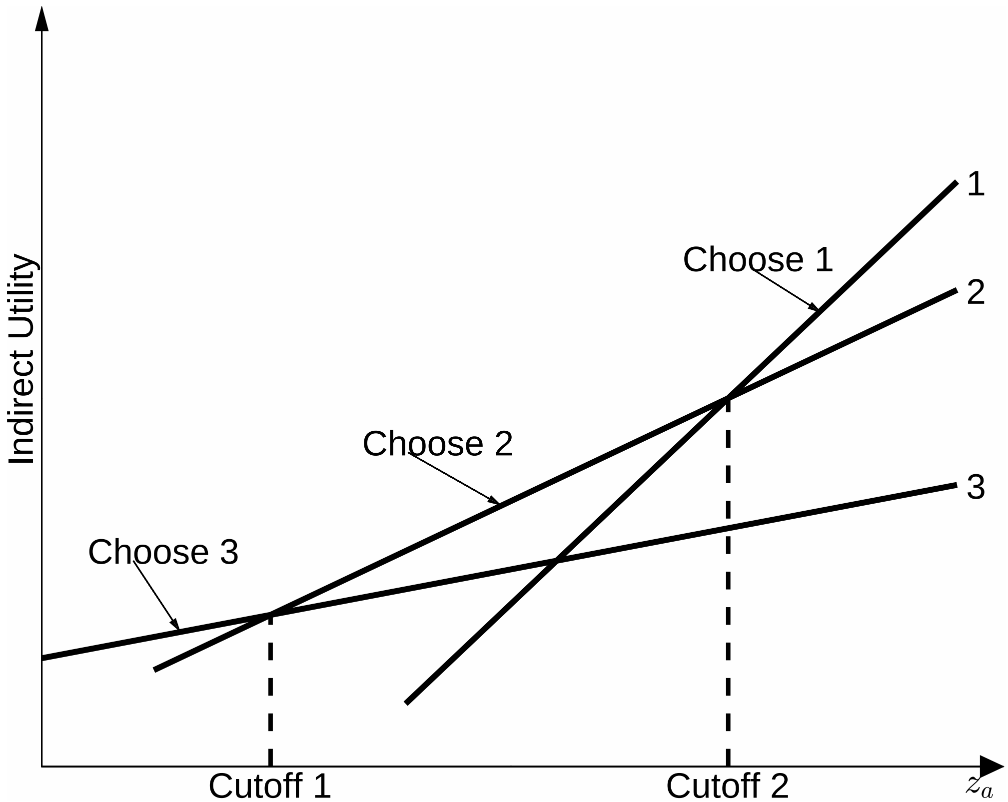
[42] All regions must be occupied due to the presence of a fixed factor.

[43] Technically, $\frac{w^j}{\left(P_a^j\right)^{\mu}\left(P_n^j\right)^{1-\mu}}$ is not a satisfying real wage measure, as different workers have different price indexes when preferences are non-homothetic. I call it real wage nevertheless for convenience.

these results in the following proposition.

Proposition 1 *Among rural regions, regions with lower relative price of the agricultural good: 1) pay lower real wages, measured as nominal wage deflated by a price index with fixed weights; 2) attract agricultural workers with lower productivity.*

Figure 2.8: Location Choice Between Rural Areas



Lee (2010) show that ability sorting can rise between cities when big cities provide more product varieties but also have higher rents. Since the housing demand is fixed, big cities attract more high ability (income) worker because they value the consumption variety more than low ability (income) workers. Sorting is not perfect because high ability and low ability workers complement each other in production in all cities. This results in lower return for high ability workers in big cities as low ability workers must be compensated for the high housing price. For similar logic, Black et al. (2009) also show that cities with higher housing prices should have lower return to education, which they find support in U.S. data. They however assume both types of workers live in all cities, without considering the location choice problem. In my model, the disadvantage of near regions in food price is compensated by higher real income. In developing countries, food consumption is more important than housing consumption, while housing consumption might still play a role. The Malawi data shows that food accounts for over 50% of total expenditure while housing rent is negligible. Engel's law stating that the expenditure share of food is declining in income is also a robust feature of data, see eg. Houthakker (1957). Spatial price differences induced by transport costs then imply workers are sorted across regions, which should help explain the spatial income profile document in Section 2.3.

From the profit maximization conditions of agricultural firms, the wage rate in rural regions is closely related to average product, $w^j = \alpha P_a^j y_a^j$. Plugging it back into the indirect utility function, we have

$$V(j, z_a, z_n) = \left(\frac{P_a^j}{P_n^j}\right)^{1-\mu} \left(y_a^j z_a - \bar{a}\right), \forall j > 0.$$

Since the relative price of food is decreasing in j, we must have y_a^j increasing in j, otherwise no worker will choose the remote areas. As the production technology is the same everywhere, it must be the case that remote areas have more land per unit of efficiency labor. Since agricultural workers are sellers of the agricultural goods, they naturally prefer higher relative price of food. For them to accept a lower relative price of food in remote areas, they must be compensated by higher land intensity. The rental rate of land must be lower in remote areas as land intensity is high while food price is low. I summarize these results in the following proposition.

Proposition 2 *More remote rural areas have higher land-labor (in efficiency units) ratios and lower land rental rates.*

The model however doesn't have an unambiguous prediction regarding how nominal wages vary between regions. In a von Thünen model with Cobb-Douglas preferences and homogeneous workers, Nerlove and Sadka (1991) show that nominal wages can be declining in the distance from the urban center, only if remote regions in general have a lower price index. My model also has a similar property. As $\frac{w^j}{\left(P_a^j\right)^{\mu}\left(P_n^j\right)^{1-\mu}}$ is decreasing in j, nominal wage can be declining in j, if the increase in the price of the non-agricultural good is small enough compared to the reduction in the price of the agricultural good.[44] If the condition is not satisfied, the model allows nominal wage to be increasing in j or behave randomly as long as real wage is declining in j. This indicates that for workers around the boundary $\tilde{z}_a^j$, workers in nearer regions might have lower nominal income. Perfect ability sorting doesn't necessarily mean there is perfect separation in nominal income between regions, a prediction which is counterfactual.

Another way to look at nominal wage is to use the fact that $w^j = \alpha P_a^j y_a^j$. As P_a^j is decreasing in j and y_a^j is increasing in j, nominal wage is also decreasing in j if output per efficiency unit doesn't increase fast enough. This again requires the price of non-agricultural good doesn't increase too fast as a small increase in output is enough to make the marginal worker indifferent. The model doesn't have an unambiguous prediction for output per worker either. As we go farther away from the urban center, y_a^j is increasing while worker's productivity is decreasing, output per worker in a region might decrease or increase. The same result applies to land per worker. Income per worker also changes in an ambiguous way as it is a product of P_a^j, y_a^j, and average worker efficiency. It is more likely to be decreasing in j than nominal wage and output per worker, which explains the spatial income profile documented in Section 2.3.

The expenditure share of food doesn't have a clear spatial pattern. It is given by

$$\text{food share} = \mu + (1-\mu)\frac{\bar{a}}{y_a^j z_a} \tag{2.16}$$

It is lower for higher ability agricultural workers. But these workers choose nearer areas where average agricultural output per efficiency unit is lower, which raises the expenditure share. The fact that it is increasing in the distance to city again requires the difference in ability to play a dominant role.

Sector Choice I next consider the choice between rural and urban regions. Let the optimal

---

[44] Notice that the Engel curve evidence in Section 3 points to the opposite.

choice between rural regions be j^*, had the worker chosen agriculture. The non-agricultural sector will be chosen if,

$$\frac{w^0 z_n}{\left(P_a^0\right)^\mu \left(P_n^0\right)^{1-\mu}} - \bar{a}\left(\frac{P_a^0}{P_n^0}\right)^{1-\mu} > \frac{w^{j^*} z_a}{\alpha\left(P_a^{j^*}\right)^\mu \left(P_n^{j^*}\right)^{1-\mu}} - \bar{a}\left(\frac{P_a^{j^*}}{P_n^{j^*}}\right)^{1-\mu}$$

The location choice can be rewritten as,

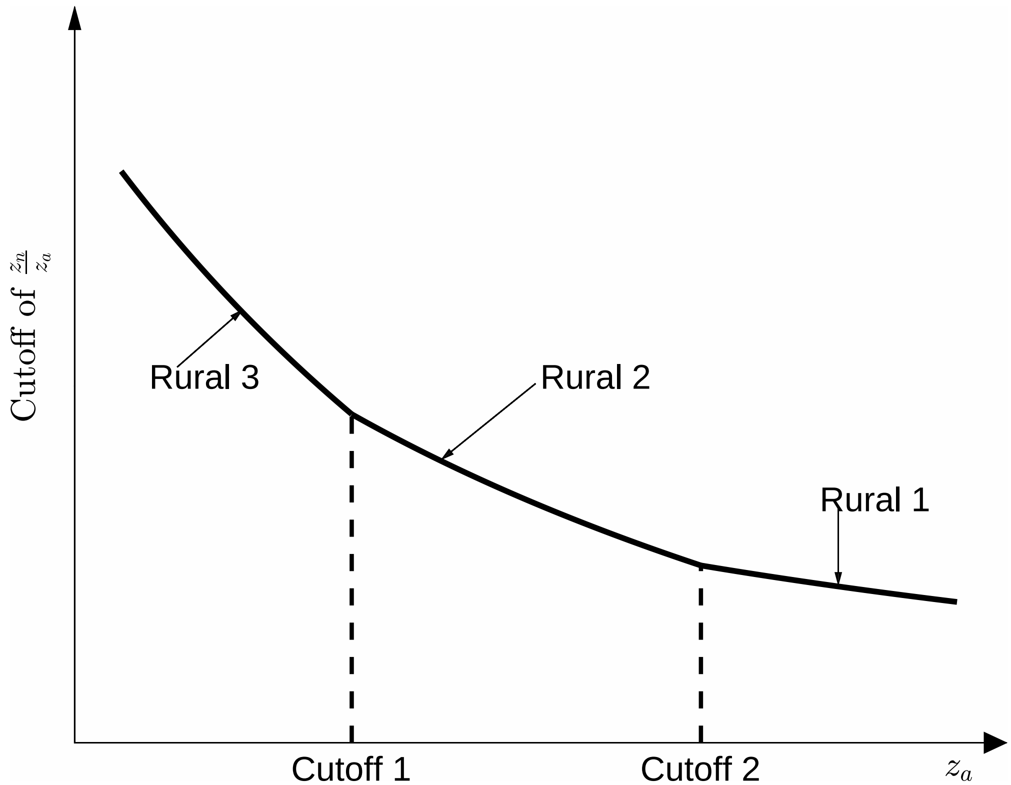
$$\frac{\bar{w}^0}{\bar{w}^{j^*}}\frac{z_n}{z_a} > \frac{1}{\alpha} + \frac{\bar{a}\left[1 - (\tau_a^{j^*}\tau_n^{j^*})^{\mu-1}\right]\left(\frac{P_a^0}{P_n^0}\right)^{1-\mu}}{\bar{w}^{j^*} z_a}, \tag{2.17}$$

where $\tau_a^{j^*}$ and $\tau_n^{j^*}$ are trade costs of region j^* as defined above, and $\bar{w}^j = \frac{w^j}{\left(P_a^j\right)^\mu \left(P_n^j\right)^{1-\mu}}$ is the real wage rate. This condition is different from the standard Roy model where sector choice is made to maximize nominal wages (Lagakos and Waugh, 2013) for three reasons. First, when prices differ between sectors, what matters is real wages instead of nominal wages. This is reflected in the LHS of 2.17. Second, workers receive average product instead of marginal product in agriculture, as seen in the first term on the RHS of 2.17. This raises the bar for the workers to choose non-agriculture.[45] Third, the second term on the RHS of Eq. 2.17 captures the interaction between subsistence consumption and spatial price differences. This interaction induces the sorting of better agricultural workers to non-agriculture. For each z_a, the sector choice 2.17 can be perfectly described by a cutoff of z_n, denoted as $\tilde{z}_n(z_a)$. The higher z_a is, the lower the ratio of $\tilde{z}_n(z_a)$ to z_a. It is thus harder to induce low ability agricultural workers to choose non-agriculture. This is due to two effects. First, there is a direct effect as low ability workers with low income value low price of agricultural good more. Second, there is the indirect effect as low ability workers choose remote regions with even lower prices of agricultural good. This is depicted in Figure 2.9 where there are three rural regions. I summarize this result in the following proposition.

Proposition 3 *The ratio $\frac{z_n(z_a)}{z_a}$ which describes the rural-urban choice for a worker with agricultural productivity z_a is decreasing in z_a.*

[45] It can also be viewed as a source of labor market frictions that lowers agricultural wages relative to non-agricultural wages.

Figure 2.9: Rural-Urban Choice



Given the description of the location choice, labor supply in the urban center is given as follows.

$$L_n^0 = \int_0^\infty \int_{\tilde{z}_n(z_a)}^\infty z_n dG(z_a, z_n). \tag{2.18}$$

In rural region j, labor supply is given as

$$L_a^j = \int_{\tilde{z}_a^{j-1}}^{\tilde{z}_a^j} \int_0^{\tilde{z}_n(z_a)} z_a dG(z_a, z_n), \tag{2.19}$$

where I assume $\tilde{z}_a^0 = 0$ and $\tilde{z}_a^J = \infty$ to save on notation. The number of workers in each regions are given as

$$N^j = \begin{cases} \int_0^\infty \int_{\tilde{z}_n(z_a)}^\infty dG(z_a, z_n), & \text{if } j = 0, \\ \int_{\tilde{z}_a^{j-1}}^{\tilde{z}_a^j} \int_0^{\tilde{z}_n(z_a)} dG(z_a, z_n), & \text{otherwise.} \end{cases} \tag{2.20}$$

We are now ready to define a competitive equilibrium for this economy.

Definition 1 *The competitive equilibrium is defined as as a set of good prices* $\{P_a^j, P_n^j\}_{j=0...J}$, *wages $\{w^j\}_{j=0\ldots J}$, and land rental rates $\{q^j\}_{j=1\ldots J}$ that: 1) clear all the markets given worker's and firm's optimal decision take the prices as given, and 2) satisfy the relationship between prices described by 2.3.*

Factor market clearing is given by equating the factor demand 2.5, 2.7, and 2.8 to labor supply 2.18 and 2.19, and land supply T_j. Given Walras's law, I normalize $P_n^0 = 1$ and consider only the market clearing condition for agricultural good,

$$\mu \frac{P_n^0}{P_a^0} Y_n^0 + (1-\mu) N^0 \bar{a} = \sum_{j=1}^{J} \frac{(1-\mu)(Y_a^j - N^j \bar{a})}{\tau_a^j}, \tag{2.21}$$

where I have made use of the aggregation property of the linear demand system implied by the Stone-Geary preference.

2.5 The Quantitative Analysis

2.5.1 Calibration and Model Performance

This section quantifies the model and evaluates the importance of ability sorting. The purpose is to understand how adding spatial price differences reinforces the selection effect of Lagakos and Waugh (2013). The preference and technology parameters are calibrated to U.S. data. For the U.S., it is assumed that transport costs are zero such that all workers face identical prices. The model reduces to the baseline model of Lagakos and Waugh (2013) except that agricultural production uses land. Following them, the joint distribution of individual productivity draw is assumed to be

$$G(z_a, z_n) = C[F(z_a), H(z_n)],$$

$$\text{where } F(z_a) = \exp^{(z_a)^{-\theta_a}} \text{ and } H(z_n) = \exp^{(z_n)^{-\theta_n}},$$

$$\text{and } C[u,v] = -\frac{1}{\rho} \log \left[1 + \frac{(\exp^{-\rho u} - 1)(\exp^{-\rho v} - 1)}{\exp^{-\rho} - 1} \right].$$

The parameters θ_a and θ_n controls the dispersion of the productivity draws and ρ determines the correlation of the productivity draw. Using U.S. data, Lagakos and Waugh (2013) find $\theta_a = 5.3$, $\theta_n = 2.7$, and $\rho = 3.5$. These numbers are also used here. Compared to Lagakos and Waugh (2013), I have land in my model. The land share $1 - \alpha$ is set to be 0.18, according to the estimate of Valentinyi and Herrendorf (2008) for U.S. agriculture. Land endowment is set to be 3.39 to match a land per agricultural worker of 169.3 acre (Adamopoulos and Restuccia, 2014). The preference parameters are set to target a long-run agricultural employment share of 0.5 percent and a current agricultural employment share of 2 percent, which leads to $\mu = 0.0037$ and $\bar{a} = 5.54$. Productivity A is normalized to 100.

The transportation costs are calibrated to the Malawi economy. Similar to Gollin and Rogerson (2014) who also focus on the spatial misallocation induced by transportation costs, I assume there are only two rural regions. Total land is set to match the gap in arable land per worker between Malawi and U.S. in Restuccia et al. (2008), $T^{MWI}/T^{US} = 0.73/1.62$. In line with previous discussion on spatial price differences, I assume the first rural region corresponds to the area within 50 kilometers from an urban center and the second rural region includes all the rest rural areas. Transportation costs are assumed to be the same for both goods. I set $\tau_a^1 = \tau_n^2 = 1.1$ and $\tau_a^2 = \tau_n^2 = 1.5$, which are broadly consistent with the Engel curve evidence in Section 2.3 and in line with the numbers used by Gollin and Rogerson (2014). In the data, total land cultivated by farmers in the remote region is slightly larger, which leads to $T_1 = 0.45T^{MWI}$ and $T_2 = 0.55T^{MWI}$. Aggregate productivity A is then calibrated to match the aggregate income difference between Malawi and U.S., which requires $A^{MWI} = 6.37$.

Table 2.2 presents the simulation results for the aggregate economy. To emphasize the role of ability sorting, the model is contrasted to the model of Lagakos and Waugh (2013). Since the Lagakos and Waugh (2013) model doesn't have land, I first introduce land to their model then add transportation costs. Both sets of results are reported. Compared to Lagakos and Waugh (2013), adding land differences significantly increase the explaining power for sectoral productivity and the employment share of agriculture. However, it helps

Table 2.2: Model Performance: Aggregates

	Data	L&W (2013)	+land	+sorting
U.S./Malawi ratio				
Aggregate productivity	28.8	28.8	28.8	28.8
Ag productivity	88.2	39.5	54.0	53.3
Non-ag productivity	7.9	17.2	8.8	8.0
Malawi				
Ag employment	85.4%	58.6%	78.8%	80.8%
Rural-urban income gap	3.01	1.56	1.60	1.95

Note: The productivity data come from Restuccia et al. (2008). Rural-urban income gap comes from my calculation using the Malawi data.

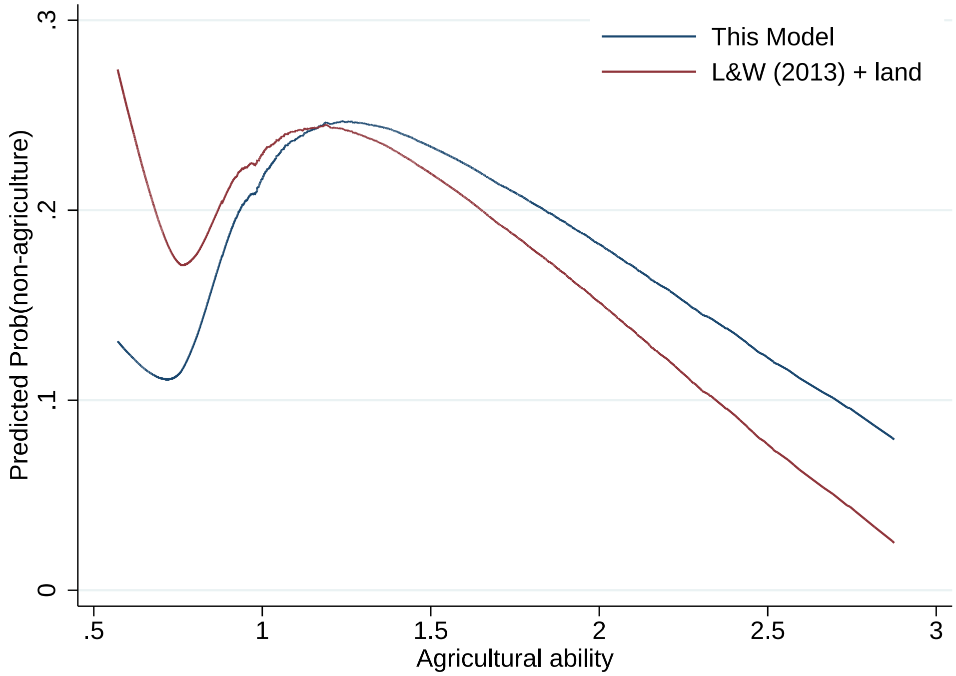
little in terms of explaining the rural-urban income gap.[46] Adding land to Lagakos and Waugh (2013) increases the explained percentage of APG relative to U.S. from 11.7% to 15.1%.[47] Ability sorting increases that percentage to 42.7%, a significant improvement.

Ability sorting explains the productivity gap relative to U.S. in non-agriculture slightly better, but becomes slightly worse in explaining that in agriculture. This is because I hold the aggregate productivity gap constant. With the new model, aggregate efficiency A is higher than without ability sorting. This raises productivity in both sectors. Average worker ability decreases in agriculture and increases in non-agriculture due to ability sorting. Figure 2.10 plots the probability of choosing non-agriculture for workers. Low ability agricultural workers are less likely to choose non-agriculture in my model compared to Lagakos and Waugh (2013), while the opposite is true for high ability agricultural workers. Overall, workers

[46]The agricultural productivity gap (APG) is probably a more suitable measure for my model. Malawi has an unusually high APG. According to raw national accounts data, the gap is 12.5 in 2005, which reduces to 6.2 after adjusting for measurement errors and human capital (Gollin et al., 2014). This is high even compared to countries of similar income levels. For example, according to Restuccia et al. (2008), Malawi's GDP per capita is 1171 in 2005 international dollars. Uganda's GDP per capita is 1224 while its APG is only 2.49 after adjustment. In the Gollin et al. (2014) sample, only one country (Lesotho) has an APG higher than Malawi. In line of this, I believe the rural-urban income gap calculated in data probably gives a better description of the actual rural-urban differences.

[47]The model is calibrated to explained the U.S. APG which is 1.43. The explained percentage is calculated as $\frac{\log(\text{predicted APG})-\log(\text{U.S. APG})}{\log(\text{actual APG})-\log(\text{U.S. APG})}$.

Figure 2.10: Sector Choice in the Model



Note: The probability is estimated from regressing actual sector choice on agricultural ability using robust locally weighted scatter plot smoothing.

choose agricultural more, which leads to a 2 percentage points increase in the agriculture employment share.

I next contrast the within rural area model predictions to data. This could be a test of the model because my calibration doesn't target these data moments. As seen in Table 2.3, the model overestimates the employment share in the near rural region but underestimates the employment share in the remote rural region. This is reasonable since the one city assumption will overestimate the attractiveness of the near region when remote regions actually have access to other cities as well. Even though the model overestimates the attractiveness of the near region, it correctly predicts that the employment to land ratio in the near region is smaller than in the remote region. This is in direct contrast to Gollin and Rogerson (2014) who study a similar model without heterogeneous ability. If workers are homogeneous, the near region will have a larger employment to land ratio to counteract the benefit of being closer to the urban center. Otherwise, all workers will flow to the near region. In my model, what matters is the efficiency labor unit instead of number of workers. Ability sorting allows

Table 2.3: Model Performance: Within Rural Area

	Data	Model
Employment share		
Urban region	14.8%	19.2%
Rural region 1	27.9%	39.2%
Rural region 2	57.3%	41.7%
Urban/Rural income ratio		
Region 1	2.67	1.45
Region 2	3.24	2.74

the near region to have a lower employment to land ratio but higher labor (in efficient units) to land ratio. The model correctly predicts a lower income gap in the near region, which is a motivation for the model. The predicted within-rural area income gap is larger than that in data, again reflecting the over-estimation of the attractiveness of the near region. Overall, the model is more successful in explaining the income gap between the urban center and the remote region than explaining that between the urban center and the near region. This somehow suggests that transportation costs between the near region and the urban center might be underestimated.

2.5.2 Counterfactual Experiments

I next run a set of counterfactual experiments with the model. I focus on how changes in productivity and transportation costs affect the equilibrium and the associated welfare effects. Since the ability sorting mechanism depends both on aggregate productivity and transportation costs, the experiments also evaluates the relative importance of them. I consider the following sets of changes: 1) reduce transportation costs in the near region by 10%; 2) reduce transportation costs in the near regions by 10%; 3) increase the land share in the near region to 0.5, and 4) raise A by 10%. The first three can be considered as infrastructure development targeted at different regions, while the last can be viewed as an improvement in production efficiency coming from other sources.

Since utility is not defined when food consumption is below subsistence, I adopt the following transformation of the Stone-Geary preference used above,

$$U = \begin{cases} C_a, & \text{if } C_a < \bar{a} \\ \bar{a} + \frac{\left[\left(\frac{C_a - \bar{a}}{\mu}\right)^{\mu}\left(\frac{C_n}{1-\mu}\right)^{1-\mu}\right]^{\sigma} - 1}{\sigma}, & \text{if } C_a \geq \bar{a} \end{cases} \tag{2.22}$$

where $0 < \sigma < 1$ ensures that utility is strictly concave when $C_a \geq \bar{a}$ and that it increases when the consumption of the non-agricultural good goes from zero to positive. Following Gollin et al. (2007), I choose $\sigma = 0.0001$ to approximate a log utility. Under the transformation, the positive results still hold as consumption behavior and location choice is not affected. My welfare measure is the equivalent variation defined as the percentage change in welfare that satisfies

$$V(I', P_a', P_n') = V((1+x)I, P_a, P_n), \tag{2.23}$$

where prime indicates variables after the change. Given the non-homothetic preference, welfare gains differ across workers even they are perfectly mobile. I document the average welfare change for all workers and for workers in the bottom 20% and 40%. Focusing on workers in the bottom is motivated by Dollar and Kraay (2002); Dollar et al. (2016) who study whether growth is shared among the poor.

Table 2.4 presents the simulation results. For the first experiment, reducing transportation costs in the near region raises agricultural productivity but lowers non-agricultural productivity. This is because workers migrate in response to their comparative advantage due to lowering spatial price differences such that the non-agricultural sector on average has worse workers. Consistent with that, rural-urban income gap has a 5.8% drop. The size of the drop in non-agricultural productivity is larger than the increase in agricultural productivity, implying the reduction in transportation costs actually make the misallocation within rural areas worse. Overall, this leads to an increase in aggregate productivity less than 2%. The urban center and the near rural region both experience increase in employment at the expense of the remote region. Welfare gains benefit the bottom workers less than others. Since most bottom workers locate in the remote region, this reflects that the reduction in

Table 2.4: Counterfactual Experiments

	Baseline	Ex 1	Ex 2	Ex 3	Ex 4
U.S./Malawi ratio					
Aggregate productivity	28.8	28.3	28.5	28.7	23.6
Ag productivity	53.3	53.2	54.0	53.1	48.0
Non-ag productivity	8.0	8.3	8.8	8.1	8.2
Malawi					
Urban Employment	19.2%	20.7%	19.7%	19.4%	25.7%
Rural 1 Employment	39.2%	40.8%	36.0%	42.2%	40.2%
Rural 2 Employment	41.7%	38.6%	44.3%	38.3%	34.1%
Rural-urban income gap	1.95	1.84	1.88	1.93	1.86
Welfare gain					
Overall		2.5%	0.6%	0.3%	11.4%
Bottom 20%		1.3%	-0.1%	0.0%	13.5%
Bottom 40%		1.2%	0.6%	0.0%	13.4%

the relative price of the agricultural good hurts net suppliers in the remote region, they are partly compensated by an increase in the land-labor (in efficiency unit) ratio. The bottom 20% workers fare slightly better than the bottom 40% workers as some of them are not selling their output and are not affected by the price change.

Reducing transportation costs in the remote region reduces both agricultural and non-agricultural productivity, though aggregate productivity only increases by around 1%. This again is due to the reallocation of workers from the near region to the urban center and the remote region. It reduces the average worker ability in the urban center as in the first experiment. It also increases the share of employment in the low productivity remote region. Aggregate productivity increases because the non-agricultural sector has a larger share of employment. Rural-urban income gap is reduced by a little over 2%. The welfare gains are much smaller than that in the first experiment. Bottom workers still benefit less than the average workers. The bottom 20% workers even suffer a welfare loss due to the reduction. This reflects the reduction in land-labor ratio in the remote region, which hurts the agricultural workers in subsistence. Net suppliers in the remote region are compensated by a higher relative price of the agricultural good even though relative price has decreased in the urban center.

The third experiment, raising the land share of the near region, has similar effects on productivity as the first experiment, though with much smaller magnitude. Employment in the urban center increases a little, while there is a large shift of employment from the remote region to the near region. This is largely mechanical due to the increase in size of the near region. Employment density actually decreases in the near region relative to the remote region.[48] Due to the increase in the price of the agricultural good, the remote region becomes more attractive to low ability workers. The welfare gains are small, with the bottom workers experience almost zero gains on average. This again is caused by the decrease in land-labor ratio associated with the increase in employment density in the remote region.

The fourth experiment has the largest effects on the equilibrium outcome and welfare, because it is an economy wide change affecting all regions. Productivity growth reduces the importance of food consumption, which weakens the ability sorting channel. This again leads to a slight decrease in non-agricultural productivity despite the increase in production efficiency. Agricultural productivity increases by a little over 10% due the increase in production efficiency and the increase in worker efficiency, while aggregate productivity increases by almost 20%. The near region shows a small reduction in employment, while employment in the urban center increases by 5.5 percentage points at the loss of the remote region. Welfare gains are also large, with the average change being 11.4%. Different from the first three experiments, the bottom workers benefit more from the increase in production efficiency. Even though the price of the agricultural good decreases, they benefit from the increase in the land-labor ratio. Due to the increase in production efficiency and the increase in land-labor ratio, over 90% workers under subsistence are able to move out of subsistence.

The experiments I carry out do not consider the cost side at all. The results from these

[48] In the table, the employment share of the remote region has declined. But note that the land share has also decline.

experiments are thus not sufficient to guide infrastructure investment. Nevertheless, the three experiments regarding transportation costs show very different effects on productivity and welfare from the other experiment. Reductions in transportation costs are more effective in reducing the regional income gap than raising sectoral productivity. They also benefit the bottom workers less than the rest, contrary to the a universal productivity growth. These findings echo the study of Gollin and Rogerson (2014) who also emphasize the geographic factors within rural area. It shows that studying how limited resources for infrastructure development should be allocated can be as important as simply promoting infrastructure development.

Dollar and Kraay (2002) and Dollar et al. (2016) argue that the poor benefit equally from economic growth as others by showing that income growth of the bottom workers are similar to that of all workers in data. My welfare analysis shows that there is more subtlety to the story. I find that infrastructure investment aiming at reducing transportation costs benefits the poor less or even hurt the poor, while that aiming at increasing aggregate production efficiency benefits the poor more than the others. The results of Dollar and Kraay (2002) and Dollar et al. (2016) then should be viewed as a combination of effects coming from improvements in different aspects of the economy. It is however not very meaningful to just asking whether growth is good or bad for the poor. Instead we need to ask what kind of growth-promoting policies are good for the poor.

2.6 Empirical Support for the Sorting Mechanism

I test the model predictions in this section using the third Integrated Household Survey (IHS3) of Malawi introduced in Section 2.3. Previous studies have provided evidence for worker selection between rural and urban areas. Young (2013) show that despite an average rural-urban consumption gap of 1.52 in 65 developing countries, around a quarter of individuals who lived in urban areas prior to the age of 12 migrate to rural areas as adults. Using Brazilian panel data, Alvarez (2017) find that workers switching from agriculture to non-agriculture don't experience a wage increase if they don't change their occupations, confirming the presence of sector-specific skills.

The Malawi data shows that 11% of workers originated from urban areas migrate to rural areas. The number is smaller than what Young (2013) has found but the rural-urban income gap of 3.01 in Malawi is also larger than that in the countries studied by him. Although providing some support, this is not a test of the selection mechanism. Since individual productivity is unobserved, such a test is challenging. Good panel data such as that used by Alvarez (2017) is required. The Malawi data doesn't qualify for this task. Given the multi-region setting of my model, the data does provide a way to get around this problem. Detailed data on agricultural production allows me to directly measure agricultural worker's ability by estimating agricultural production functions controlling for other inputs (Jacoby and Minten, 2009; Shenoy, 2017). With the ability measures the model prediction of within rural region ability sorting can be readily examined. This is the main objective of this section. I also provide evidence for the spatial profile of land rental rate predict by the model in next subsection. It however cannot be viewed as an piece of evidence of the ability sorting mechanism as it is also implied in Gollin and Rogerson (2014) where the sorting mechanism is missing.

2.6.1 Ability Sorting between Rural Areas

Ability sorting within rural regions can be tested if we have a measure of agricultural ability. I construct such a measure from estimating a plot-level production function. I start with decomposing total output on a plot into the contribution of inputs and a residual as follows,[49]

$$y_{ij} = \beta_0 + \beta_s s_{ij} + \beta_k k_{ij} + \beta_l l_{ij} + \beta_m m_{ij} + a_{ij},$$

[49] Output and inputs except for capital are constructed by aggregating over different items using a common set of market prices, following Restuccia and Stantaeulàlia-Llopis (2017). So the production function is measured in physical units, devoid of the effect of local prices. Capital is measured using self-evaluation of different agricultural instruments and structures owned by the household. The use of self-reported values helps capture the quality differences in capital. See Section 2.9.2 for a detailed description of data construction. I also only study the rainy season as agricultural production in Malawi mainly happens during the rainy season.

where i and j indicate farmer and plot respectively, the four inputs are land (s), capital (k), labor (l), intermediate input (m) all in logs, output(y) is also in logs, and a is the residual. The residual combines all factors affecting the output on a plot besides the inputs, including the quality of land, the weather shocks, measurement errors, and the managerial ability of the farmer that is a very important determinant in agricultural production (Welch, 1970). The purpose is to separate the managerial ability from other factors included in the residual and take it as my measure of agricultural productivity.

The residual productivity of a plot can be decomposed into a farmer-plot-specific component and a farmer-specific component,

$$a_{ij} = \Phi_i + \Psi_{ij}.$$

The farmer-specific component Φ_i affects all plots managed by the same farmer. It includes the farmer's productivity, measurement errors at the farmer level, and local weather conditions.[50] The farmer-plot-specific component Ψ_{ij} includes the quality of the land, measurement errors at the farmer level, and other random shocks. The data provides detailed information on land quality through a set of plot characteristics: the type of soil and soil quality, the extent of erosion, and the type of irrigation system.[51] Weather conditions are also well documented, including information on the amount of rain during the last completed rainy season and the timing of rain.[52] Let X_{ij} be the set of land quality and weather controls, the production function can then be estimated as,

$$y_{ij} = \beta_0 + \beta_s s_{ij} + \beta_k k_{ij} + \beta_l l_{ij} + \beta_m m_{ij} + X_{ij}\gamma + u_i + \epsilon_{ij}, \tag{2.24}$$

[50]Local weather might affect different plots differently. The effect of the local weather conditions included in the farmer-specific component then is the average effect over all plots, and the deviation from the average enters the farmer-plot-specific component.

[51]All these measures are categorical such that I can control for the differences in land quality in a flexible way.

[52]The survey asks whether the amount of rain is too much, the right amount, or too little. It also asks whether the rain began (ended) too early, at the right time, or too late.

where u_i is a farmer-specific unobserved effect including the farmer's productivity and measurement errors at the farmer level, and ϵ_{ij} includes all other random errors. Data quality of IHS3 is very high. One example for that is the measurement of plot size. In surveys on agricultural production, plot size is often self-reported which can deviate from true plot size (Carletto et al., 2013). This problem is overcome by using GPS measured plot size provided for almost 98% of all land plots.[53] For this reason, I assume the effect of measurement errors are small and take u_i to be my measure of agricultural productivity.

Given the above discussion, land quality is well measured while weather conditions at the plot level are not. ϵ_{ij} then mainly include unmeasured weather effects. If input decisions are mainly made before the realization of weather shocks, we have the exogeneity condition $E(\epsilon_{ij}|s_{ij}, k_{ij}, l_{ij}, m_{ij}, X_{ij}) = 0$. The inputs however are affected by u_i as farmers make input decisions knowing their own managerial talent. The fixed effects estimation then should lead to consistent estimates of the parameters. However, a few variables are invariant at the farmer level, including capital stock and weather conditions. Under fixed effects estimation, the coefficients for these variables are not identified and u_i will also contain the effects of capital stock and weather conditions. In view of this, I use the correlated random effects estimation proposed by Mundlak (1978). The Mundlak approach projects the unobserved effects on the observed controls,

$$u_i = \bar{Z}_i\theta + \eta_i,$$

where $\bar{Z}_i$ is the average of all controls at the farmer level, and η_i is the true random effect. Plugging this into equation 2.24, the estimation equation now reads

$$y_{ij} = \beta_0 + \beta_s s_{ij} + \beta_k k_{ij} + \beta_l l_{ij} + \beta_m m_{ij} + X_{ij}\gamma + \bar{Z}_i\theta + \eta_i + \epsilon_{ij}. \tag{2.25}$$

The model is then estimated using random effects estimation. The parameters β and γ are the same as those from the fixed effects estimation. θ equal to the difference between the fixed effects and random effects estimators.

[53]For the rest, I use self-reported plot size.

From the regression, my measure of agricultural ability is constructed as

$$\hat{a}_i = \theta_s \overline{s}_i + \theta_l \bar{l}_i + \theta_m \overline{m}_i + \eta_i. \tag{2.26}$$

The agricultural ability has two components. The first component is the projection of u_i on the inputs, which reflects how the input responds to agricultural productivity at the farmer level. Since the true output elasticities are the fixed effects estimates, these should be conceived as part of farmer's agricultural productivity correlated with the average input level. I don't include capital because the output elasticity of capital is not identified under fixed effects estimation, as capital is invariant at farmer level. To the extent that capital declines in the distance to city and higher farmer managerial talent leads to higher capital usage, I will underestimate the decline of farmer productivity in distance to the city. The second component is the random effect η_i, which is the part of productivity orthogonal to inputs. This could be due to distorted factor markets, which is prevalent in Malawi given the allocation of land is mostly made by local chiefs and intermediate inputs are heavily subsidized Restuccia and Stantaeulàlia-Llopis (2017).

Table 2.5: Production Function Estimates

	Estimates	Standard error
Land	0.23	0.025
Intermediate	0.17	0.010
Labor	0.43	0.029
Mean land	0.10	0.040
Mean intermediate	0.070	0.016
Mean labor	-0.35	0.033
Mean capital	0.14	0.014
No. of Obs	8791	
No. of Groups	3651	
Overall R-squard	0.33	

Remark: Standard errors are clustered at EA level.

I include in the regression only farmers managing at least 2 plots. I end up with 3651 farmers managing 8791 plots. Most farmers only manage 2 plots. Average number of plots per farmer is 2.4. Table 2.5 presents the estimation results. All estimates are highly significant. The output elasticities show that agricultural production in Malawi is relatively land- and labor-intensive. The estimates for the average of land size and intermediate input are positive, showing better farmers use more land and intermediate inputs. The estimates for average labor however is negative, indicating better farmers also save on labor. If we take the estimates on capital as the true output elasticity, we can not reject the hypothesis that the production function exhibits constant returns to scale.[54] If we believe the estimate for capital is larger than the true output elasticity, the production function should have minor decreasing returns to scale. This is in contrast with Shenoy (2017) who find significant decreasing returns in Thailand data. However, his estimation is at the farm level while my estimation is at the plot level. The constraint due to farmer's span of control should not matter that much given the limited size of most plots in the data.

To check the validity of the estimated agricultural productivity, I examine whether two proxies for human capital, age and years of schooling, are good predictors of the estimated agricultural productivity. Regression shown that both are highly significant predictors. The estimates for years of schooling is 0.01 and that with age is -0.002.[55] This shows education increases farming ability, while increase in age is associated with lower farming ability. This makes sense if physical strength matters for farming or young people are better at adapting to new technology. These results are also consistent with Shenoy (2017).

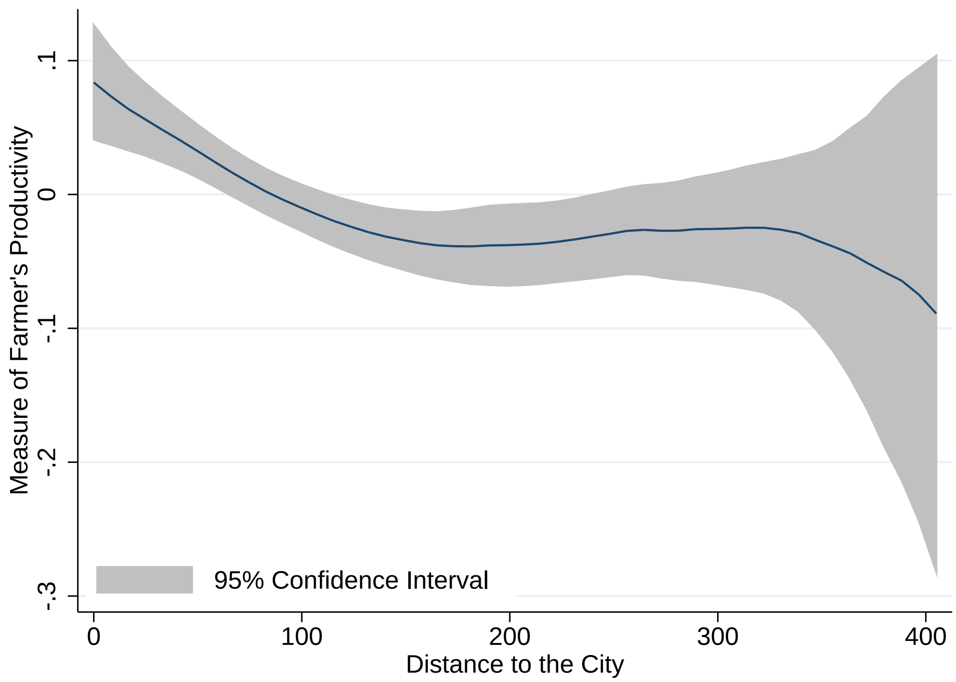
Figure 2.11 plots my measure of farmer's ability against the distance to nearest urban center. Farmer's ability quickly decreases initially then the effect of distance levels off in the more remote areas. It shows a declining trend in the farthest areas though the confidence interval is very large. The spatial profile is consistent with the model prediction and the spatial profile of income. The spatial gap of ability is much lower than that of income. The large gap in income reflects the fact that more able farmers also employ more inputs. Given

[54]A Wald test of constant returns to scale has a p-value of 0.29.

[55]The standard errors are 0.002 and 0.0004 respectively.

that I find relatively large returns to scale, we should expect the induced effect on inputs to be large.[56]

Figure 2.11: Distance to City and Farmer Productivity



Note: Farmer's productivity comes from the production function estimation described in the text. It is further smoothed using the kernel-weighted local linear regressions with Epanechnikov kernel and bandwidth equal to 50.

The spatial profile of farmer's ability might also reflect genuine benefits of being near the cities. One of the benefits might be lower intermediate prices in the near region. The production function estimation however already controls for input usage such that the effect of input prices should not enter the measure of farmer productivity. I also include in the production function estimation an intermediate input price index as control. To construct the index, I regress household level input prices (in logs) on dummies for the type of the input and dummies for EAs. The estimates for the EA dummies are then taken to be the price index. Another benefit of being near the cities is that it allows farmers to learn better farming practices that is not reflected in the use of intermediate inputs. To control for that,

[56]An estimation of the effects however is not possible because these farmers also face different prices.

I also add the access to agricultural extension services in an EA to the regression.[57] The findings are similar without or without these controls.

Since I use a common set of prices to value physical output of different crops, this could distort my finding as I also emphasize the spatial price differences. If the farmers in the remote areas grow more crops that are undervalued with the common prices, it will also generate a spatial profile as found above.[58] To exclude the effect of prices, I estimate the production function using plots growing only "local maize" or "hybrid maize", two of the most common crops in Malawi. The spatial profile of measured farmer's ability is particularly strong for hybrid maize, which is consistent with Schultz (1964) who argues that human capital is particularly important when new technology is adopted. The labor measure I use weighs the input of men, women, and children using their wages. In the baseline estimation I include farmers who actually reside in urban centers. The spatial pattern is robust to the use of a non-weighted labor measure or use observations only from the rural area. I also estimate the model using only plots with GPS-measured land size. The results still hold.

I interpret my findings as providing support to the ability sorting mechanism. Alternatively, farmer's productivity can be higher in regions closer to the urban center along the lines of Lagakos and Waugh (2013) if they are more selected as those regions might also have a higher share of non-agricultural production (Fafchamps and Shilpi, 2003). To rule out this possibility, I control for the share of income coming from agriculture in each EA in my estimation. My findings again are robust to the additional control, lending further support to the ability sorting mechanism.

[57] The survey asks about the distance to the office/residence of the nearest Assistant Agricultural Extension Development Officer.

[58] The problem is similar to that in constructing real GDP: using prices of the rich countries tend to underestimate the international income gap while using prices of the poor countries tend to overestimate the gap because consumers spend more on cheaper goods. Since the data applies to the suppliers, the opposite is true because they produce more of the expensive goods given the trading opportunity. Since medium prices are used as the common prices, they tend to be closer to prices in regions near the cities due to the goods are more likely to be available in the cities. This leads to lower prices for crops grown in remote regions.

2.6.2 Testing the Model Prediction for Land Rental Rate

The model predicts that land rental rate should be lower in remote prices. I examine whether the pattern exists in data in this subsection. For each plot, IHS3 asks the respondents to estimate the selling price and the one-year rent of the plot if they were to sell or rent out the land. Given that less than 10% of all plots are purchased or rented, it is not possible to use actual rents.[59] I use the self-estimated land price or rent divided by land size to be my measure of land value. I net out the effect of land quality by regressing the log of the value measures on the set of land quality measures discussed above. The regression residuals are then plotted against the distance to the city in Figure 2.12. Both price and rent show a clear declining trend in the distance to the city except a small bump for prices in the middle range. This is consistent with the model. Jacoby (2000) also provides similar evidence for Nepal. This pattern however is not unique to my model. It is also predicted by Gollin and Rogerson (2014).

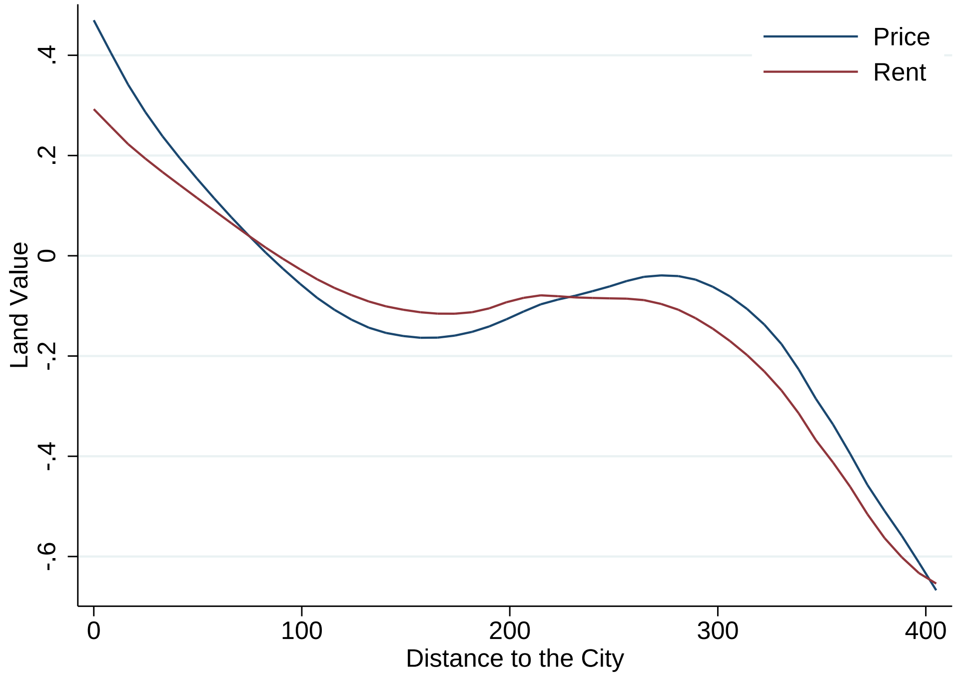
2.7 Other Supporting Evidence: Women in Agriculture

I provide the last piece of evidence regarding the employment of women in agriculture. Lagakos and Waugh (2013) find ample evidence to show that women tend to have lower agricultural productivity due to their lower physical strength and agricultural production is strength intensive. The comparative disadvantage of women in agricultural production makes them less likely to work in agriculture when the agricultural sector shrinks along economic development. The share of women in agriculture will increase as the employment share of agriculture increases. They show that this is borne out in data, supporting that workers select into different sectors according to their comparative advantage.

Another implication of women's comparative advantage in non-agriculture is that women should always be more likely to choose non-agriculture than men. To see whether that is also borne out in data, I use the International Labor Organization (ILO) database of labor

[59] Actual purchase price is not provided.

Figure 2.12: Distance to City and Land Value



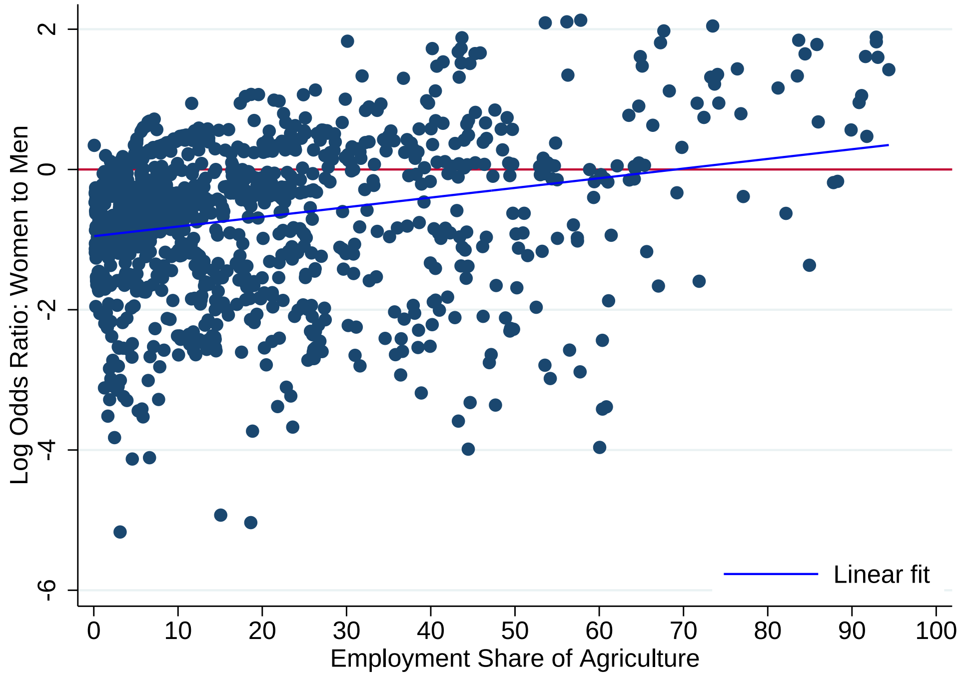
Note: The one-year rental and selling price are self-estimated by survey respondents. I divide them by land size after netting out the differences in land quality. The unit measures are then smoothed using the kernel-weighted local linear regressions with Epanechnikov kernel and bandwidth equal to 50.

statistics, which provides information on employment by sex and industry for 205 countries or regions running from 1948 or 2008. In any given year the number of countries with data available is very small, especially during the early years.[60] The data also come from different sources, including household survey, labor force survey, official estimates, and population census. I pool all available data together to maximize the use of available information, but the findings are robust to the use of data for a single year or data coming from a particular source.

Figure 2.13 plots the log odds ratio of choosing agriculture for women versus men against the employment share of agriculture. The odds ratio is defined as ratio of the probability of choosing agriculture relative to non-agriculture for women to that for men. It depicts how women have different tendencies to choose agriculture over non-agriculture than men. When

[60] In 1945, only one country (New Zealand) has data.

Figure 2.13: Log Odds Ratio of Choosing Agriculture: Women VS. Men



Note: Data come from International Labor Organization. The scatter plot pools available data from different years.

plotted in logs, a dot above zero indicates a violation of the prediction that women are less likely to choose agriculture. It can be seen that many countries have odds ratio above 1 and the chance of observing that is increasing in the employment share of agriculture. I run a logit regression of the odds ratio above 1 on the employment share of agriculture. The estimated coefficient is 0.045 and the p-value is less than 0.001. Predicted probability of the odds ratio above 1 is 0.21 when agriculture accounts for 10% of total employment, it jumps to 0.86 when the share increases to 80%.

It is hard to explain these findings if only comparative advantage is at work. My model of spatial ability sorting provides an explanation for that. Instead of relying on comparative advantage, the mechanism I emphasize relies on absolute advantage. When it is at work, women are more likely to choose agriculture than men if they are less productive than men in both sectors. What we see in data combines both the effects of comparative advantage and absolute advantage. The impact of absolute advantage is stronger in countries with large agricultural employment as 1) food consumption is more important in poor countries and 2) poor countries tend to have larger spatial differences in prices. This explains why the probability of observing an odds ratio larger than 1 is increasing in the employment share of agriculture.

For the comparative advantage mechanism to work, female workers need to be less productive than men in agriculture. Lagakos and Waugh (2013) argue that this is true as women has less physical strength than men and agricultural work is strength intensive. For the absolute advantage mechanism to work, it suffices that women are also less effective workers than men in non-agricultural work. This could be true if physical strength is also an input in non-agricultural work, albeit less important an input than in agriculture. Consider the framework of Pitt et al. (2012) where workers have two traits, skills (H) and brawn (B), the sectoral productivity which is taken as primitives in my model is a function of the two traits,

$$z_i = v_i H^{\alpha_i} B^{1-\alpha_i}, i = a, n \tag{2.27}$$

where $\alpha_a < \alpha_n$ indicates agricultural production is more brawn intensive, and v_i changes the scale of the productivity measure. If women have similar skills but less brawn than men, it would imply they have comparative advantage in non-agriculture. It also implies that women are less productive than men in non-agriculture, although to a less extent than in agriculture. This conclusion could hold even if women have more skills than men, as long as the advantage is not large enough. Women are also more likely to have absolute disadvantage in developing countries if the technology used in these countries are more brawn intensive. This happens if developing countries choose brawn intensive technologies in response to the lack of skills of their workers, as forcefully argued by Caselli and Coleman (2006). Since the non-agriculture sector is a combination of manufacturing and service, if women have comparative advantage in services as argued by Ngai and Petrongolo (2017), they will also tend to have a disadvantage in the non-agricultural sector in developing countries where the services sector is relatively small.

Absolute advantage of men is a sufficient but not necessary condition for the ability sorting mechanism to work. If women are paid less in both agriculture and non-agriculture due to discrimination, the ability sorting mechanism is still at work. Given data on wages only, we can not separate the gender wage gap due to ability differences or discrimination. Both will induce ability sorting nevertheless. Using data from International Labor Organization, Oostendorp (2009) find that the average within occupation gender wage gap is 0.11 log points across 63 countries. Surprisingly, Oostendorp (2009) also find that the gender wage gap tend to be lower in developing countries. Given that the data available are more likely for non-agricultural occupations, this could be readily explained by the ability sorting mechanism if women in non-agriculture is more selected in developing countries.

2.8 Summary

Why do the large rural-urban income gaps persist in developing countries? Young (2013) find that around a quarter of urban workers migrate to rural areas despite the large rural-urban income gaps in a sample of 65 countries.[61] The voluntary urban-rural migration suggests that the rural-urban income gaps might not be explained by labor market barriers preventing rural workers from moving to urban areas. An alternative explanation proposed in Lagakos and Waugh (2013) and Young (2013) is that workers self-select into different sectors according to their comparative advantage. More productive workers prefer urban areas more because the urban sector uses skilled labor more intensively. This chapter proposes a new explanation complementing that of Lagakos and Waugh (2013) and Young (2013). Urban areas are attractive to better workers due to their consumption value: high income workers spend a large fraction of income on non-food items which are cheaper in urban areas. The rural-urban income gaps do not reflect distortions in the allocation of workers, but come from frictions in the good markets. Quantitatively, the new mechanism plays a larger role in explaining the income gap than worker selection based on comparative advantage.

To distinguish my explanation from that of Lagakos and Waugh (2013) and Young (2013), I look beyond the rural-urban dichotomy and examines the spatial income inequality within

[61]Going from urban to rural areas, there is no drop in the variance of consumption, so the urban-rural migration can not be explained by risk factors.

rural areas. Using Malawi data, I find that income declines as we go further away from the urban center. While the rural-urban distinction is in the center stage of the study of economic development, this finding suggests that examining spatial income inequality beyond that simple dichotomy could be rewarding. A successful theory of the "dual economy" should explain both the rural-urban income gaps and the income differences between near and remote rural regions. While we can use labor market barriers to explain these spatial income gaps, they are not easily explained by worker selection. This paper provides an explanation of both the rural-urban income gap and the income profile within rural areas without relying on labor market barriers.

My analysis provides further support for investment in transportation infrastructure in developing countries (World Bank, 1994). I extend the analysis of Gollin and Rogerson (2014) to a heterogeneous worker framework which allows me to discuss the distributional effects of infrastructure development. Previous studies (Jacoby, 2000; Renkow et al., 2004; Jacoby and Minten, 2009) find that improving market access of the remote rural areas generally benefit the poor more. My general equilibrium analysis however shows that these investments might benefit the poor less than average workers when endogenous price changes are taken into account. The governments should have this in mind if poverty reduction is also a policy target.

2.9 Appendix

This section provides additional results not reported in the main text. The first subsection presents the results for the Engel curve estimation. The second subsection discusses the construction of data. The last subsection provides a proof for ability sorting under general non-homothetic preferences.

2.9.1 Engel Curve Estimates

Table 2.6 presents the results from the Engel curve estimation. For each food item, the first column gives the average share of that item in total expenditure, the second item the income coefficient and the third column the price coefficient. Estimates for other control variables are omitted.

Table 2.6: Engel Curve Estimates

	Average Share	Income(β)	Price(γ)
Food	0.483	-0.115 (0.0028)	
Maize ufa mgaiwa (Normal flour)	0.070	-0.043 (0.0015)	0.158 (0.0009)
Maize ufa refined (Fine flour)	0.052	-0.026 (0.0014)	0.086 (0.0007)
Dried fish	0.024	-0.009 (0.0005)	0.024 (0.0001)
Sugar	0.020	-0.003 (0.0004)	0.009 (0.0001)
Tomato	0.017	-0.009 (0.0004)	0.031 (0.0001)
Rice	0.013	-0.002 (0.0005)	0.004 (0.0001)
Nkhwani	0.014	-0.008 (0.0004)	0.030 (0.0002)
Bean, brown	0.013	-0.006 (0.0004)	0.011 (0.0001)
Goat	0.012	-0.011 (0.0006)	0.018 (0.00004)
Salt	0.011	-0.008 (0.0002)	0.023 (0.0001)

Remark: Standard errors in the parentheses. Reported standard errors for β and γ come from the second and third stage of my estimation approach, without considering the effect of generated variables used in that stage.

2.9.2 Data Construction

The construction of data follows De Magalhães and Stantaeulàlia-Llopis (2015) who use the Integrated Survey on Agriculture (ISA) to document the cross-sectional facts on income and consumption for several Sub-Sahara African countries including Malawi,[62] and Restuccia and Stantaeulàlia-Llopis (2017) who use the same Malawi data to study the misallocation of land and capital in agriculture.

2.9.2.1 Consumption

Total consumption is the sum of durable and non-durable consumption. Non-durable consumption includes food, clothing, services, utility, school, and medical expenditures. Durable consumption includes housing services and furniture. The expenditures on different items are for different time length. All of them are annualized to construct the total consumption. Food consumption includes 135 items, which can be purchased, home produced, or received as gifts. I evaluate the non-purchased food items at the medium reported price prevailed within a region (defined as enumeration areas within a certain distance range from a certain urban center). The food items are reported in different units. I convert them all into kilograms in two steps: 1. estimate the quantity of purchased items in modal unit using the medium price in modal unit for a region; 2. construct household conversion rate using the constructed quantity in modal unit and reported quantity in other units, use the medium household conversion rate as the conversion rate for a region. House services includes the self-reported renting value when the dwellings are owned by the households.

2.9.2.2 Income

Income consists of labor market income, agricultural net production, fishery net production, business income, capital income, and net transfers.

The most important category in agricultural net production is non-permanent crops. As not all output are sold, I evaluate the output not sold using the medium reported sales prices prevailed in a region in a similar fashion as in the construction of total consumption. Agricultural production is reported for two seasons: the rainy season and the dimba (dry) season. Most agricultural production happens in the rainy season, which makes the con-

[62] The Malawi data I use is part of the ISA project, see Section 2.3.

struction of prices for the dimba season hard due to data scarcity. I thus use the prices from the rainy to evaluate output in the dimba season. The costs for agricultural production include land rents, hired labor, transportation costs associated with sales, expenditures on fertilizers, seed, and pesticides/herbicides. Subsidies on intermediate inputs are excluded from the costs. Other agricultural production include tree/permanent crops, livestock sales, and livestock products. For all types of agricultural net production, costs associated with renting-in agricultural equipment and structure capital is subtracted from output.

Labor market income is summed over all household members report working. Workers can have multiple occupations. Wage payment can be made in cash or in kind. Household members report the last payment they receive for a reference period and how much time they work during a year, which can be used to construct total wage income in a year. Business income is aggregate over different types of non-agricultural business, such as processing/selling agricultural by-products, street or market trading business, etc. For each business, the data distinguishes months with zero, low, medium, or high volume of sales and how many months each situation spans in a year. Cost is reported only for the last month. It is re-scaled using information on revenue to estimate costs in other months.[63] Total net income from a business is then calculated by adding up profit (revenue-cost) over different sales situations, weighted by number of months of those situations. Fishery net production is similarly calculated by subtracting costs from total output. The output is valued at reported prices or imputed median price, if the households do not report any sales. Costs include rented equipment, fuel, oil, and maintenance, hired labor, and other costs. There are two seasons, high and low. Total net production is aggregate over the two seasons and over different types of fish.

2.9.2.3 Production Function Estimation

Output is constructed by aggregating output from different crops using the medium prices constructed in similar fashion as described above. Plot size is GPS measured with minimal measurement error. Physical capital is aggregated over different types of agricultural

[63] That is, given the revenue in last month, we can calculate a cost share of revenue. I assume the same share applies to all months.

machinery and structures measured by self-evaluated prices. The use of self-evaluation of asset values takes into account the differences in the quality of the machinery and structures. Intermediate input is also aggregated over different items using medium prices. Given that intermediate inputs in Malawi are heavily subsided, I only use prices constructed from input purchases without using coupons. Labor is aggregated over different sources and uses: family labor, hired labor, and exchange labor used for non-harvest work and harvest work. Labor can be supplied by men, women, and children under 15 years old. To adjust for the differences in human capital of different workers, I use wage ratios constructed from information on hired labor as weights in summing over different types of workers. Capital income mainly comes from saving and investment, rents of house and equipment, and asset sales. It also includes rental income from agricultural land and fishery equipment. Finally, net transfers is calculated as income transfers and gifts received from rural areas/urban areas/other countries minus income transfers and gifts given out to rural areas/urban areas/other countries. It might be associated government programs, social safety nets, or private transfers.

2.9.3 Ability Sorting Under Non-homothetic Preferences

This section proves that the ability sorting result holds under general non-homothetic preferences. Consider a non-homothetic preference of the general form $U(C_a, C_n)$. The agricultural good is the necessity with its income elasticity of demand less than 1. Workers choose between two regions j and j'. Regions differ in prices: $P_a^j > P_a^{j'}$ and $P_n^j < P_n^{j'}$. Regions might also differ in wages, denoted w^j and $w^{j'}$. Consider the location choice of two workers i and i' with $z_i > z_{i'}$. To prove ability sorting, it suffices to show that: 1) if worker i' chooses region j, worker i also chooses region j; 2) if worker i chooses region j', worker i' also chooses region j'. I will only prove the first part below, the second part follows the same logic.

Let $u_{i'}^{j'}$ be the utility worker i' derive in region j'. That worker i' prefers region j over region j' implies

$$e(P_a^j, P_n^j, u_{i'}^{j'}) < w^j z_{i'}, \tag{2.28}$$

where $e(\cdot)$ is the expenditure function. For small changes in prices, the change in expenditure is given by

$$de(P_a, P_n, u) = C_a dP_a + C_n dP_n$$

which can be further written as

$$d\ln e(P_a, P_n, u) = \frac{P_a C_a}{e} d\ln P_a + \frac{P_n C_n}{e} d\ln P_n = s_a(P_a, P_n, u) d\ln P_a + s_n(P_a, P_n, u) d\ln P_n,$$

where s_a and s_n are the expenditure share of the two goods.

For worker i', the change in expenditure going from region j' to j, holding utility constant, is given by

$$\ln e(P_a^j, P_n^j, u_{i'}^{j'}) - \ln e(P_a^{j'}, P_n^{j'}, u_{i'}^{j'}) = \int_{\ln P_a^{j'}}^{\ln P_a^j} \int_{\ln P_n^{j'}}^{\ln P_n^j} d\ln e(P_a, P_n, u_{i'}^{j'})$$

Similarly, the change in expenditure for worker i is given by

$$\ln e(P_a^j, P_n^j, u_i^{j'}) - \ln e(P_a^{j'}, P_n^{j'}, u_i^{j'}) = \int_{\ln P_a^{j'}}^{\ln P_a^j} \int_{\ln P_n^{j'}}^{\ln P_n^j} d\ln e(P_a, P_n, u_i^{j'})$$

Since worker i has higher income than worker i', we must have

$$s_a(P_a, P_n, u_i^{j'}) < s_a(P_a, P_n, u_{i'}^{j'}), \text{ and } s_n(P_a, P_n, u_i^{j'}) > s_n(P_a, P_n, u_{i'}^{j'})$$

Given that P_a is increasing and P_n is decreasing from j' to j, this implies that,

$$\ln e(P_a^j, P_n^j, u_i^{j'}) - \ln e(P_a^{j'}, P_n^{j'}, u_i^{j'}) < \ln e(P_a^j, P_n^j, u_{i'}^{j'}) - \ln e(P_a^{j'}, P_n^{j'}, u_{i'}^{j'})$$

Combining this with the fact $e(P_a^{j'}, P_n^{j'}, u_i^{j'}) = w^{j'} z_i$ and $e(P_a^{j'}, P_n^{j'}, u_{i'}^{j'}) = w^{j'} z_{i'}$, we have

$$\frac{e(P_a^j, P_n^j, u_i^{j'})}{w^j z_i} \frac{w^j}{w^{j'}} < \frac{e(P_a^j, P_n^j, u_i^{j'})}{w^j z_{i'}} \frac{w^j}{w^{j'}} \Rightarrow \frac{e(P_a^j, P_n^j, u_i^{j'})}{w^j z_i} < \frac{e(P_a^j, P_n^j, u_i^{j'})}{w^j z_{i'}} < 1$$

where the last inequality comes from 2.28. We thus prove worker i will also be better off in j because the change in wage is more than compensating the changes in prices.

CHAPTER 3

Capital-Labor Substitution, Agriculture, and Development

3.1 Introduction

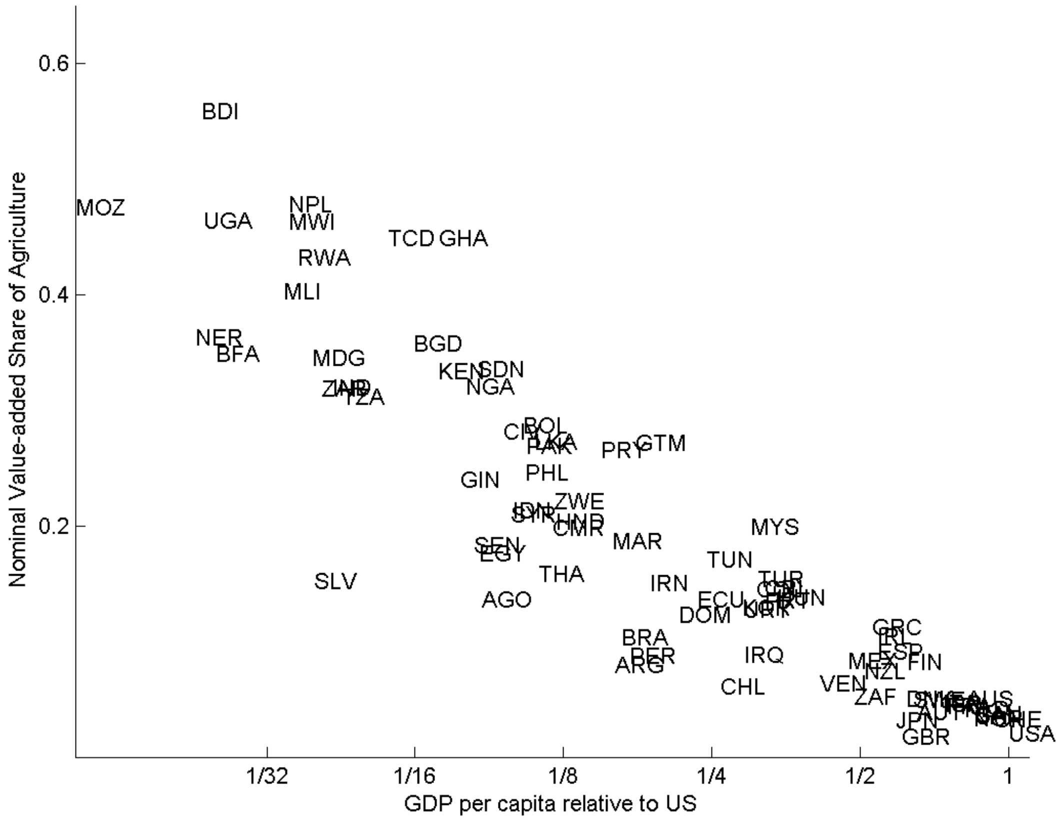
This chapter continues the discussion on agriculture. While Chapter 2 mainly focuses on the productivity gap between agriculture and non-agriculture within a country. The importance of agriculture is also reflected in the fact that agriculture accounts for a large part of the international income differences: poor countries are much more unproductive in agriculture relative to non-agriculture while allocating a large share of workers to agriculture (Caselli, 2005; Restuccia et al., 2008). Understanding the differences in sectoral allocation and agricultural productivity across countries thus can provide key insights to understanding economic development. In this chapter, I study the role of capital deepening in explaining these development facts. I allow for sectoral differences in the process of capital deepening by adopting CES production functions instead of the commonly used Cobb-Douglas production functions.

Figure 2.1 already shows that both the employment share of agriculture and the productivity gap between agriculture and non-agriculture decline with GDP per capita.[1,2] These two patterns are linked to each other through the cross-country pattern in the nominal

[1]The productivity gap is related to but different from the wage gap between sectors as studied in many micro studies. The wage gap might be a driving force behind the productivity gap. But productivity gap might still exists if there were no wage gap such as in this model. For recent studies on the wage gap, see Vollrath (2014) and Alvarez (2017).

[2]While the productivity gap can be a statistical figment due to errors in measuring output and inputs, Gollin et al. (2014) show that the pattern persists after carefully dealing with the measurement issues. See the discussion in Chapter 2.

Figure 3.1: Nominal Value-added Share in Agriculture



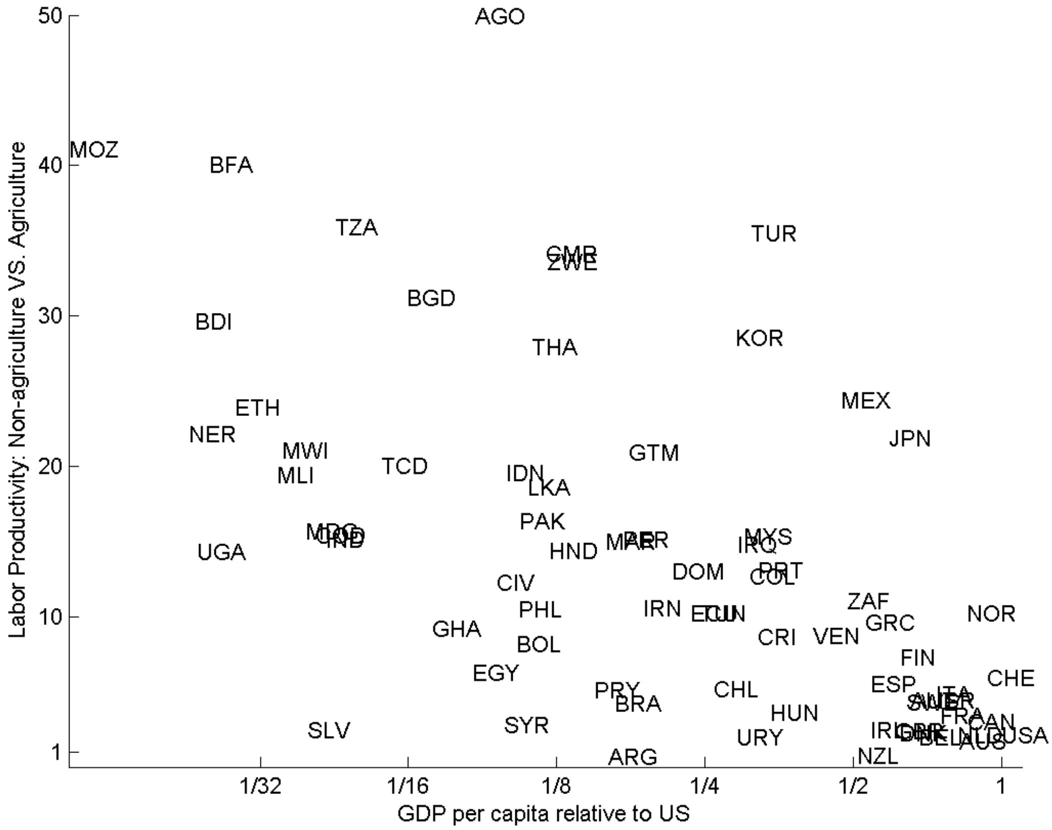
value-added share of agriculture presented in Figure 3.1.[3,4] While the value-added share also declines in the level of development, it is much smaller than the employment share in most countries and the gap is larger in poor countries. This explains why the agricultural productivity gap in current prices (APG hereafter) exists and why it negatively correlates with economic development. The same pattern can also be seen for the sectoral productivity gap measured in international prices which is presented in Figure 3.2. This directly translates into a larger dispersion in agricultural labor productivity across countries.

The model used in this chapter is a static version of an otherwise standard model of structural transformation as reviewed in Herrendorf et al. (2014). The model incorporates

[3]Remember labor productivity is defined as the ratio of value-added to employment.

[4]The data is for 1985 and comes from several sources, which is discussed in the data appendix in Section 3.8.

Figure 3.2: Agricultural Productivity Gap in International Prices



both driving forces of structural transformation discussed in the literature: non-homothetic utility (Kongsamut et al., 2001) which leads to the demand side explanation, and differential sectoral TFP (Ngai and Pissarides, 2007) which leads to the supply side explanation. The deviation is to go from Cobb-Douglas to CES production technology, that is, allowing for sectoral differences in the elasticity of substitution between capital and labor. In two recent studies, Alvarez-Cuadrado et al. (2013, 2014) show that this could be a neglected source of structural transformation: the process of capital deepening is faster in sectors with larger elasticity of substitution such that labor moves out those sectors. Their studies are motivated by ample evidence on changes in sectoral capital intensity and factor shares in many countries (Zuleta and Young, 2007; Herrendorf et al., 2015). Similar evidence has been found in agriculture. For example, Schultz (1964) argues that the modernization of traditional agriculture is mainly accomplished through the massive substitution of machines for labor.[5] Herrendorf et al. (2015) recently show that the U.S. sectoral technology is far from Cobb-Douglas: the elasticity is much larger in agriculture than in non-agriculture. I argue that this difference might have more bearing on the development facts regarding agriculture besides structural transformation and evaluate the mechanism quantitatively.

The role of capital deepening with sectoral difference in elasticity of substitution can be seen clearly from the following optimization condition,

$$\frac{\theta}{1-\theta}\left(\frac{K}{N}\right)^{\frac{1}{\sigma}} = \frac{w}{r}, \tag{3.1}$$

where capital intensity ($\frac{K}{N}$) is related to the relative factor price ($\frac{w}{r}$). Capital deepening in an economy increases the wage-rental rate ratio. The factor that the agricultural sector has a larger elasticity implies that capital intensity increases faster in agriculture, which can lead to larger changes in agricultural labor productivity in real prices and shift of employment out of agriculture.[6] In terms of APG, notice that the relative productivity ratio between the two sectors ($\frac{y_n}{py_a}$)[7] equals to the ratio of sectoral labor shares of income,[8]

$$\frac{y_n}{py_a} = \frac{Labor\ Share_a}{Labor\ Share_n}. \tag{3.2}$$

Changes in factor intensity will lead to changes in this ratio, which is decreasing in wage-rental ratio if capital-labor substitution is easier in agriculture.[9]

Quantitative study of the model shows that compared to Cobb-Douglas, CES technology provides little help in explaining sectoral allocation and real labor productivity. Large TFP differences across sectors and countries are needed to explain why agriculture performs so

---

[5] An important example of this process is the use of tractors (Manuelli and Seshadri, 2014).

[6] The effect on productivity holds true if capital intensity in agriculture is not substantially lower than that in non-agriculture, see Proposition 4.

[7] I add the relative price of the agricultural good p to indicate that it is measured in current prices.

[8] This identity only requires a competitive labor market equating wages between the two sectors and holds irrespective of other distortions in the economy. See Gollin et al. (2014) for a discussion.

[9] Notice the labor share can be expressed as a function of wage-rental ratio: $Labor\ Share = \frac{1}{(\frac{1-\theta}{\theta})^{\sigma}(\frac{w}{r})^{\sigma-1}+1}$.

badly in poor countries. This is similar to the findings of Herrendorf et al. (2015) who use U.S. time series data. The reason for the finding is that agriculture also has a large capital share other than a large elasticity of substitution. When the model is specified with Cobb-Douglas technology, sectoral differences in capital shares also help explaining the sectoral allocation, as emphasized in Acemoglu and Guerrieri (2008). Capital-labor substitution however helps explaining the agricultural productivity gap. In fact, it over-predicts the cross-country difference in APG if we use the measurement error adjusted numbers from Gollin et al. (2014). I find that APG can differ by a factor of 3 between the poorest and richest countries, while in the data the gap in APG is in the range of 2 after adjusting for measurement errors.

The model predicts incredibly high labor shares in agriculture for poor countries, suggesting the explaining power of the model for APG might be overstated. I discuss possible remedies to the model specification, in particular allowing for changes in factor-augmenting productivity which can not be identified due to the paucity of data. Since reliable data on sectoral labor shares is not available, I cannot directly test the model predictions. I however emphasize that differences in sectoral labor shares have the potential to explain the agricultural productivity gap in poor countries, contrary to the conclusion drawn in Gollin et al. (2014). The large differences in sectoral labor shares can still be consistent with a relatively stable aggregate labor share as documented in Gollin (2002). Poor countries have large employment share in agriculture which tends to lower aggregate labor share.[10] It is however counteracted by larger labor share of income in agriculture in poor countries. The two forces work together to keep aggregate labor share relatively stable.

Related literature on agriculture and development has already been extensively reviewed in Section 2.2. In particular, the literature on the role agriculture in comparative development has mainly focused on finding sources of low total factor productivity (TFP) in agriculture. Capital deepening is often thought of playing a limited role in explaining these development facts (see, for example Lagakos and Waugh (2013) and Liao and Wang (2014)).

[10]Note that labor share is larger in non-agriculture than in agriculture in rich countries.

However, the literature has mainly assumed Cobb-Douglas production functions. This chapter on the other hand provides a comparison between CES technology and Cobb-Douglas technology, in the same fashion as Herrendorf et al. (2015) who explore how structural transformation is affected by the property of technology. Wingender (2015) also focuses on technology. He emphasizes sectoral differences in the elasticity of substitution between high skill and low-skilled labor and finds that APG can be explained by the differences in skill composition between sectors. This chapter finds that CES technology mainly helps explaining the pattern of APG, contrary to what is argued by Gollin et al. (2014). Though capital-labor substitution still can not explain APG in rich countries, Herrendorf and Schoellman (2015) find that the gap in the US can be fully explained by measurement errors in agricultural value-added and in differences in worker's human capital.

The model studied in this chapter also follows a long tradition in the structural transformation literature emphasizing the role of demand and supply factors in shaping the economic structure (Kongsamut et al., 2001; Ngai and Pissarides, 2007; Acemoglu and Guerrieri, 2008). In particular, my emphasis on capital-labor substitution echoes recent studies of Alvarez-Cuadrado et al. (2013, 2014), who point out that differences in elasticity of substitution between sectors can be a source behind structural transformation. I apply this idea to a two-sector model to study its effect on both structural transformation and agricultural productivity.

The next section presents the model and derives some theoretical results. Section 3.3 calibrates the model to U.S. economy and Section 3.4 presents the quantitative results. Section 3.5 presents some supporting evidence for the capital deepening mechanism. Section 3.6 further discusses some extensions of the model. Section 3.7 concludes and Section 3.8 is a data appendix.

3.2 The Model

Technology There are two sectors in the economy: an agriculture sector (a) and a non-agriculture sector (n). Both sectors have a representative firm employ capital (K) and labor(N) to produce a single good used for final consumption. The production function has a CES form. In non-agriculture, it is given as

$$Y_n = A\left[\theta_n K_n^{\frac{\sigma_n - 1}{\sigma_n}} + (1-\theta_n) N_n^{\frac{\sigma_n - 1}{\sigma_n}}\right]^{\frac{\sigma_n}{\sigma_n - 1}}, \tag{3.3}$$

where σ_n is the elasticity of substitution and $0 < \theta_n < 1$ is a share parameter equals to the share of capital in output when capital-labor ratio is 1. A is a Hicks-neutral productivity parameter which can be viewed as the economy-wise total factor productivity (TFP). Similarly, the agriculture sector has the following technology

$$Y_a = A\kappa\left[\theta_a K_a^{\frac{\sigma_a - 1}{\sigma_a}} + (1-\theta_a) N_a^{\frac{\sigma_a - 1}{\sigma_a}}\right]^{\frac{\sigma_a}{\sigma_a - 1}}, \tag{3.4}$$

where the new parameter κ gives the relative TFP of the agricultural sector. Note that I don't introduce factor-augmenting productivity. This simplification is a result of data constraint. In the quantitative analysis, I calibrate the share parameters to target factor shares and the productivity parameters to labor productivity in data. I however don't have data on capital productivity, which can be used to discipline factor-augmenting productivity. I further discuss this in Section 3.5.

Preference A representative consumer maximizes utility given income from capital and labor. The preference is assumed to have a direct addilog form (Houthakker, 1960) over the two goods

$$U(c_a, c_n) = \omega\frac{c_a^{1-\alpha_a} - 1}{1-\alpha_a} + (1-\omega)\frac{c_n^{1-\alpha_n} - 1}{1-\alpha_n}, \tag{3.5}$$

where $0 < \omega < 1$, and I require $\alpha_a > 0$ and $\alpha_n > 0$ to guarantee diminishing marginal utility. The utility function nests the usual CES utility function when $\alpha_a = \alpha_n$, and it becomes the Cobb-Douglas utility function when we further require $\alpha_a = \alpha_n = 1$.

For my purpose, the addilog preference has the following advantages over the Stone-Geary preference widely used in the literature. First, it avoids the problem that utility might not be well-defined when the subsistence consumption is not met, which is indeed the case for the poorest countries in my data sample. Related to that, it also guarantees a positive non-agricultural employment in even the poorest countries, as is observed in data. Second, the preference is not just asymptotically non-homothetic. The income elasticity of the two goods is given by

$$e_{a,E} = \frac{\alpha_n}{\alpha_n s_a + \alpha_a s_n}, \text{ and } e_{n,E} = \frac{\alpha_a}{\alpha_n s_a + \alpha_a s_n},$$

where $s_i = \frac{p_i c_i}{E}$ is the expenditure share of good $i \in \{a, n\}$. If $\alpha_a > \alpha_n$, the income elasticity of the agricultural good is smaller than 1 and approaches from above to $\frac{\alpha_n}{\alpha_a}$ as the expenditure share of the agricultural goods decreases to 0, while the income elasticity of the non-agricultural good approaches 1 from above. This is consistent with empirical evidence as the income elasticity of food is shown to be higher in poor countries and lower in rich countries, and does not approach 1 as the Stone-Geary preference would predict (Seale Jr. et al., 2003; Muhammad et al., 2011).[11] Lastly, even though the Stone-Geary preference admits an elasticity of substitution less than 1, which is key to the supply side explanation of structural transformation (Baumol, 1967; Ngai and Pissarides, 2007). The elasticity also approaches 1 in the limit. This is not the case for the addilog preference, whose elasticity of substitution is given by

$$e_{an,p} = \frac{1}{\alpha_a} + \left(1 - \frac{\alpha_n}{\alpha_a}\right) \frac{(1-\alpha_a)s_a(\alpha_a s_n + s_a)}{(\alpha_n s_a + \alpha_a s_n)[2s_a(1-\alpha_a) + \alpha_a]}.$$

The elasticity of substitution depends on the expenditure shares. One set of sufficient conditions for the elasticity to be less than 1 is $\alpha_a > 1$ and $s_a < \frac{\alpha_a}{2(\alpha_a - 1)}$, which is satisfied in the data. As the economy grows, the expenditure share of the agricultural good becomes small, which gurantees $[2s_a(1-\alpha_a) + \alpha_a] > 0$. In this case, $\alpha_a > 1$ is sufficient for the two goods to be gross complements.

[11] The structural transformation literature has pointed out the importance of income effect in explaining sectoral reallocation, even in developed countries (Boppart, 2014; Comin et al., 2015). Some other proposed preferences that allow for long-run income effects are the constant differences of elasticity of substitution preference (Święcki, 2014) which belongs to the class of indirect addilog preferences of Houthakker (1960), the price independent generalized linear preference (Boppart, 2014), and the non-constant CES preference (Comin et al., 2015).

Finally, the budget constraint of the representative agent is given by

$$pc_a + c_n = rK + wN, \tag{3.6}$$

where the non-agricultural good is taken to be the numeraire, p is the relative price of the agricultural good, and r and w are the rental rate of capital and the wage rate respectively.

Equilibrium All markets are competitive. The equilibrium is defined in the usual way: 1) firms maximize profits taking technology and prices as given; 2) the representative consumer maximizes utility given the budget constraint; and 3) markets clear.

The optimality condition for utility maximization is given by

$$\frac{\omega}{1-\omega}\frac{Y_a^{-\alpha_a}}{Y_n^{-\alpha_n}} = p, \tag{3.7}$$

where I have plugged in the market clearing conditions for the two goods. For the representative firm in agriculture, profit maximization gives

$$p\theta_a \left(A\kappa\right)^{\frac{\sigma_a-1}{\sigma_a}} \left(\frac{Y_a}{K_a}\right)^{\frac{1}{\sigma_a}} = r, \;\; p(1-\theta_a)\left(A\kappa\right)^{\frac{\sigma_a-1}{\sigma_a}} \left(\frac{Y_a}{N_a}\right)^{\frac{1}{\sigma_a}} = w, \tag{3.8}$$

Similarly, the profit maximization conditions in non-agriculture are given by

$$\theta_n A^{\frac{\sigma_n-1}{\sigma_n}} \left(\frac{Y_n}{K_n}\right)^{\frac{1}{\sigma_n}} = r, \;\; (1-\theta_n)A^{\frac{\sigma_n-1}{\sigma_n}} \left(\frac{Y_n}{N_n}\right)^{\frac{1}{\sigma_n}} = w, \tag{3.9}$$

The equilibrium is fully described by the system of equations (3.3), (3.4), (3.7), (3.8), (3.9), and the two market clearing conditions for the capital and labor,

$$K_a + K_n = K, \;\; N_a + N_n = N. \tag{3.10}$$

Discussion The model allows all the driving forces of structural transformation emphasized in the literature.[12] The new source introduced is capital-labor substitution (Alvarez-

[12] Since the model economy is closed, the effect of international trade is not considered (Uy et al., 2013).

Cuadrado et al., 2013, 2014). I next discuss how it might help explain the three development facts regarding agriculture. Before that, the following lemma is going to be useful.

Lemma 1 *The elasticity of substitution for the aggregate economy $\sigma^{agg} = -\frac{d\log K/N}{d\log r/w}$ is given by*

$$\sigma^{agg} = \frac{\eta_a(1-\eta_a)}{\eta(1-\eta)} s_a\sigma_a + \frac{\eta_n(1-\eta_n)}{\eta(1-\eta)} s_n\sigma_n + \left(1 - \sum_i \frac{\eta_i(1-\eta_i)}{\eta(1-\eta)} s_i\right)\varepsilon,$$

where $\eta_i = \frac{rK_i}{rK_i+wN_i}$ is the capital share of output in sector i, $\eta = \frac{rK}{rK+wN}$ is the aggregate capital share, and ε is the elasticity of substitution between the two goods.

The proof follows Oberfield and Raval (2014). We can write the sectoral and aggregate elasticity of substitution as follows.

$$\begin{aligned}
\sigma_i - 1 &= -\frac{\mathrm{d}\log rK_i/wN_i}{\mathrm{d}\log r/w} = -\frac{\mathrm{d}\log \eta_i/(1-\eta_i)}{\mathrm{d}\log r/w} = -\frac{1}{\eta_i(1-\eta_i)}\frac{\mathrm{d}\eta_i}{\mathrm{d}\log r/w}, \\
\sigma^{agg} - 1 &= -\frac{\mathrm{d}\log rK/wN}{\mathrm{d}\log r/w} = -\frac{\mathrm{d}\log \eta/(1-\eta)}{\mathrm{d}\log r/w} = -\frac{1}{\eta(1-\eta)}\frac{\mathrm{d}\eta}{\mathrm{d}\log r/w}.
\end{aligned}$$

Notice that $\eta = \sum_i \eta_i s_i$, we have

$$\frac{\mathrm{d}\eta}{\mathrm{d}\log r/w} = \sum_i \frac{\mathrm{d}\eta_i}{\mathrm{d}\log r/w} s_i + \sum_i \frac{\mathrm{d}s_i}{\mathrm{d}\log r/w}\eta_i$$

while changes in expenditure share is given by

$$\frac{\mathrm{d}s_i}{\mathrm{d}\log r/w} = (\varepsilon - 1)(\eta_i - \eta)$$

Combining all these equations together gives Lemma 1.

It is easy to show that the aggregate elasticity of substitution given in Lemma 1 is larger than 0,[13] such that an increase in aggregate capital intensity lowers the relative price of capital. I next proceed to study the effect of an decrease in the relative factor price (r/w)

Section 3.6.4 discusses how opening to trade might change the results of the baseline model.

[13]We only have to show $1 - \sum_i \frac{\eta_i(1-\eta_i)}{\eta(1-\eta)} s_i > 0$, which follows from the definition of η.

caused by capital deepening. I will consider the empirically relevant case described in the following assumption.[14]

Assumption 1 *$\sigma_a > 1 > \sigma_n$, and $\alpha_a > 1 > \alpha_n$.*

Under Assumption 1, a decrease in the relative price of capital lowers labor share in agriculture but raises that in non-agriculture. Since we have

$$\frac{Y_n/N_n}{pY_a/N_a} = \frac{1-\eta_a}{1-\eta_n},$$

from the first order condition for labor. This just says that non-agricultural labor productivity relative to agricultural labor productivity in nominal prices increases due to the increase in capital intensity. This explains why APG is lower in developed countries.

It can also be shown that the employment share of agriculture also decreases in aggregate capital intensity. To see this, notice we can write the price of agricultural good as

$$p = \left(\frac{Y_a}{N_a}\right)^{-\frac{1}{\sigma_a}} \left(\frac{Y_n}{N_n}\right)^{\frac{1}{\sigma_n}}.$$

Optimal consumption allocation can be rewritten as

$$\omega \left(\frac{Y_n}{N_n} N_n\right)^{\alpha_n} = (1-\omega) p \left(\frac{Y_a}{N_a} N_a\right)^{\alpha_a}.$$

From these two equations we can express the employment share of agriculture as an implicit function of labor productivity in the two sectors. Since labor productivity is determined by the sectoral capital-labor ratio, which itself is determined by the relative price of capital from firm's optimal decision. Given the relative price, we can then solve for the employment

[14]The estimates presented in the next section are consistent with the assumption.

share of agriculture from the following equation

$$\Phi \frac{\left[\theta_n \left(\frac{1-\theta_n}{\theta_n}\frac{r}{w}\right)^{1-\sigma_n} + (1-\theta_n)\right]^{\frac{\alpha_n\sigma_n-1}{\sigma_n-1}}}{\left[\theta_a \left(\frac{1-\theta_a}{\theta_a}\frac{r}{w}\right)^{1-\sigma_a} + (1-\theta_a)\right]^{\frac{\alpha_a\sigma_a-1}{\sigma_a-1}}} = \frac{N_a^{\alpha_a}}{(N-N_a)^{\alpha_n}},$$

where Φ is a constant. Since the left hand side decreases due to the decrease in relative factor price, N_a also decreases. An increase in capital intensity thus lowers the share of labor in the agricultural sector. This establishes capital deepening as a source of structural transformation when the elasticity of substitution differs across sectors.

Finally, I examine the effect of capital deepening on relative productivity measured in real prices. Agains, since labor productivity is determined by capital-labor ratios which itself is determined by the relative price, we have

$$\begin{aligned}\frac{\partial \log Y_n/N_n}{\partial \log r/w} &= \frac{\partial \log Y_n/N_n}{\partial \log K_n/N_n}\frac{\partial \log K_n/N_n}{\partial \log r/w} = -\frac{rK_n}{Y_n}\sigma_n,\\ \frac{\partial \log Y_a/N_a}{\partial \log r/w} &= \frac{\partial \log Y_a/N_a}{\partial \log K_a/N_a}\frac{\partial \log K_a/N_a}{\partial \log r/w} = -\frac{rK_a}{Y_a}\sigma_a\end{aligned}$$

Changes in relative labor productivity thus depend on the size of the elasticity and capital share of income in each sector. Agricultural labor productivity increases relative to non-agricultural labor productivity as long as $\eta_a\sigma_a > \eta_n\sigma_n$. Given that $\sigma_a > \sigma_n$, agricultural labor productivity increases relative to non-agriculture in response to an increase in aggregate capital intensity as long as capital share of income in agriculture is not substantially smaller than that in non-agriculture. This condition will hold in rich countries as agriculture is more capital intensive than non-agriculture in those countries. We are however not sure whether it is also the case in poor countries due lack of information on sectoral factor shares. The effects of capital deepening in this economy are summarized in the following proposition.

Proposition 4 *Given Assumption 1, capital deepening induces labor to reallocate to non-agriculture, lowers the agricultural productivity gap, and raises labor productivity more in agriculture if $\eta_a\sigma_a > \eta_n\sigma_n$.*

3.3 Calibration

I calibrate the model parameters to U.S. data in this section. I estimate the elasticity of substitution between capital and labor following Herrendorf et al. (2015). The two parameters that determine the curvature of the utility function are estimated from the associated demand system using a GMM approach. The rest of the parameters are then calibrated to key data moments of the U.S. economy.

3.3.1 Elasticity of Substitution in Production

The elasticity of substitution is estimated using the procedures in Herrendorf et al. (2015). They estimate a system of equations including the production function and the two first order conditions as follows.

$$
\begin{aligned}
\log\left(\frac{Y_{it}}{\bar{Y}_i}\right) &= \frac{\sigma_i}{\sigma_i - 1}\log\left[\bar{\theta}_i\left(\exp(\gamma_{ik}(t-\bar{t}))\frac{K_{it}}{\bar{K}_i}\right)^{\frac{\sigma_i-1}{\sigma_i}} + (\overline{1-\theta_i})\left(\exp(\gamma_{in}(t-\bar{t}))\frac{N_{it}}{\bar{N}_i}\right)^{\frac{\sigma_i-1}{\sigma_i}}\right] + \epsilon_{yit} \\
\log(r_{it}) &= \log\left(\frac{\bar{\theta}_i\bar{Y}_i}{\bar{K}_i}\right) + \frac{\sigma_i-1}{\sigma_i}[\gamma_{ik}(t-\bar{t})] + \frac{1}{\sigma_i}\log\left(\frac{Y_{it}/K_{it}}{\bar{Y}_i/\bar{K}_i}\right) + \epsilon_{rit} \\
\log(w_{it}) &= \log\left(\frac{\overline{1-\theta}_i\bar{Y}_i}{\bar{N}_i}\right) + \frac{\sigma_i-1}{\sigma_i}[\gamma_{in}(t-\bar{t})] + \frac{1}{\sigma_i}\log\left(\frac{Y_{it}/N_{it}}{\bar{Y}_i/\bar{N}_i}\right) + \epsilon_{wit}
\end{aligned}
$$

where i indexes sector and t indexes time, $\{\epsilon_{yit}, \epsilon_{rit}, \epsilon_{wit}\}$ is a set of random errors, the variables with a hat are the geometric average over the sample period except for $\bar{t}$, which is an arithmetic average of the time index, γ_{ik} and γ_{in} are the growth rates of factor-augmenting productivity. The functions are normalized such that the share parameters are calibrated: $\bar{\theta}_i$ denotes the geometric average of observed capital shares over the period. The system is estimated using the non-linear, feasible, generalized three-stage least squares estimation procedure implemented in Eviews. The estimation uses the one-period lags of all endogenous variables and a time trend as instruments and allows a AR(1) structure in the error term.

Herrendorf et al. (2015) estimate the system for 3 broad sectors: agriculture, manufacturing, and services. They use U.S. data over the period 1947-2010. I use their data and

Table 3.1: Production Function Estimates

	Agriculture	Non-agriculture
σ	1.60	0.88
	(0.069)	(0.039)
γ_k	0.022	-0.021
	(0.0029)	(0.010)
γ_n	0.050	0.025
	(0.0043)	(0.0054)
$\bar{\theta}$	0.61	0.32

Note: standard errors in parentheses.

combine manufacturing and services into one single non-agricultural sector.[15] The estimation results are provided in Table 3.1. Capital and labor are substitutes in agriculture with an elasticity of 1.6. The elasticity for non-agriculture is only 0.88. These estimates are consistent with that in Herrendorf et al. (2015) (Table 1). The estimates for agriculture are close to their estimates for agriculture, while the estimates for non-agriculture is close to their estimates for total economy.[16] Given the small share of agriculture during most of the period, this is hardly a surprise.

The estimates of the elasticity of substitution also find supports in other studies. Studies of U.S. agriculture looking at the changes in factor cost shares, capital-labor ratio, and relative prices over time also find an elasticity of substitution around 1.5 (Lianos, 1971; Kislev and Peterson, 1981, 1982). This fits the description of a wave of mechanization after the second world war in Schultz (1964). The ease of capital-labor substitution, however, is not confined to the U.S. Thirsk (1974) finds that it is also the case in Colombia. He also summarizes studies on U.S., Europe, India, and Brazil to conclude that "the elasticity of capital-labor substitution in agriculture exceeds unity and is probably close to one and a half" (pp. 80). More recently, Xu (1999) studies China's rapid growth in agricultural productivity and finds an elasticity of substitution of 1.4.

On the other hand, an estimate of elasticity of substitution for non-agriculture is rare in the literature. There is, however, ample evidence showing it is smaller than 1 for the

[15] Data is downloaded at `https://www.aeaweb.org/articles?id=10.1257/mac.20130041`.

[16] Their estimate for agriculture is 1.58, and that for the total economy is 0.84.

aggregate economy (Chirinko, 2008). Remember the aggregate elasticity of substitution from Lemma 1. Aggregate elasticity is a weighted sum of the sectoral elasticity and the elasticity of substitution in consumption. Given that that in data factor shares varies between 0.3 and 0.7, the weight for the elasticity in consumption is close to 0.[17] Aggregate elasticity of substitution is then close to a weighted average of the elasticity of substitution in the two sectors. The fact it is less than 1 combined with an estimate larger than 1 for the agricultural sector implies inelastic substitution in the non-agricultural sector.

3.3.2 Preference Parameters

I estimate the two preference parameters α_a and α_n using a GMM approach.[18] Given the preference parameters and data on the budget constraint, we can solve for the optimal consumption decisions numerically. The predicted expenditure shares ($\hat{s}_{it}$) are not going to exactly match the true expenditure shares (s_{it}). I treat the gap between data and model prediction as a result of measurement errors and non-optimizing consumer behavior. These errors should be uncorrelated with the production side of the economy. We thus can use supply side variables as instruments. Given the instrument, the moment condition is given as follows

$$E\left[x_{it}\epsilon_{it}\right] = 0,$$

where $\epsilon_{it} = s_{it} - \hat{s}_{it}$ is the prediction error in expenditure share for sector i in period t, x_{it} is any instrument used. The sample analog to the moment condition is

$$m(\Omega; x_{it}) = \frac{1}{T}\sum_{t=1}^{T} x_{it}(s_{it} - \hat{s}_{it}),$$

[17]The maximum of $\eta(1-\eta)$ is 0.25 when $\eta = 0.5$, while the minimum is 0.21 when $\eta = 0.3$, or 0.7. This says $\frac{\eta_i(1-\eta_i)}{\eta(1-\eta)}$ will be close to 1.

[18]I also recover the parameters by estimating the associated demand system, following the approach in Deaton (1974). The estimates are similar to that reported below. The parameters can also be estimated by using the cointegration structure implied in the optimality condition (Ogaki, 1992; Clarida, 1994). I however fail to detect a cointegration relationsip in the data.

where T is the total number of periods in data, Ω is the set of parameters to be estimated. The GMM estimator then minimizes the following object

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmin}} T \cdot \sum_k m(\Omega; x_{at}^k)^2,$$

where k indexes the instruments used. Also notice that the estimation is confined to the agricultural sector. The other sector does not provide additional information as the expenditure shares sum up to 1 both in data and in the model (Deaton, 1974).

I use Herrendorf et al. (2013)'s 3-sector consumption value-added data, which covers the period 1947-2010. Manufacturing and services are again combined into one non-agricultural sector. The instruments used include a constant term, and labor productivity in both sectors, which comes from Herrendorf et al. (2015) as described above. The Herrendorf et al. (2013) dataset only considers value-added used for consumption while neglecting investment. This could be a problem for the model to match data as the size and structure of investment might change from country to country. Section 3.6.1 gives a discussion on how to adjust for investment.

The estimation results are presented in Table 3.2. I also report the implied income elasticity of the agricultural good for two expenditure shares 0.1 and 0.5, which respectively are reasonable numbers for U.S. around 1950 and a poor country like Malawi today. The income elasticity is reported to make sure the estimates are within reasonable bounds. The literature only provides income elasticity estimates for final consumption, but not consumption value added. I thus compare these estimates to income elasticity of food. Notice that the income elasticity of food tends to be much larger than that for agricultural value added, as food includes not only agricultural value-added but also services associated with processing, packaging, and distributing food (Bunkers and Cochrane, 1957). Column 1 of Table 3.2 shows that the estimate for α_n approaches 0, indicating that a quasi-linear utility function fits the data. Income elasticity of the agricultural good takes two extreme values. When income is low, it is 1 while it jumps to 0 when income is high enough. These estimates provides support for the preference used in Gollin et al. (2004, 2007).

Table 3.2: Estimates of Preference Parameters

	1 (No constraint)	2 ($\alpha_n > 0.1$)
α_a	1.74	1.87
α_n	~ 0	0.102
RMS s_a	0.0516	0.0521
Implied income elasticity		
$s_a = 0.1$	0	0.06
$s_a = 0.5$	0	0.1

Note: the estimates for ω is not reported.

The implied income elasticity in Column 1 however is far below Bunkers and Cochrane (1957)'s estimates for farm product, which casts doubt on the extreme estimates for α_n. Similar to the Stone-Geary preference, it is still possible under the quasi-linear preference that a poor country has all employment in agriculture, which is not observed in data. In view of this, I provide another set of estimates by imposing an additional constraint on the estimation. In Column 2 I require $\alpha_n > 0.1$. Judging by the mean squared error of the estimated equation for expenditure shares, the model fit does not change much. The implied income elasticity now are 0.06 and 0.1 respectively, which are more reasonable. I hence use these estimates in the baseline numerical experiment.[19]

3.3.3 Other Parameters

Given the production and preference parameters, the other parameters are calibrated to match data moments of the U.S. economy. The two productivity parameters are normalized to be 1. This only changes the unit of output. The two share parameters in the production function are calibrated to match a capital share of 0.61 in agriculture and 0.32 in non-agriculture (see Table 3.1). The weight in preference ω is calibrated to match a current employment share of 2.85% in agriculture. Labor endowment is normalized to be 1 such that the results are on a per capita basis. K is set to deliver a capital-output ratio of 2.5.

Even in the U.S., the agricultural productivity gap is still larger than 1, which is incon-

[19]Note smaller income elasticity of food helps explain a large agricultural sector in poor countries. The additional constraint hence might lower the explaining power of the model.

Table 3.3: Baseline Calibration

Parameter	Value	Target
		Production
A,k,N	1	Normalization
K	3.95	Capital-output ratio (2.5)
θ_a	0.46	Capital share in agriculture (0.61)
σ_a	1.60	Table 3.1
θ_n	0.36	Capital share in non-agriculture (0.32)
σ_n	0.88	Table 3.1
		Preference
α_a	1.87	Table 3.2
α_n	0.102	Table 3.2
ω	0.003	Employment share in agriculture (2.85%)
		Barrier
ξ	0.59	Agricultural productivity gap (1.4)

sistent with observed factor shares. To match this gap, I add a wedge to the non-agricultural labor market following Adamopoulos and Restuccia (2014). In particular, I assume a gap of $\frac{1}{1-\xi}$ between non-agricultural and agricultural wages. Now let w denote the wage prevalent in the agricultural sector, the only change to the equilibrium is that the labor demand in non-agriculture now reads,[20]

$$(1-\theta_n)A^{\frac{\sigma_n-1}{\sigma_n}}\left(\frac{Y_n}{N_n}\right)^{\frac{1}{\sigma_n}} = \frac{w}{1-\xi}, \tag{3.11}$$

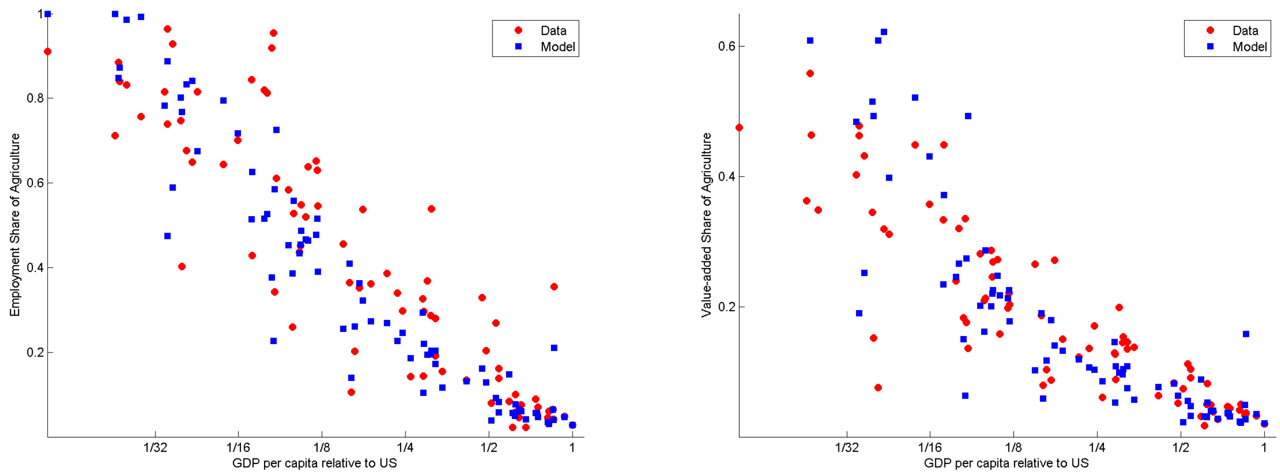
with ξ chosen to match a agricultural productivity gap of 1.4 as observed in data. The calibrated parameters are summarized in Table 3.3.

3.4 Quantitative Results

I am now ready to examine the model's quantitative performance. The data I use are compiled from Penn World Table, Food and Agriculture Organization of United Nations, United Nations National Accounts, and International Labor Organization. It covers 80

[20]Another interpretation is this procedure corrects the measurement errors in the data. Herrendorf and Schoellman (2015) show that after accounting for the measurement errors and differences in human capital agricultural wage and non-agricultural wage are not so different from each other.

Figure 3.3: Sectoral Allocation, Model VS. Data



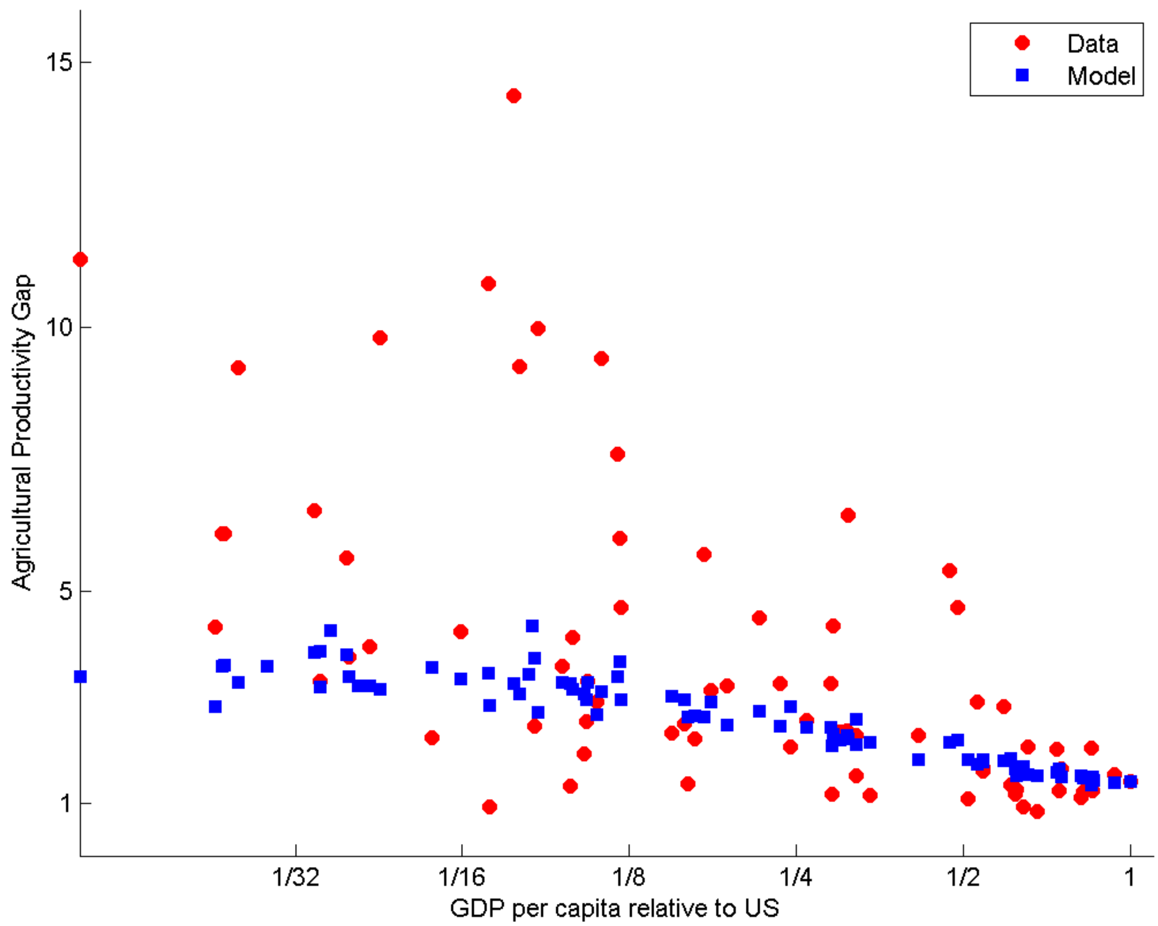
Note: left panel is for employment shares, and the right panel is for value-added shares.

countries in 1985, including information on sectoral productivity, employment and value-added shares, and aggregate capital and labor endowment.[21] For each country, I calibrate the two productivity measures to match observed labor productivity in the two sectors, and the aggregate capital intensity is calibrated to directly match that in data. A comparison between model prediction and data are presented in Figure 3.3 and 3.4. Figure 3.3 shows that the model outcome tracks well the employment and value-added share in data. Figure 3.4 shows that the model predicts a negative relationship between the agricultural productivity gap and income level, which can only be driven by difference in capital-labor substitution between sectors.

To examine the model fit quantitatively, I regress data on model prediction for each of the measures shown in Figures 3.3 and 3.4. The regression results are shown in Table 3.4. I look at the regression coefficient and R-squared respectively. The model predicts the employment share quite well, with the regression coefficient close to 1 and a R-squared of 0.87. The model prediction for value-added share is worse than that for employment. In particular, a regression coefficient less than 1 indicates that the model over-predicts the value-added share for most countries. This directly translates to an underestimation of the

[21]Section 3.8 provides a more detailed description.

Figure 3.4: Agriculture Productivity Gap, Model VS. Data



agricultural productivity gap. R-squared for APG is also very low, confirming the impression from Figure 3.4. Also presented in Table 3.4 is a comparison with Cobb-Douglas technology allowing for sectoral differences in the capital share.[22] It shows that allowing for capital-labor substitution helps explaining the sectoral allocation pattern, though not by a large margin. The model with CES technologies helps getting the employment shares right but performs worse for the value-added shares. As I emphasize above, Cobb-Douglas technology cannot explain APG at all.

The model underestimates the agricultural productivity gap and the data is also much

[22]To save space, I don't present the figures for Cobb-Douglas technology.

Table 3.4: Model Fit

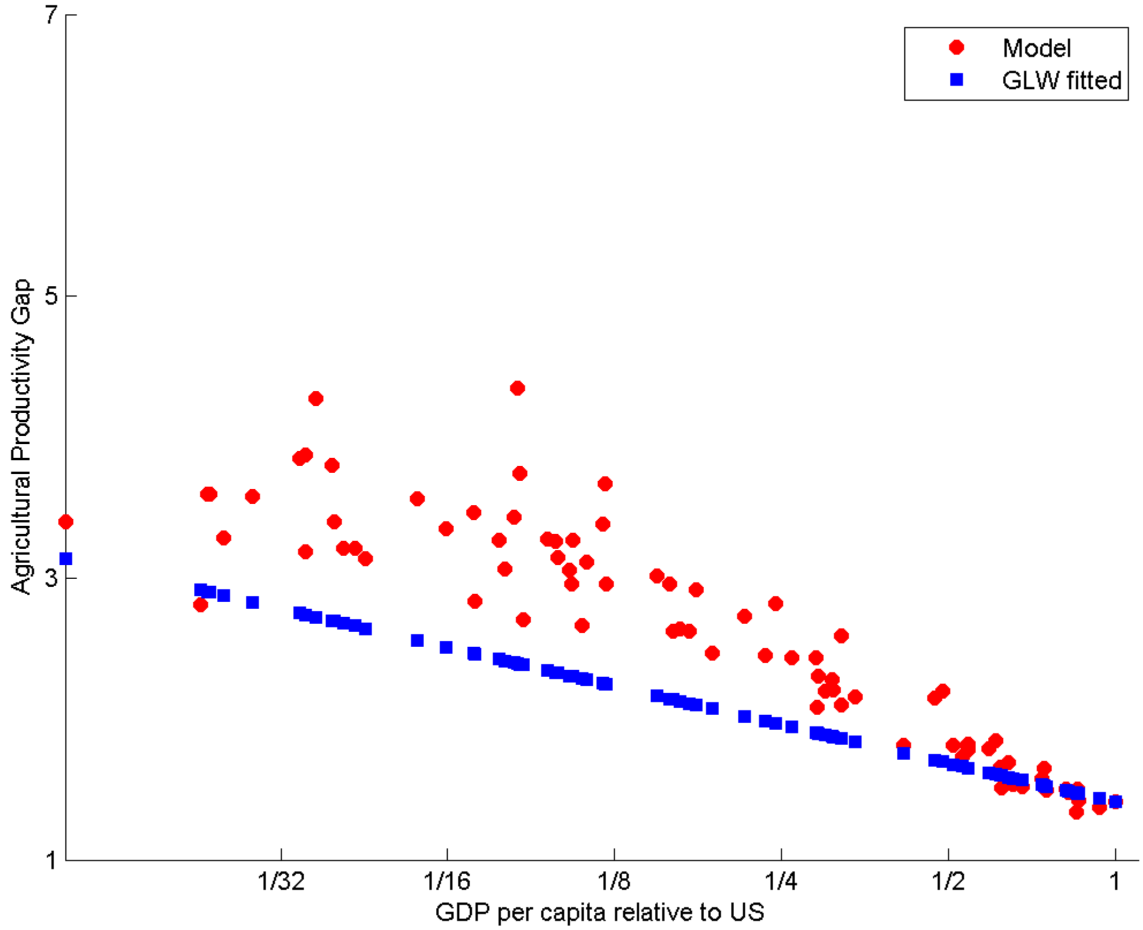
	CES		C-D	
	Coefficient	R-squared	Coefficient	R-squared
Employment	1.01	0.88	1.41	0.84
	(0.043)		(0.069)	
Value-added	0.63	0.74	0.77	0.80
	(0.043)		(0.043)	
APG	2.42	0.08	-	-
	(0.31)			

Note: This table compares the model fit with CES and CD production functions. The numbers reported come from regressing actual data on predicted value without an intercept.

more volatile than the model predicts. This is partly driven by measurement errors in APG. Gollin et al. (2014) show that after adjusting for measurement errors in both input and output, the agricultural productivity gap shrinks a lot, which is particularly the case in developing countries. Figure 3.5 contrasts the model prediction to the adjusted APG in Gollin et al. (2014). What is presented is the model prediction against a linear relationship between APG and GDP per capita using the adjusted numbers from Gollin et al. (2014), so we are not targeting the extreme values for particular countries but an average over countries. Surprisingly, the model actually over-predicts APG in the poorest countries. While Gollin et al. (2014) show that the gap in APG between the poorest and richest countries is on average around 2, the model predicts a gap close to 3.

As shown above, APG is tightly linked to sectoral labor share of income under competitive markets. Figure 3.6 plots the model prediction of labor shares for the two sectors and the aggregate economy. This over-prediction of APG in poor countries is echoed by the implausible predictions for labor shares, with agricultural labor share in poor countries between 0.8 and 0.9. Even though we don't have good data on sectoral labor shares covering countries of different income levels, these numbers are not credible. Fractional evidence in Fuglie (2010) shows that labor share in agriculture tends to be around 0.6 in developing countries and it doesn't fall as country develops. Using the labor share of 0.6, APG in poor countries will be reduced by a third, closer to the adjusted numbers reported in Gollin et al. (2014). I argue that the over-prediction of APG and agricultural labor share are more likely

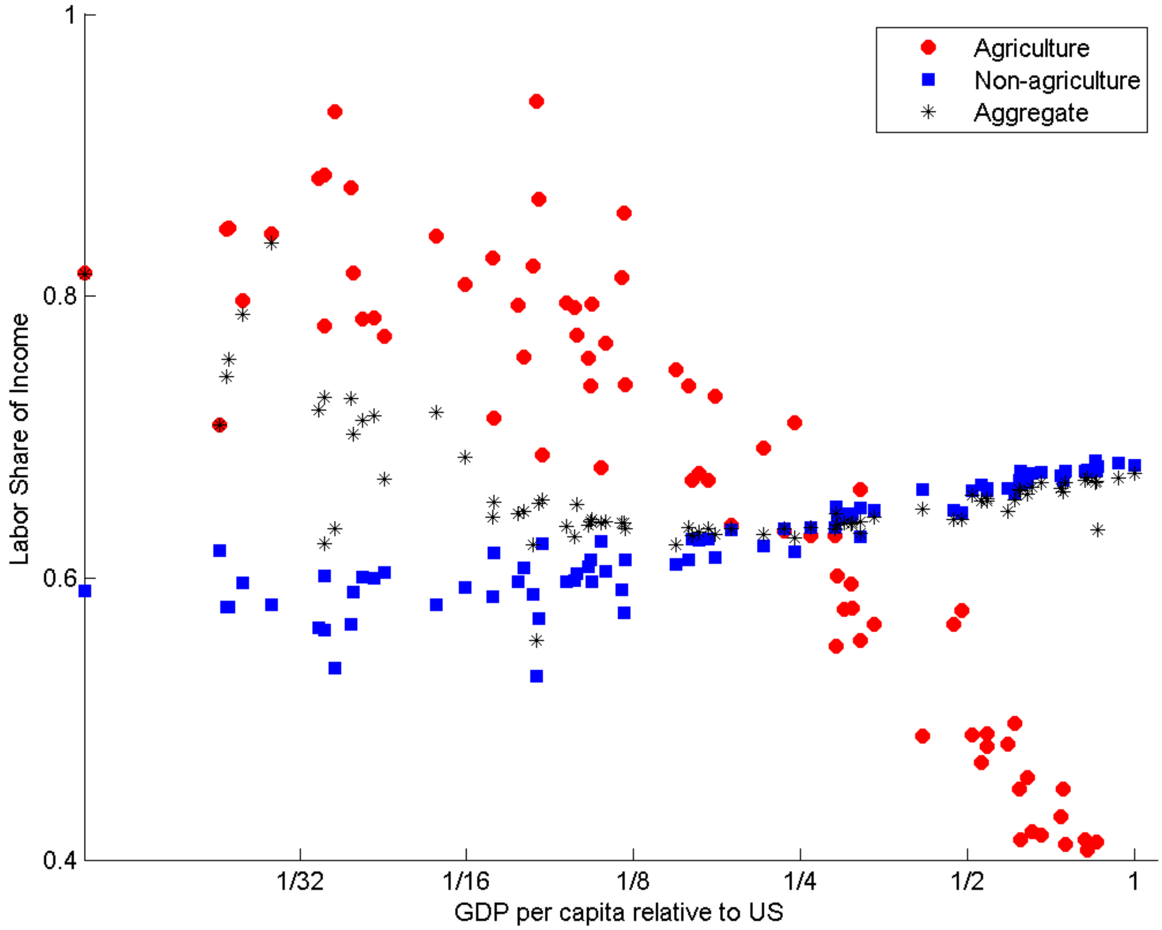
Figure 3.5: Agriculture Productivity Gap, Model VS. GLW



Note: The blue dots are fitted APG for each country using adjusted numbers from Gollin et al. (2014).

due to model mis-specification, and should not be seen as a rejection of the capital-labor substitution mechanism. In particular, land as an important input in agricultural production is not separated from physical capital in the model. Also, factor-augmenting productivity change is not modeled due to data constraints. Section 3.6 discusses some extensions that take these into consideration. For countries above a certain income level, the model predicts that the aggregate labor share does not change much across countries. This is consistent with the findings of Gollin (2002). In the model, both sectoral labor share of income and sectoral composition change along economic development. Poor countries have both a larger

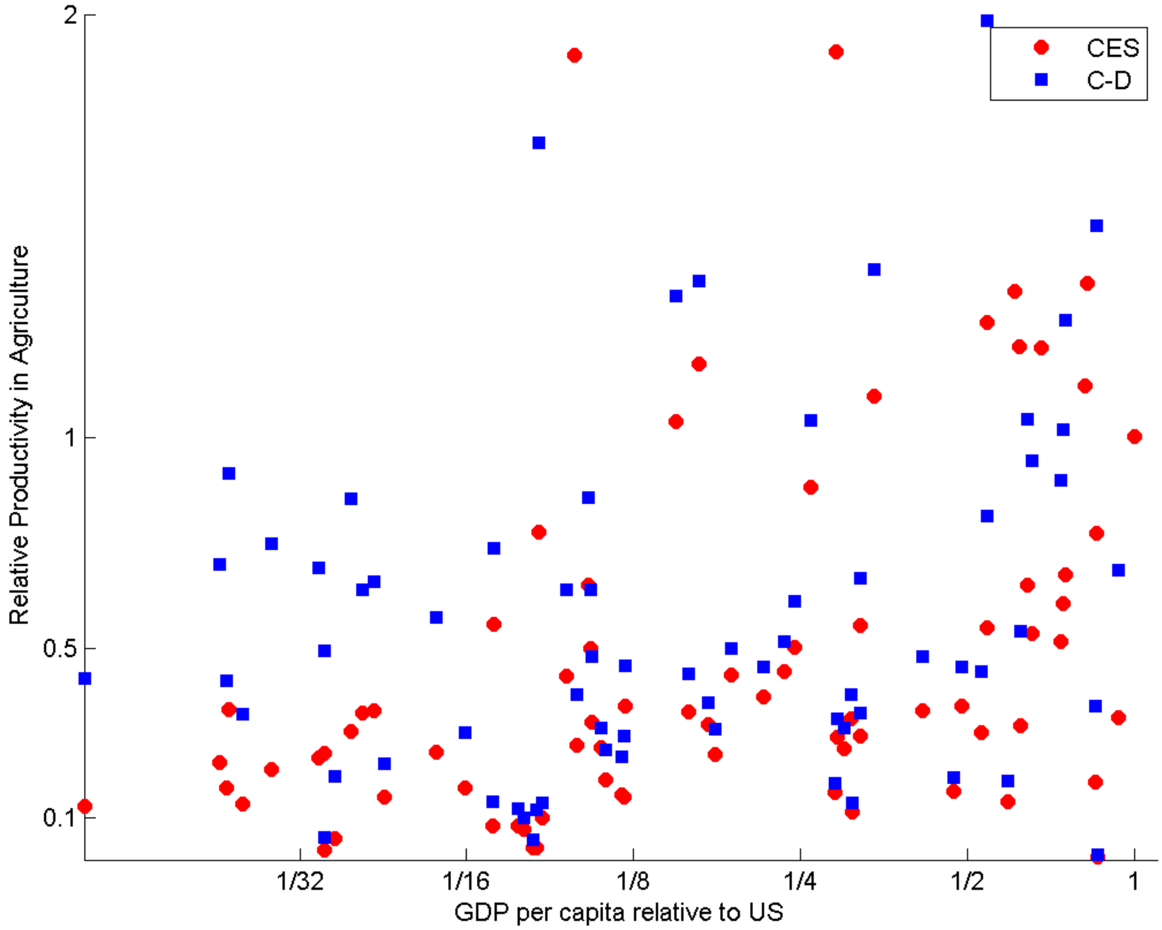
Figure 3.6: Model Prediction of Labor Share of Income



agricultural sector and larger labor share of income in agriculture, which helps to keep the aggregate labor share relatively stable across countries.

Lastly, I examine whether capital-labor substitution helps explain low agricultural productivity in poor countries. Figure 3.7 compares the relative TFP in agriculture (κ) for different model specification. Two points are worth mentioning. First, large TFP differences are needed to get the agricultural productivity right, which is why recent studies tries to model TFP differences (e.g., Adamopoulos and Restuccia (2014)). Second, the Cobb-Douglas technology actually performs better than CES, as the inferred relative TFP is larger under the former. This is because capital intensity is also much larger in agriculture.

Figure 3.7: Relative TFP in agriculture, CES VS. C-D



The same difference in capital intensity leads to larger productivity difference in agriculture. To summarize, besides APG, CES does not help explain other development facts regarding agriculture. This is consistent with the findings of Herrendorf et al. (2015) using U.S. time series data.

3.5 Labor Share of Income and Sectoral Capital Intensity

This section presents further evidence for the capital-labor substitution mechanism. I first examine whether APG is tightly linked to the the ratio of sectoral labor shares as predicted by the model. Data on sectoral labor share is not available for a large number of countries,

Figure 3.8: APG VS. Ratio of Labor Share, EU-KLEMS data

in particular the poor countries. I hence use the EU KLEMS database (O'Mahony and Timmer, 2009), which covers the OECD countries for the period 1970-2005. The OECD countries are mostly rich countries but the data does show some variation in APG across countries and over time. Figure 3.8 plots APG against the ratio of non-agricultural labor share to agricultural labor share.[23] The ratio of labor share does track APG closely.[24] APG however is in general larger than the ratio of labor shares, indicating possible measurement errors or distorted labor market.

Labor shares are closely tied to capital-labor ratios. In particular, the model predicts that capital intensity in non-agriculture relative to that in agriculture decreases with aggregate capital intensity. I next examine this using the dataset of Butzer et al. (2010), who

[23] Labor share is defined as labor compensation (LAB) divided by gross value added(VA). Labor productivity is calculated as value added divided by total hours worked by persons engaged (H_EMP). Agriculture corresponds to the industry agriculture, hunting, forestry and fishing (AtB), and non-agriculture is defined as the difference between total industries (TOT) and agriculture.

[24] The correlation coefficient between the two is 0.67, and it's statistically significant at 1%.

Table 3.5: Relative Capital Intensity VS. Income Levels

	(1)	(2)	(3)
Log of GDP per capita	-7.03	-24.12	-2.99
	(3.29)	(5.32)	(0.41)
Time Dummies	No	Yes	-
R-squared	0.23	0.24	0.68
# of Obs.	520	520	80
Data Source	Butzer et al. (2010)	Butzer et al. (2010)	Model

Note: the estimate for the constant term is omitted here, standard errors in parentheses.

provide internationally comparable measures of fixed capital in agriculture for a set of 30 countries from 1967-2003. These countries include not only rich countries but also some poor countries. I complement this data with total capital stock from PWT and employment share of agriculture from WDI to reach a measure of relative capital intensity. Non-agricultural capital stock is defined as the difference between total and agricultural capital stock.[25] Since employment share of agriculture is often missing for poor countries and for early periods, I end up with an unbalanced panel with 520 observations, covering 28 countries from 1980-2003. Even the sample covers mainly rich countries, it does include observations for poor countries such as Indonesia, Sri Lanka, and Pakistan, where agriculture takes up around half of total employment. I regress the relative capital intensity on the log of GDP per worker, which is derived from PWT. The results are shown in Table 3.5. Column 1 shows that the relative capital intensity decreases along economic development as the model predicts. The estimate is statistically significant at 1% level.[26] Column 2 adds time dummies to control for measurement errors in each year, as the two capital stock might still not be comparable after our adjustment.[27] Adding time dummies increases both the level and statistical significance of the estimate. To make a comparison, I also run the same regression using model simulated data in Column 3. Curiously, the decline in relative capital intensity is larger in

[25] To make the two capital stocks as comparable as possible, I convert PWT's capital stock evaluated at current price PPPs (ck) into 1990 prices using the capital price for US (pl_k).

[26] The significant effect mostly comes from cross-section variation, as the between group estimator produces much larger and more significant estimates then the within group estimator.

[27] See footnote 25. In particular, total aggregate stock is comparable across countries but not so much over time. We should expect the time dummies to catch this effect.

data, despite the model produces a larger effect on APG. I don't want to overemphasize this contradiction as the data is not of highest quality. Also, the model abstracts some features that can alter the relationship between capital intensity and labor shares. The next section discusses some of these features. I conclude this section by emphasizing that the evidence presented here does lend support to the model prediction at least qualitatively.

3.6 Discussion

The baseline model is surely oversimplified. I now discuss some possible extensions. In particular I will focus on how the results on APG will be altered.

3.6.1 Land in Agriculture Production

I have only considered reproducible capital which can be freely allocated between sectors. In reality land plays an important role in agricultural production. For instance, Valentinyi and Herrendorf (2008) show that the high capital share in agriculture is mainly explained by land rents. Adding land to production probably helps bring the predicted APG for poor countries down because land is going to take up a share of output, which can come from labor. I make the extension here by assuming agricultural production employs capital, labor, and land under a nested CES production function. There are 3 possible nesting structure. I consider 2 of them as follows,

$$Y_a = Ak\left[\theta_a\left(\gamma K_a^{\frac{\eta-1}{\eta}} + (1-\gamma)L^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}\frac{\sigma_a-1}{\sigma_a}} + (1-\theta_a)N_a^{\frac{\sigma_a-1}{\sigma_a}}\right]^{\frac{\sigma_a}{\sigma_a-1}}, \tag{3.12}$$

$$Y_a = A\kappa\left[\gamma L^{\frac{\eta-1}{\eta}} + (1-\gamma)\left(\theta_a K_a^{\frac{\sigma_a-1}{\sigma_a}} + (1-\theta_a)N_a^{\frac{\sigma_a-1}{\sigma_a}}\right)^{\frac{\sigma_a}{\sigma_a-1}\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}. \tag{3.13}$$

The first technology combines land and physical first as in Adamopoulos and Restuccia (2014). This makes sense as land is often treated as a form of capital, and both land and physical capital are operated by labor. The second technology nests physical capital and

Table 3.6: Estimates for 3-factor Agricultural Production

	Eq 3.12	Eq 3.13
σ_a	1.85 (0.049)	0.93 (0.041)
η	0.44 (0.046)	1.08 (0.029)
γ_k	0.012 (0.0016)	0.031 (0.021)
γ_n	0.023 (0.0035)	0.096 (0.020)
γ_l	0.034 (0.0025)	-0.041 (0.022)

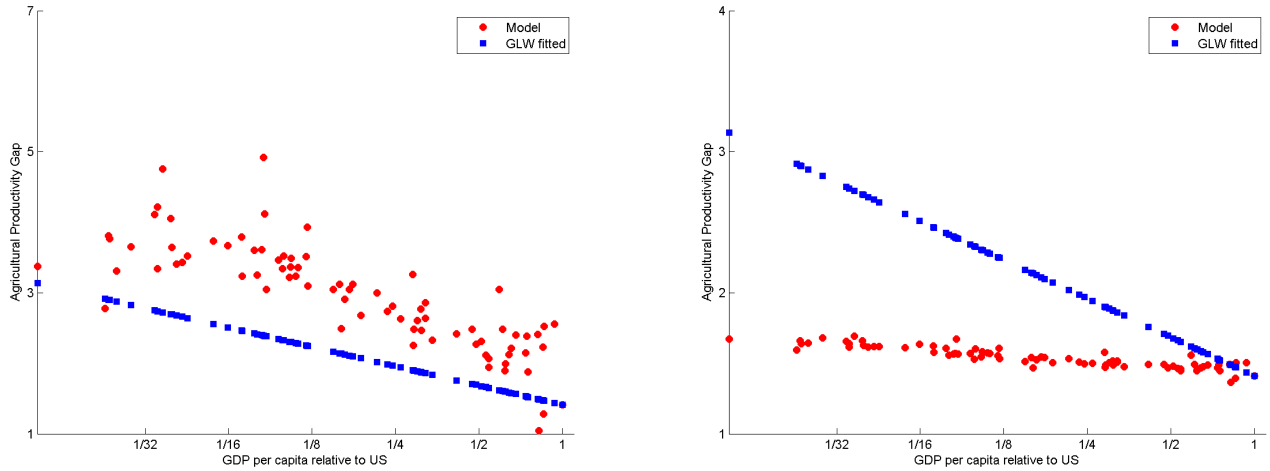
Note: standard errors in parentheses.

labor together. It recognizes the fact that capital and labor might provide services together, which is then applied to land (Kislev and Peterson, 1982).

I still use the data and the estimation procedure of Herrendorf et al. (2015) to estimate the production functions except now I break down capital income into those from reproducible capital and land rents. The normalization is altered following León-Ledesma et al. (2011) who also estimate a two-level nested CES production function using a normalized system. The system is estimated combining the two sectors together, though I only report the results for agriculture. The estimates for the non-agricultural sector change only a little. Under the first nesting structure, we can see that the elasticity of substitution between capital and labor is larger with land added to production, while land is a complement to physical capital. Under the second nesting structure, capital and labor become complements though the elasticity of substitution is still larger than in non-agriculture. The estimates for productivity growth rate however seems less probable. In particular, land productivity declines at 4.1% per year, which is significant at 10% level. The first technology seem to be a more reasonable choice.

I next repeat the numerical exercise using the new production technologies. Figure 3.9 presents the results on APG. Naturally, we see that the model with the first technology produces even larger APG for developing countries. On the other hand, the explanatory power of the model for APG is substantially reduced when we use the second technology. Adding land to agricultural production thus delivers different results, depending on the

Figure 3.9: APG with Land in Agricultural Production



Note: left panel is for the first technology (Eq. 3.12), and the right panel is for the second technology (Eq. 3.13).

specification of technology. Under the more reasonable case capital-labor substitution is still an important source of large APG in developing countries.

3.6.2 Factor-augmenting Productivity

I have omitted factor-augmenting productivity but only relied on Hicks-neutral productivity. This is partly due to data constraint. The literature has long recognized that factor-augmenting productivity can be a key driving force behind changes in factor shares, which is particularly a problem in production function estimation. I now discuss how the results will be affected if we allow for factor-augmenting productivity. I first consider another extreme case where only labor-augmenting productivity is allowed. This is the natural case to consider as labor-augmenting productivity is consistent with balanced growth path in one-sector models. I modify the production functions as follows,

$$Y_n = \left[\theta_n K_n^{\frac{\sigma_n - 1}{\sigma_n}} + (1-\theta_n)(AN_n)^{\frac{\sigma_n - 1}{\sigma_n}}\right]^{\frac{\sigma_n}{\sigma_n - 1}}, \tag{3.14}$$

$$Y_a = \left[\theta_a K_a^{\frac{\sigma_a - 1}{\sigma_a}} + (1-\theta_a)(A\kappa N_a)^{\frac{\sigma_a - 1}{\sigma_a}}\right]^{\frac{\sigma_a}{\sigma_a - 1}}. \tag{3.15}$$

The effect of labor-augmenting productivity is intuitive. If poor countries also have low productivity, factor shares might change less as what matters is the ratio of capital to effective labor.[28] This potentially can help get the predictions on APG right, or even reverse the results. A problem with this modification however is that the calibration strategy in Section 3.4 might not produce sensible results. Notice that there we need large TFP differences to match labor productivity in poor countries. To match the same data in the new specification can be unfeasible as it is possible that using capital alone is more than necessary to produce the output a developing country can produce.[29] Labor-augmenting productivity hence will be negative if we require labor productivity in the model to match data.[30] Indeed, this is a serious problem in the sample of countries, the calibration is successful for only 16 out of 80 countries.[31] I note that predicted APG is still negatively correlated with income levels, even though now the correlation is less systematic than the baseline case.

We have seen in Section 3.5 that relative capital intensity in data is larger than model prediction and the gap is larger for developing countries. For example, relative capital intensity in U.S. is 0.74 in the model, while it is 1.13 in data. For a poor country like Morocco, they are 4.34 and 46! This could be due to factor-augmenting productivity differences. But will this overturn the results on APG? The answer is no. The reason is that no matter how factor prices change, the changes in capital intensity due to factor-augmenting productivity are going to increase the labor share of agriculture relative to that of non-agriculture if the model is going to match data on sectoral capital intensity. On the other hand, changes in factor prices might change the results. For that, I consider the following thought experiment. Without re-calibrating the model, let's assume that the model predicts a constant aggregate

[28]That is, capital intensity might increase less if increase in effective labor reduces the need of substituting capital for labor.

[29]This is because neither capital or labor is necessary for producing output in CES production functions.

[30]At a deeper level, the problem is not concave if we have to change A and κ to match data on labor productivity. For example, if we increase rental rate of capital, we can show that the demand for capital might also increase due to corresponding shifts in production function.

[31]To solve the problem, I use a grid search over employment share in agriculture, given which all other variables can be solved. The grids range from 0 to 1 with a grid size of 0.0001. An equilibrium exists if the equilibrium conditions holds roughly.

labor share of 0.67 for all countries and the allocation of capital and labor is the same as in data.[32] This pins down the wage-rental rate ratio, as[33]

$$\frac{w}{r} = \frac{ls}{(1 - ls)} \frac{K}{N_a + \frac{N_n}{1-\xi}},$$

where ls is the aggregate labor share. Given that the allocation of capital and labor is exactly what we observe in data, what are the predictions for APG? To get that, I apply the following formula to all the countries with comprehensive data,

$$APG = \frac{\frac{K_n}{N_n}\frac{r}{w} + \frac{1}{1-\xi}}{\frac{K_a}{N_a}\frac{r}{w} + 1}.$$

I find that predicted APG is still negatively correlated with income levels. In particular, it is higher than 10 for 14 out of 27 countries, and it can be as large as 150! I find these numbers implausibly large, indicating substantial errors in the capital measures.

The first experiment in this subsection shows that even labor-augmenting productivity might play a role, it cannot account for the large international differences in labor productivity alone. The second experiment shows that a model calibrated to target capital allocation in the data provided by Butzer et al. (2010) will still over-predict APG for poor countries, indicating potentially large measurement errors in the capital measures. The conclusion is that factor-augmenting productivity might get the predicted APG right. The data from Butzer et al. (2010) however is not good enough for a quantitative study.

3.6.3 Investment

The model is static such that investment is not explicitly considered. With investment, output will be different from consumption. The model thus implicitly requires the investment rates in the two sectors follow a certain rule. To see that, notice the first order condition

[32] One property of a successful model would be a relative stable aggregate labor share (Gollin, 2002). The baseline model shows the predictive power for labor allocation. Capital allocation can also be targeted if we were to use that information.

[33] Notice the calibration of ξ will not be affected by adding factor-augmenting productivity.

with investment changes to

$$\frac{\omega}{1-\omega}\frac{((1-inv_a)Y_a)^{-\alpha_a}}{((1-inv_n)Y_n)^{-\alpha_n}} = p,$$

where inv_i is investment rate in sector i. Ignoring investment doesn't affect model prediction only if $(1-inv_a)^{\alpha_a} = (1-inv_n)^{\alpha_n}$, which implies $inv_n > inv_a$. As investment is commonly believed to come from the manufacturing sector (Herrendorf et al., 2014), the bias due to ignoring investment might be small. In general, raising investment rate in the non-agricultural sector increases the demand for the non-agricultural good, which further increases the rental rate of capital, as non-agriculture is more capital intensive than agriculture in poor countries. The changes in factor prices will increase predicted APG as it further deviates the wage-rental rate ratio from the U.S. benchmark. Adding investment hence will not overturn the prediction for APG but reinforces it.

3.6.4 Trade and International Capital Flow

How does opening to international markets change my results? I next analyze two cases: one with international trade and the other with international capital flow. I focus on what will happen to developing countries.

What happens if a poor country opens to international trade? Given the low agricultural productivity, relative price of the agricultural good is high in poor countries (Restuccia et al., 2008; Lagakos and Waugh, 2013). This means poor countries will import food and export industrial products.[34] In my model, opening to international trade lowers the wage-rental rate ratio, which tend to raise model predicted APG in poor countries. To see that, notice that we can write the resource constraint for capital as

$$\left(\frac{K_n}{N_n}\right)N_n + \left(\frac{K_a}{N_a}\right)N_a = K.$$

[34] In data poor countries don't import as much food as they should, as trade costs are prohibitively high in agriculture (Tombe, 2015). The fact that poor countries allocate a lot workers to agriculture is a reflection of high trade costs.

Given that the country will export non-agricultural good and import agricultural good, we expect N_n to increase and N_a to decrease. Since $\frac{K_n}{N_n} > \frac{K_a}{N_a}$ in developing countries, we will have both of them decrease, otherwise the resource constraint cannot be met.[35] This leads to a decrease in labor share in non-agriculture and an increase in agriculture, contributing to a even larger APG. Opening to international trade thus only strengthens my results. The fundamental reason behind this result is international trade benefits capital in capital-scarce countries, which leads to a lower wage-rental rate ratio than under a closed economy. This seemingly counter-intuitive result makes sense because the source of comparative advantage mostly comes from TFP differences rather than factor endowment. In rich countries, the opposite will happen. The dispersion in APG across countries will increase.

What about opening to the international capital market? The model predicts that capital will move from rich to poor countries as returns are higher in poor countries. The wage-rental rate ratio will increase in poor countries, which will reduce the APG. However, the difference in relative factor price will not disappear because of capital flow alone. Factor prices will not be equalized across countries because countries differ in TFP. Workers in rich countries benefit from high TFP and receive higher wages. Efficient allocation of resources requires both capital and labor move to rich countries instead of capital moving to poor countries. Opening to international capital inflow however contributes to a decrease in the dispersion of APG, which might help resolving the over-prediction of the model.

3.7 Summary

This chapter examines whether differential capital-labor substitution across sectors helps explain some of the development facts regarding agriculture. I find that it helps explain the negative relationship between the agricultural productivity gap and GDP per capita. While Gollin et al. (2014) conclude that difference in sectoral labor shares cannot explain the agricultural productivity gap, I find that the *changes in sectoral labor shares* might

[35] This relies on a shift of labor from agriculture to industry. This is necessary because if it's not the case, increased non-agricultural production require a larger increase in capital intensity in non-agriculture, which leads to an even larger increase in agricultural labor share. The resource constraint again cannot be met.

nevertheless help explain the cross-country pattern of APG. Since sectoral labor share data covering a large set of countries are not readily available, we however cannot directly test this model prediction. Fractional evidence in Fuglie (2010) supports the model. Quantitatively, my baseline model even over-predicts the negative correlation, which can be attenuated by adding more realistic features to the model. In terms of sectoral allocation of resources and sectoral labor productivity, I find that the effect of capital-labor substitution is quantitatively small, conforming the findings of Herrendorf et al. (2015).

The evidence presented in this chapter is hardly conclusive due to the lack of good data. What's clear is that relative capital intensity and labor share of income do change over time, and the change is different across industries. This cautions using Cobb-Douglas production functions in the study of agriculture and development. I show that it is in particular the case when we try to attribute labor productivity differences across sectors to distortions.

3.8 Data Appendix

For the quantitative analysis, I construct a dataset countries for the year 1985. The dataset includes each country's agricultural and non-agricultural real GDP per capita, employment and value-added shares of both sectors, relative prices, capital stock, and land endowment per capita. Data on agricultural output come from Food and Agriculture Organization of the United Nations (FAO, Rao (1993)). Aggregate output data come from Penn World Table 8.1 (Feenstra et al., 2015). While both the FAO and PWT provide real output, they are not comparable because the international prices used are not comparable. I use the procedure in Caselli (2005) to adjust the FAO data to make it comparable to PWT. The non-agricultural output then is derived by subtracting agricultural output from aggregate output. The employment data also come from FAO and PWT. Because FAO reports economically active population while PWT only considers people who work, I use International Labor Organization's (ILO) data on economically active population to adjust the PWT data and make economically active population as my measure of employment. The source of value-added shares is the United Nations National Accounts, which I download from the online data appendix of Herrendorf et al. (2014). Relative prices are derived as the ratio of PPPs for the two sectors. Finally, I use PWT's capital stock measure. The land data also comes from FAO. The dataset has 80 countries with non-missing values for all variables.

CHAPTER 4

Value-added, Production Networks, and Misallocation

4.1 Introduction

Both the previous two chapters provide explanations for the agricultural productivity gap, which has often been as an indicator of severe labor misallocation in developing countries Gollin et al. (2014). While my study does not favor the misallocation explanation for the productivity gap, a recent literature pioneered by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) does argue forcefully and provide much evidence that resource misallocation could be an important reason behind the staggering productivity differences between developed and developing countries. In particular, Hsieh and Klenow (2009) develop an accounting framework to estimate the efficiency loss resulted from misallocation across micro production units such as firms, which is followed by later studies such as Oberfield (2013), Hsieh and Klenow (2014), and Gopinath et al. (2017). While their approach relies on value-added production functions and ignores intermediate input use in production, later studies also have pointed out that the effect of given micro distortions can be magnified through the linkages across sectors (Jones, 2011, 2013). Without taking firm productivity and distortions as given, this chapter asks and answers a different question: will a model with production networks produce larger efficiency loss than a model with value-added production functions *conditional on observing the same data*?

To answer this question requires correct specification of both models. While we can naturally view sectors as either categories of final expenditure or value-added that are connected through complicated input-output linkages,[1] Herrendorf et al. (2013) show that it

[1] An example given in Herrendorf et al. (2013) is a cotton shirt. While it is a product of the manufacturing

is not easy to specify tractable models that can be easily transformed into the other form. If the models are mis-specified, the question raised can not be properly answered because we can not distinguish between model mis-specification and the use of different production functions. The first contribution of this chapter is to show that the widely used model of production networks with Cobb-Douglas production technologies at the sectoral level *a la* Long and Plosser (1983) (the output model hereafter) is theoretically consistent with the model of Hsieh and Klenow (2009) (the value-added model) when there are no distortions in intermediate input use.[2] This provides the basis for the comparison of these two models.

With the equivalence of the two models, I proceed to show that efficiency loss produced in these two models are identical when there are no distortions in the intermediate input use. Despite the magnification of distortions through the production network, the output model doesn't produce larger efficiency loss as firm productivity measured under an output production function is substantially less dispersed that that measured under a value-added production function, which lowers the effect of misallocation. This scaling effect has been pointed out by Bruno (1978) in comparing the real value-added production and the output production function. It has also recently been emphasized by Gandhi et al. (2013, 2017) as a reason that the value-added framework might overstate the extent of misallocation in data, though they do not consider production networks.

The scaling effect can be intuitively understood as follows. A rise in productivity will induce the firm to employ more intermediate inputs and raise its output and value-added. Value-added-based productivity will register a larger increase in productivity because it doesn't take into account the increase in intermediate input use, which is correctly accounted for in the output-based productivity measures. It is essentially the magnification effect of production networks, now implicitly reflected in measured firm productivity, while the same effect is explicitly spelled out in the output model. It is another manifestation that the two

sector under the final expenditure view, it is a combination of raw cotton from agriculture, processing from manufacturing, and retail services from the services sector under the value-added view.

[2] A minor modification is required: the elasticity of substitution between products of firms within a sector should be different. Hsieh and Klenow (2009) assume a constant elasticity. I'll come back to this point when I discuss empirical implementation of the models.

models are just different representations of the same production process.

When distortions in intermediate input use are introduced, the value-added model is no longer a correct representation of data and will not produce the correct measure of efficiency loss. I explore how the results might be biased in this scenario if the value-added model is nevertheless specified. I go on to analyze a case where there is no misallocation across sectors. This is the case considered by Hsieh and Klenow (2009) and many others.[3] It also allows us to identify the production network and makes the implementation feasible. Surprisingly, efficiency loss in the value-added model actually has the theoretically correct form in terms of revenue and physical productivity of firms, despite a value-added representation of the data does not exist. Though measured efficiency loss will be biased when the model is implemented in data. The reason for the bias is that both revenue productivity and physical productivity are incorrectly measured in the value-added model. Since I assume that intermediate inputs observed in the data are recorded in the same prices,[4] one way to understand this mismeasurement is to think of the commonly used value-added measure (output minus intermediate input) as a real value-added measure constructed with constant prices while the firms actually face different nominal prices due to the presence of distortions. As argued by Bruno (1978), when prices are not constant, both marginal products and total factor productivity measured using real value-added will be biased. The same logic underlies the findings of Basu and Fernald (1995) that productivity spillover across manufacturing industries exist in the value-added data but not in the output data. They consider imperfect competition as a particular source of price distortion, while the distortions in this chapter are generic and I apply the idea to misallocation accounting.

While efficiency loss is mismeasured in the value-added model with distortions in intermediate input use, the model does not predict a unambiguous direction for the bias. It depends on how the distortions are distributed in data. I next empirically measure efficiency loss in China under these two models using the Chinese Annual Survey of Industrial Production in

[3]In the contrary, Jones (2011) studies only misallocation across sectors.

[4]That is, the distortions considered in the chapter are implicit, or the data is net of taxes if the distortions come from actual taxes.

2005, which is also used by Hsieh and Klenow (2009) among others. When the value-added model is calibrated in a way consistent with the theoretical discussion, I find that there are biases in both directions across 4-digit industries of China. The biases however are relatively small. The difference between the measures from the two models is less than 5 percentage points for over 95% of the industries. This is because measured distortions in the intermediate input markets are substantially smaller than that in the capital and labor markets, a finding that is consistent with Krishna and Tang (2018). The standard deviation of the distortions in intermediate input market (in logs) is only 0.19, while that in the primary input market is 0.85.

Existing literature however might have overstated the cost of misallocation due to misspecified parameter values. A key parameter for measuring efficiency loss is the elasticity of substitution between varieties within a sector. This parameter is often assigned wrong values in the literature in two ways: 1. it should be sector-specific, and 2. it is different in the two models. While the bias coming from the first source is small, efficiency loss can be substantially overstated when an output elasticity is assigned to a value-added model. In the Chinese data, if I let the elasticity to be 3 in both models,[5] measured efficiency loss is larger in the value-added model in over 95% of the industries and the value-added measures are on average 14 percentage points larger than the output measures. The reason behind this overstatement is that the output elasticity is much smaller than the value-added elasticity. The distinction between these two parameters is emphasized by Herrendorf et al. (2013) who discuss the two views of sectors. It however has not be fully appreciated in the literature. While the elasticity is often estimated for goods (Broda and Weinstein, 2006; Hendel and Nevo, 2006), these estimates are often used in the value-added model without deliberation.

This chapter contributes to a burgeoning literature on misallocation, in particular those studies using micro data to measure aggregate efficiency loss following Hsieh and Klenow (2009).[6] It is also closely related to a newer literature studying misallocation in production

[5]This is the baseline case in Hsieh and Klenow (2009). The overstatement still exists if other values are assigned.

[6]It is part of the "indirect approach" under the taxonomy of Restuccia and Rogerson (2013). Studies

networks.[7,8] This chapter focuses on the measurement of efficiency loss under production networks and compared it to the value-added model of Hsieh and Klenow (2009). It is close to Krishna and Tang (2018), who also measure misallocation for China and India using an output model similar to mine. They find that measured efficiency losses in the two countries are not necessarily higher than those in Hsieh and Klenow (2009). This chapter provides a theoretical underpinning for their findings. The comparison between the use of output and value-added measures also echoes the early studies of Bruno (1978, 1984), and Basu and Fernald (1995). Different from them, this chapter explicitly considers production networks and studies misallocation across firms.[9] The two views of sectors are also extensively discussed in Herrendorf et al. (2013) who try to disentangle different sources of structural transformation. This chapter applies the idea to the measurement of misallocation and explicitly models the input-output linkages.

The rest of the chapter is structured as follows. The next section presents the output model, which is followed by the value-added framework in Section 4.3. Section 4.4 shows that the output model can be transformed into the value-added model and efficiency loss measured in the two models are identical if there are no distortions in intermediate input use. Section 4.5 discusses how the value-added model will produce biased results when those distortions are present. Section 4.6 evaluates the two models empirically using Chinese data. Section 4.7 concludes.

---

along this line include Dollar and Wei (2007), Alfaro et al. (2009), Kalemli-Özcan and Sorensen (2012), Ziebarth (2012, 2013), Brandt et al. (2013), Oberfield (2013), Hsieh and Klenow (2014), Adamopoulos et al. (2015), Chen and Irarrazabal (2015), ,Gopinath et al. (2017), and Restuccia and Stantaeulàlia-Llopis (2017), among others.

[7]This line of research includes Jones (2011, 2013), Bartelme and Gorodnichenko (2015), Bigio and LaO (2016), Baqaee and Farhi (2017a), Caliendo et al. (2017), Liu (2017), and Krishna and Tang (2018).

[8]Other important contributions to the misallocation literature include Banerjee and Duflo (2005), Restuccia and Rogerson (2008), Guner et al. (2008), Banerjee and Moll (2010), Bartelsman et al. (2013), Buera and Shin (2013), Asker et al. (2014), Midrigan and Xu (2014), Moll (2014), David et al. (2016), David and Venkateswaran (2017), among many others. Sectoral interlinakges have also been studied by Hirschman (1958), Hulten (1978), Long and Plosser (1983), Horvath (1998, 2000), Ciccone (2002), Acemoglu et al. (2012), Baqaee (2016), and Baqaee and Farhi (2017b), to name a few.

[9]The other studies mainly focus on measuring productivity growth over time, while this chapter compares productivity across firms.

4.2 The Output Model

This section presents the output model, which comes from adding a production network *a la* Long and Plosser (1983) to the Hsieh and Klenow (2009) framework. I view this model as a natural representation of production in an economy, as firms treat primary inputs and intermediate inputs symmetrically in reality. From the point of view of the aggregate economy, the concept of value-added however is more natural as the intermediate inputs cancel out in the aggregation process. This makes sure that the comparison between the output model present here and the value-added model in the next section is legitimate: both model takes capital and labor to produce a final product (GDP) while one model considers endogenous intermediate input use and the other model does not. The output model corresponds to the "final consumption expenditure approach" of Herrendorf et al. (2013), while the value-added model corresponds to the "consumption value-added approach".[10]

The are S sectors in the economy. A final product is produced by competitive firms from sectoral output,[11]

$$Y = \sum_{s=1}^{S} C_s^{\theta_s}, \text{ with } \sum_{s=1}^{S} \theta_s = 1. \tag{4.1}$$

Let the final output be the numeraire, the demand for sectoral output from final good producers is given by $P_s Y_s = \theta_s Y$. The sectoral output is used both for final consumption and as intermediate inputs in production, with the market clearing condition,

$$Q_s = C_s + M_s, \tag{4.2}$$

where M_s is intermediate input demand from all sector, $M_s = \sum_{q=1}^{S} M_{sq}$, where M_{sq} is the demand of sector q for the output of sector s. The sectoral output is produced using different

[10] Note Herrendorf et al. (2013) only considers consumption as their purpose is to estimate utility functions without considering which sector investment sources from. In this chapter both consumption and investment are taken as part of total expenditure.

[11] I use C to denote consumption and will call it consumption, but it should be understood as expenditure including both consumption and investment, see Footnote 10.

intermediate varieties Q_{si},

$$Q_s = \left(\sum_{i=1}^{N_s} Q_{si}^{\frac{\sigma_s - 1}{\sigma_s}} \right)^{\frac{\sigma_s}{\sigma_s - 1}}, \tag{4.3}$$

where N_s is the total number of varieties in sector s. I assume each variety is produced by one firm such that firm and variety can used interchangeably. Note that the elasticity of substitution between varieties is sector specific. This is not necessary for the output model but I will later show that the elasticity cannot be constant in both models. The inverse demand for a variety is given by $P_{si} = P_s Q_s^{\frac{1}{\sigma_s}} Q_{si}^{-\frac{1}{\sigma_s}}$, with the sectoral price index given as $P_s = \left(\sum_{i=1}^{S} P_{si}^{1-\sigma_s} \right)^{\frac{1}{1-\sigma_s}}$.

The firms produce with two primary inputs, capital and labor, and intermediate inputs from all sectors, according to a Cobb-Douglas technology

$$Q_{si} = A_{si} \left(K_{si}^{\alpha_s} L_{si}^{1-\alpha_s} \right)^{\eta_s} \left(\prod_{q=1}^{S} M_{qsi}^{\lambda_{qs}} \right)^{1-\eta_s}, \tag{4.4}$$

with $\sum_{q=1}^{S} \lambda_{qs} = 1$. Different from Hsieh and Klenow (2009), I assume the firms take prices as given instead of actively set prices.[12] They face idiosyncratic distortions in the factor markets, trying to maximize,

$$\pi_{si} = P_{si} Q_{si} - (1 + \tau_{Ksi}) R K_{si} - (1 + \tau_{Lsi}) W L_{si} - (1 + \tau_{Msi}) \sum_{q=1}^{S} P_q M_{qsi}.$$

The key to equivalence results presented below is there are no distortions in intermediate input use, which I make explicit in the following assumption.

Assumption 2 *$\tau_{Msi} = 0, \ \forall \ s \in \{1, 2, ...S\}$ and $i \in \{1, 2, ...N_s\}$.*

To prove the two models represent the same production structure, it is however not necessary to introduce the distortions but adding them shows that the results are not affected

[12] One way to think about this is there are many firms with access to identical technology with tiny entry costs. In equilibrium there will be only one firm operating but act like competitive producers. Otherwise, I can assume there are many firms producing one variety and they all have the same technology and face the same distortions. Assuming monopoly pricing does not change main conclusions in this chapter but only alters the constant terms in the aggregate production derived below.

by distortions in primary factor markets. It also facilitates the presentation of results below. Following Hsieh and Klenow (2009), I assume that the payments to factors are recorded in market prices net of the wedges. One way to think of this to view the wedges as the shadow prices for the constraints the firms face in the factor markets, such as a collateral constraints for renting capital. This however runs the risk of having firms with negative profits if they face negative wedges. Giving the firms monopoly power can ease the problem but will not solve it altogether. To assume the taxes and subsidies are rebated to the firms lump sum however guarantees all firms receive zero profit. Either way, the allocation of production factors is not affected. Finally, factor market clearing for capital and labor requires,

$$\sum_{s=1}^{S}\sum_{i=1}^{N_s} K_{si} = K, \text{ and } \sum_{s=1}^{S}\sum_{i=1}^{N_s} L_{si} = L. \tag{4.5}$$

Within a sector, sectoral intermediate demand is the sum of demand from all variety producers $M_{qs} = \sum_{i=1}^{N_s} M_{qsi}$.

I next proceed to show that the model allows an aggregate production function with a Cobb-Douglas form. Following Hsieh and Klenow (2009), I define marginal revenue products as follows,

$$MRPK_{si} \equiv \alpha_s \eta_s \frac{P_{si}Q_{si}}{K_{si}} = (1+\tau_{Ksi})R, \tag{4.6}$$

$$MRPL_{si} \equiv (1-\alpha_s)\eta_s \frac{P_{si}Q_{si}}{L_{si}} = (1+\tau_{Lsi})W, \tag{4.7}$$

$$MRPM_{qsi} \equiv \lambda_{qs}(1-\eta_s)\frac{P_{si}Q_{si}}{M_{qsi}} = P_q, \tag{4.8}$$

where the second equality simply comes from the first order conditions of the profit maximizing firms. With the wedges, the marginal revenue products are no longer the same across firms, which is required for efficient resource allocation. I also define value productivity and physical productivity as follows,

$$TFPR_{si} \equiv \frac{P_{si}Q_{si}}{\left(K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}\right)^{\eta_s}\left(\prod_{q=1}^{S}M_{qsi}^{\lambda_{qs}}\right)^{1-\eta_s}}, \tag{4.9}$$

$$TFPQ_{si} \equiv \frac{Q_{si}}{\left(K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}\right)^{\eta_s}\left(\prod_{q=1}^{S}M_{qsi}^{\lambda_{qs}}\right)^{1-\eta_s}} = A_{si}. \tag{4.10}$$

Revenue productivity is simply a firm's marginal cost of a production bundle such that we can express price as $P_{si} = \frac{TFPR_{si}}{A_{si}}$.[13] It also contains all the information on the wedges, each of which weighted by the output elasticity of the corresponding production factor. Without the wedges, revenue productivity is equalized across firms within a sector, which leads to the efficient allocation of resources. Under efficient allocation, more productive firms employs more inputs and produce more, which lowers the price of its product eventually to the point of equal revenue productivity across firms.

For a sector as a whole, value productivity can be similarly defined as[14]

$$TFPR_s \equiv \frac{P_sQ_s}{(K_s^{\alpha_s}L_s^{1-\alpha_s})^{\eta_s}\left(\prod_{q=1}^{S}M_{qs}^{\lambda_{qs}}\right)^{1-\eta_s}}, \tag{4.11}$$

which allows us to express sectoral total factor productivity as,

$$TFP_s \equiv \frac{Q_s}{(K_s^{\alpha_s}L_s^{1-\alpha_s})^{\eta_s}\left(\prod_{q=1}^{S}M_{qs}^{\lambda_{qs}}\right)^{1-\eta_s}} = \frac{TFPR_s}{P_s}.$$

[13]From the first order conditions, we have $TFPR_{si} \propto \left(MRPK_{si}^{\alpha_s}MRPL_{si}^{1-\alpha_s}\right)^{\eta_s}\left(\prod_{q=1}^{S}MRPM_{qsi}^{\lambda_{qs}}\right)^{1-\eta_s}$.

[14]We have $TFPR_s \propto \left(MRPK_s^{\alpha_s}MRPL_s^{1-\alpha_s}\right)^{\eta_s}\left(\prod_{q=1}^{S}MRPM_{qsi}^{\lambda_{qs}}\right)^{1-\eta_s}$, with

$$MRPK_s \equiv \alpha_s\eta_s\frac{P_sQ_s}{K_s} = \frac{1}{\sum_{i=1}^{N_s}\frac{1}{MRPK_{si}}\frac{P_{si}Q_{si}}{P_sQ_s}},$$

$$MRPL_s \equiv (1-\alpha_s)\eta_s\frac{P_sQ_s}{L_s} = \frac{1}{\sum_{i=1}^{N_s}\frac{1}{MRPL_{si}}\frac{P_{si}Q_{si}}{P_sQ_s}},$$

$$MRPM_{qs} \equiv \lambda_{qs}(1-\eta_s)\frac{P_sQ_s}{M_{qsi}} = \frac{1}{\sum_{i=1}^{N_s}\frac{1}{MRPM_{qsi}}\frac{P_{si}Q_{si}}{P_sQ_s}}.$$

The sectoral marginal revenue products hence are the weighted average of that of the firms.

Plugging in P_s and using the prices of varieties, sectoral production takes a Cobb-Douglas form as follows.

$$Q_s = TFP_s \cdot \left(K_s^{\alpha_s} L_s^{1-\alpha_s}\right)^{\eta_s} \left(\prod_{q=1}^{S} M_{qs}^{\lambda_{qs}}\right)^{1-\eta_s}, \text{ and } TFP_s = \left[\sum_{i=1}^{N_s} \left(A_{si} \frac{TFPR_s}{TFPR_{si}}\right)^{\sigma_s - 1}\right]^{\frac{1}{\sigma_s - 1}} \tag{4.12}$$

I next study the allocation of resources across sectors and derive an aggregate production function. For this purpose, it is useful to define the Domar weight $v_s = \frac{P_s Q_s}{Y}$, which is simply the sales to GDP ratio. From the market clearing condition for sectoral output, we have[15]

$$V = \theta + BV,$$

where V is the 1 by S vector of Domar weights v_s, θ is the vector of sectoral shares in final consumption θ_s, and the S by S matrix B is the input-output matrix with its sq^{th} element b_{sq} given by $\lambda_{sq}(1-\eta_q)$. We can solve for the Domar weights as $V = (I-B)^{-1}\theta$. Without distortions in intermediate input use, the Domar weights are determined by technology parameters alone. We can then relate the intermediate input demand to sectoral output as,

$$M_{qs} = \lambda_{qs}(1-\eta_s)\frac{P_s Q_s}{P_q} = b_{qs}\frac{v_s}{v_q} Q_q.$$

The allocation of capital and labor across sectors is simply given by,

$$\begin{aligned} \frac{K_s}{K} &= \frac{\frac{1}{\tau_{Ks}}\alpha_s \eta_s v_s}{\sum_{q=1}^{S} \frac{1}{\tau_{Kq}}\alpha_q \eta_q v_q} \equiv \beta_{Ks}, \\ \frac{L_s}{L} &= \frac{\frac{1}{\tau_{Ls}}(1-\alpha_s)\eta_s v_s}{\sum_{q=1}^{S} \frac{1}{\tau_{Lq}}(1-\alpha_q)\eta_q v_q} \equiv \beta_{Ls}, \end{aligned}$$

[15]Remember the market clearing condition for sectoral output is given by

$$P_s Q_s = P_s C_s + \sum_{q=1}^{S} \lambda_{sq}(1-\eta_q) P_q Q_q,$$

where I have plugged in the sectoral demand for intermediate inputs. Dividing the equation by Y and stack it in a vector gives the result.

where $\mathcal{T}_{Ks} = \frac{MRPK_s}{R}$ and $\mathcal{T}_{Ls} = \frac{MRPL_s}{W}$ summarize the effect of distortions on sectoral factor demand. Without distortions, $\mathcal{T}_{Ks} = 1$ and $\mathcal{T}_{Ls} = 1$.

Plugging the allocation of production factors described above into the sectoral production function and taking logs, we have the log of sectoral production function in vector form[16]

$$\bar{q} = \bar{a} + \omega_q + \delta_K \log K + \delta_L \log L + B'\bar{q},$$

where $\bar{q}$ is a vector of $\log Q_s$, δ_K a vector of $\alpha_s \eta_s$, δ_L a vector of $(1-\alpha_s)\eta_s$, and ω_q is a vector of the allocation terms.[17] The vector of log output is solved as

$$\bar{q} = (I - B')^{-1}(\bar{a} + \omega_q + \delta_K \log K + \delta_L \log L)$$

Notice that we can write $C_s = \frac{\theta_s Q_s}{v_s}$. Taking logs and stacking it into a vector, we have

$$\bar{c} = \omega_c + \bar{q},$$

where $\bar{c}$ is the vector of $\log C_s$ and ω a vector of $\log \frac{\theta_s}{v_s}$. Using the final good production function, we have,

$$\log Y = \theta' \bar{c} = \theta'[\omega_c + (I - B')^{-1}(\bar{a} + \omega_q + \delta_K \log K + \delta_L \log L)]$$

which leads to the aggregate production function summarized in the following proposition.

Proposition 5 *The economy has an aggregate production function given by*

$$Y = AK^{\alpha} L^{1-\alpha}, \tag{4.13}$$

[16]The sectoral production function is given by

$$Q_s = TFP_s \cdot \left((\beta_{Ks} K)^{\alpha_s} (\beta_{Ls} L)^{1-\alpha_s}\right)^{\eta_s} \left[\prod_{q=1}^{S} \left(b_{qs} \frac{v_s}{v_q} Q_q\right)^{\lambda_{qs}}\right]^{1-\eta_s}$$

[17]The s^{th} element of ω_q is given by $\alpha_s \eta_s \log \beta_{Ks} + (1-\alpha_s)\eta_s \log \beta_{Ls} + (1-\eta_s) \sum_{q=1}^{S} \lambda_{qs} \log \left(b_{qs} \frac{v_s}{v_q}\right)$.

where total factor productivity is given by

$$A = \gamma \left(TFP_s\right)^{v_s} \prod_{s=1}^{S} \left[\left(\beta_{Ks}\right)^{\alpha_s \eta_s} \left(\beta_{Ls}\right)^{(1-\alpha_s)\eta_s} \right]^{v_s}, \tag{4.14}$$

with γ *being a constant given by* $\prod_{s=1}^{S} \left(\frac{\theta_s}{v_s}\right)^{\theta_s} \prod_{s=1}^{S} \left[\prod_{q=1}^{S} \left(b_{qs} \frac{v_s}{v_q} \right)^{1-\eta_s} \lambda_{qs} \right]^{v_s}$ *and*[18]

$$\alpha = \theta'(I - B')^{-1}\delta_K. \tag{4.15}$$

Notice that I have decomposed the aggregate total factor productivity into three terms. The first is a constant. The second is the weighted average of sectorl TFP where the weights are the Domar weights. This term gives the effect of distortions on within sector allocation. The third reflects the allocation of capital and labor across sectors. If there is no misallocation across sectors, the third term will also be constant in this Cobb-Douglas economy such that the impact of an increase in sectoral TFP is given by its sales to GDP ratio, which is Hulten's theorem (Hulten, 1978).

Efficient allocation across all variety producers requires the marginal revenue products to be equated across firms and sectors. Let Y^E be the efficient output, the efficiency loss from resource misallocation in the output model is given as

$$\left(\frac{Y^E}{Y}\right)^o = \prod_{s=1}^{S} \left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} \prod_{s=1}^{S} \left[\left(\frac{\beta_{Ks}^E}{\beta_{Ks}}\right)^{\alpha_s \eta_s} \left(\frac{\beta_{Ls}^E}{\beta_{Ls}}\right)^{(1-\alpha_s)\eta_s} \right]^{v_s} \tag{4.16}$$

where

$$TFP_s^E = \left[\sum_{i=1}^{N_s} A_{si}^{\sigma_s - 1} \right]^{\frac{1}{\sigma_s - 1}} \tag{4.17}$$

and

$$\beta_{Ks}^E = \frac{\alpha_s \eta_s v_s}{\sum_{q=1}^{S} \alpha_q \eta_q v_q}, \text{ and } \beta_{Ls}^E = \frac{(1-\alpha_s)\eta_s v_s}{\sum_{q=1}^{S} (1-\alpha_q)\eta_q v_q} \tag{4.18}$$

[18] To show that the production function has constant returns to scale, notice that the share of labor is given by $\theta'(I - B)^{-1}\delta_L$, returns to scale is then $\theta'(I - B')^{-1}(\delta_K + \delta_L)$. Since the value-added shares $\delta_K + \delta_L$ are given by $(I - B')\mathbf{1}$ where $\mathbf{1}$ is the vector of ones. It is easy to show that $\theta'(I - B')^{-1}(\delta_K + \delta_L) = 1$.

The two terms in the efficiency loss measure again reflect the role of within and across sector allocation respectively.

4.3 The Value-added Model

I next present the value-added model, which comes straight from Hsieh and Klenow (2009) with minor modifications. The model ignores sectoral linkages and assumes all firms work with a value-added production. For the whole economy, there is still a single final product that is the GDP of this economy. The final output is produced from sectoral value-added with a Cobb-Douglas technology.

$$Y = \Phi \sum_{s=1}^{S} Y_s^{\widehat{\theta}_s}, \text{ with } \sum_{s=1}^{S} \widehat{\theta}_s = 1 \tag{4.19}$$

where Φ is a normalizing constant and I have used a hat to indicate a variable that's different between the two models. Profit maximization leads to $P_{Ys} Y_s = \widehat{\theta}_s Y$. Sectoral value-added is an aggregate over the varieties,

$$Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\widehat{\sigma}_s - 1}{\widehat{\sigma}_s}} \right)^{\frac{\widehat{\sigma}_s}{\widehat{\sigma}_s - 1}}, \tag{4.20}$$

where the elasticity of substitution $\widehat{\sigma}_s$ will be different from that in the output model. These differences in parameters resemble the differences in utility functions under different views of sectors as discussed in Herrendorf et al. (2013). For example, we can view the aggregate production as a utility function without affecting the results of the model. Profit maximization again leads to the inverse demand, $P_{Ysi} = P_{Ys} Y_s^{\frac{1}{\widehat{\sigma}_s}} Y_{si}^{-\frac{1}{\widehat{\sigma}_s}}$ with the price index of sectoral value-added given by $P_{Ys} = \left(P_{Ysi}^{1-\widehat{\sigma}_s} \right)^{\frac{1}{1-\widehat{\sigma}_s}}$.

Production of the varieties uses only primary inputs,

$$Y_{si} = \widehat{A}_{si} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}, \tag{4.21}$$

where I have set the value-added shares of capital and labor to be identical in the two models. The firms face idiosyncratic distortions in the factor markets. They take prices and production technology as given and maximize,

$$\widehat{\pi}_{si} = P_{Ysi}Y_{si} - (1+\tau_{Ksi})RK_{si} - (1+\tau_{Lsi})WL,$$

where I have assumed that the size of the distortions are the same in the two models. Inferred distortions using data however might be different in these two models, depending on whether there are distortions in the intermediate input markets. I will discuss the inference later. Factor market clearing conditions are defined the same as above.

Let's proceed to define the marginal revenue products as above,

$$\begin{aligned} \widehat{MRPK}_{si} &\equiv \alpha_s \frac{P_{Ysi}Y_{si}}{K_{si}} = (1+\tau_{Ksi})R, && (4.22)\\ \widehat{MRPL}_{si} &\equiv (1-\alpha_s)\frac{P_{Ysi}Y_{si}}{L_{si}} = (1+\tau_{Lsi})W. && (4.23)\end{aligned}$$

Similarly, value productivity and physical productivity are defined as,

$$\begin{aligned} \widehat{TFPR}_{si} &\equiv \frac{P_{Ysi}Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}, && (4.24)\\ \widehat{TFPQ}_{si} &\equiv \frac{Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}} = \widehat{A}_{si}, && (4.25)\end{aligned}$$

with price given by $P_{Ysi} = \frac{\widehat{TFPR}_{si}}{\widehat{A}_{si}}$.[19] Sectoral value productivity is defined as

$$\widehat{TFPR}_s \equiv \frac{P_{Ys}Y_s}{K_s^{\alpha_s}L_s^{1-\alpha_s}}. \tag{4.26}$$

Making use of these definitions leads to the sectoral value-added production function

$$Y_s = \widehat{TFP}_s \cdot K_s^{\alpha_s}L_s^{1-\alpha_s}, \text{ and } \widehat{TFP}_s = \left[\sum_{i=1}^{N_s}\left(\widehat{A}_{si}\frac{\widehat{TFPR}_s}{\widehat{TFPR}_{si}}\right)^{\widehat{\sigma}_s-1}\right]^{\frac{1}{\widehat{\sigma}_s-1}}. \tag{4.27}$$

[19]We again have $\widehat{TFPR}_{si} \propto \widehat{MRPK}_{si}^{\alpha_s}\widehat{MRPL}_{si}^{1-\alpha_s}$.

Given the sectoral production function, the allocation of capital and labor across sectors is given by,

$$\begin{aligned}\frac{K_s}{K} &= \frac{\frac{1}{\widehat{\mathcal{T}}_{Ks}}\alpha_s\widehat{\theta}_s}{\sum_{q=1}^{S}\frac{1}{\widehat{\mathcal{T}}_{Kq}}\alpha_q\widehat{\theta}_q} \equiv \widehat{\beta}_{Ks}, \\ \frac{L_s}{L} &= \frac{\frac{1}{\widehat{\mathcal{T}}_{Ls}}(1-\alpha_s)\widehat{\theta}_s}{\sum_{q=1}^{S}\frac{1}{\widehat{\mathcal{T}}_{Lq}}(1-\alpha_q)\widehat{\theta}_q} \equiv \widehat{\beta}_{Ls},\end{aligned}$$

where $\widehat{\mathcal{T}}_{Ks} = \frac{\widehat{MRPK_s}}{R}$ and $\widehat{\mathcal{T}}_{Ls} = \frac{\widehat{MRPL_s}}{W}$ are defined as as above.[20] Combining the factor demand with the production function, the aggregate production function also has a Cobb-Douglas form, which is introduced in the following proposition.

Proposition 6 *The value-added model also admits an aggregate production function given as*

$$Y = \widehat{A}K^{\widehat{\alpha}}L^{1-\widehat{\alpha}}, \tag{4.28}$$

where total factor productivity is given by

$$A = \Phi\prod_{s=1}^{S}\left(\widehat{TFP}_s\right)^{\widehat{\theta}_s}\prod_{s=1}^{S}\left[\left(\widehat{\beta}_{Ks}\right)^{\alpha_s}\left(\widehat{\beta}_{Ls}\right)^{1-\alpha_s}\right]^{\widehat{\theta}_s}, \tag{4.29}$$

and the capital share is $\widehat{\alpha} = \sum_{s=1}^{S}\widehat{\theta}_s\alpha_s$.[21]

Again TFP is decomposed into a within sector and a between sector component. The exponents on sectoral TFP and the allocation terms are the sectoral shares in final ouptut instead of the Domar weights in the output model. This difference reflects the fact that the

[20]The marginal revenue products are defined as

$$\begin{aligned}\widehat{MRPK}_s &\equiv \alpha_s\frac{P_{Ys}Y_s}{K_s} = \frac{1}{\sum_{i=1}^{N_s}\frac{1}{\widehat{MRPK}_{si}}\frac{P_{Ysi}Y_{si}}{P_{Ys}Y_s}}, \\ \widehat{MRPL}_s &\equiv (1-\alpha_s)\frac{P_{Ys}Y_s}{L_s} = \frac{1}{\sum_{i=1}^{N_s}\frac{1}{\widehat{MRPL}_{si}}\frac{P_{Ysi}Y_{si}}{P_{Ys}Y_s}}.\end{aligned}$$

From the definition, we again have $\widehat{TFPR}_s \propto \widehat{MRPK}_s^{\alpha_s}\widehat{MRPL}_s^{1-\alpha_s}$.

[21]The share of labor is $\sum_{s=1}^{S}\widehat{\theta}_s(1-\alpha_s)$. It is obvious that they sum to 1.

shocks to productivity or resource allocation are magnified through the production network. We can similarly decompose the efficiency loss from resource misallocation into two terms, which is given by

$$\left(\frac{Y^E}{Y}\right)^v = \prod_{s=1}^{S}\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s} \prod_{s=1}^{S}\left[\left(\frac{\widehat{\beta}_{Ks}^E}{\widehat{\beta}_{Ks}}\right)^{\alpha_s}\left(\frac{\widehat{\beta}_{Ls}^E}{\widehat{\beta}_{Ls}}\right)^{1-\alpha_s}\right]^{\widehat{\theta}_s} \tag{4.30}$$

where

$$\widehat{TFP}_s^E = \left[\sum_{i=1}^{N_s} \widehat{A}_{si}^{\widehat{\sigma}_s-1}\right]^{\frac{1}{\widehat{\sigma}_s-1}} \tag{4.31}$$

and

$$\widehat{\beta}_{Ks}^E = \frac{\alpha_s\widehat{\theta}_s}{\sum_{q=1}^S \alpha_q\widehat{\theta}_q}, \text{ and } \widehat{\beta}_{Ls}^E = \frac{(1-\alpha_s)\widehat{\theta}_s}{\sum_{q=1}^S (1-\alpha_q)\widehat{\theta}_q} \tag{4.32}$$

Notice that Hsieh and Klenow (2009) focuses only on the within sector reallocation, efficiency loss measured by them corresponds to the the first part of measured efficiency loss here.

4.4 The Equivalence of The Two Models

This section presents the equivalence results. I first show that we can transform the output model to the value-added model with appropriate choice of parameters. I then show that efficiency loss measured in these two models are identical for the case with no distortions in the intermediate input markets.

4.4.1 The Proof

The proof builds the value-added model from firm value-added production and shows the two models have the same resource allocation. I proceed in three steps. I first show that given a value-added production function for the firms, there is an value-added production function for the sectors. I then show that it is also the case for the whole economy. Lastly, I prove that resource allocation in the constructed value-added model indeed conforms with the output model.

Step 1 I first show that sectoral output can be viewed as produced from sectoral value-added and intermediate inputs, with sectoral value-added defined as a CES aggregator of firm value-added while taking firm value-added production function as given. I will show in step 3 that the two models have the same allocation of primary factors, such that the firm value-added production function is properly defined.

Since the primary inputs and intermediate inputs are separated in output production, it is easy to define a firm value-added production. Let's rewrite the production function for the variety as,

$$Q_{si} = Y_{si}^{\eta_s} \left(\prod_{q=1}^{S} M_{qsi}^{\lambda_{qs}} \right)^{1-\eta_s}, \tag{4.33}$$

with $Y_{si} = A_{si}^{\frac{1}{\eta_s}} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}$ as the firm's value-added production function. For the moment, I assume the firms take value-added as given and maximize profit by changing intermediate input use only.[22] Combining the transformed production function with the demand for intermediate input of the firms, the allocation of intermediate inputs is given by

$$\frac{M_{qsi}}{M_{qs}} = \frac{Y_{si}^{\frac{\eta_s(\sigma_s-1)}{1+\eta_s(\sigma_s-1)}}}{\sum_{j=1}^{N_s} Y_{sj}^{\frac{\eta_s(\sigma_s-1)}{1+\eta_s(\sigma_s-1)}}}$$

Plugging the intermediate allocation rule back into the production functions, we can rewrite the sectoral production function as produced from sectoral value-added and intermediate inputs,

$$Q_s = Y_s^{\eta_s} \left(\prod_{q=1}^{S} M_{qs}^{\lambda_{qs}} \right)^{1-\eta_s}, \text{ with } Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\eta_s(\sigma_s-1)}{1+\eta_s(\sigma_s-1)}} \right)^{\frac{1+\eta_s(\sigma_s-1)}{\eta_s(\sigma_s-1)}}. \tag{4.34}$$

If we set $\widehat{A}_{si} = A_{si}^{\frac{1}{\eta_s}}$ and $\widehat{\sigma}_s = 1 + \eta_s(\sigma_s - 1)$, we just have shown that there exists a sectoral value-added production function which takes firm value-added as inputs.

Step 2 I next proceed to show that given the sectoral production function, final output is given by a Cobb-Douglas aggregator of sectoral value-added.

[22]The result of this maximization problem is the restricted profit function.

Plugging the demand for intermediate input $M_{qs} = \frac{(1-\eta_s)\lambda_{qs}P_sQ_s}{P_q}$ into the sectoral output production function and taking logs, we have,

$$\log Q_s = \eta_s Q_s + \sum_{q=1}^{S}(1-\eta_s)\lambda_{qs}\log(1-\eta_s)\lambda_{qs} + (1-\eta_s)(\log P_s + \log Q_s) - \sum_{q=1}^{S}(1-\eta_s)\lambda_{qs}\log P_q.$$

Solving for $\log Q_s$ and stacking it into a vector, we have,

$$\bar{q} = \bar{y} + [\mathrm{diag}((\mathbf{1} - \bar{\eta}) \oslash \bar{\eta}) - \mathrm{diag}(\bar{\eta}^{\circ -1}\mathbf{1})B']\bar{p},$$

where $\bar{y}$ is the vector of $\log Y_s$, $\bar{p}$ the vector of $\log P_s$, and $\circ$ the is the operator for the element-wise Hadamard product, $\circ - 1$ for Hadamard inverse, $\oslash$ for Hadamard division, and diag transforms a vector into a diagonal matrix with its elements on the diagonal. This equation gives the sectoral ouptu as a function of sectoral value-added and sectoral prices. Next notice that the Domar weights links final output to sectoral price and output. Taking logs of the Domar weight and stacking it into a vector, it reads

$$\bar{p} + \bar{q} = \bar{v} + \mathbf{1}\log Y,$$

where $\bar{v}$ is the vector of $\log v_s$, and $\mathbf{1}$ is a vector of ones. Combining these two equations, we can express $\bar{p}$ as a function of $\log Y$,

$$\bar{p} = (I - B')^{-1}\mathrm{diag}(\bar{\eta})[\bar{v} + \mathbf{1}\log Y - \bar{y}]$$

Since I have used the final good as numeraire such that $\prod_{s=1}^{S}\left(\frac{P_s}{\theta_s}\right)^{\theta} = 1$. Taking logs and combining it with the above equation, we have

$$\theta'(I - B')^{-1}\mathrm{diag}(\bar{\eta})[\bar{v} + \mathbf{1}\log Y - \bar{y}] - \sum_{s=1}^{S}\theta_s \log\theta_s = 0,$$

which can be used to solve for final output as,

$$\log Y = (\bar{\eta} \circ \bar{v})\bar{y} - \bar{v}'\text{diag}\bar{\eta}\bar{v} - \sum_{s=1}^{S} \theta_s \log \theta_s \tag{4.35}$$

where I have made use of the fact that $(\bar{\eta} \circ \bar{v})\mathbf{1} = 1$, which also implies that the production function has constant returns to scale.[23] Let $\widehat{\theta}_s = \eta_s v_s$, I just proved that final output can be viewed as produced from sectoral value-added with a Cobb-Douglas technology.[24]

Step 3 Finally, I show that the allocation of primary production factors are identical in the two models given the same distortions in primary factor markets, which justifies the use of firm value-added in the first step.

I start with the allocation of resources within a sector. For the output model, remember that $P_{si} = \frac{TFPR_{si}}{A_{si}}$ and $P_s Q_s^{\frac{1}{\sigma_s}} Q_{si}^{-\frac{1}{\sigma_s}} = P_{si}$. Combining these two equations, we have

$$P_{si} Q_{si} \propto \left(\frac{TFPR_{si}}{A_{si}} \right)^{1-\sigma_s},$$

where I have made use of the definition of value productivity. Combining this equation with firm's factor demand, we have within sector factor allocation in the output model given as

$$\begin{aligned} \frac{K_{si}}{K_s} &= \frac{\frac{1}{1+\tau_{Ksi}} \left(\frac{TFPR_{si}}{A_{si}} \right)^{1-\sigma_s}}{\sum_{i=1}^{N_s} \frac{1}{1+\tau_{Ksi}} \left(\frac{TFPR_{si}}{A_{si}} \right)^{1-\sigma_s}}, \\ \frac{L_{si}}{L_s} &= \frac{\frac{1}{1+\tau_{Lsi}} \left(\frac{TFPR_{si}}{A_{si}} \right)^{1-\sigma_s}}{\sum_{i=1}^{N_s} \frac{1}{1+\tau_{Lsi}} \left(\frac{TFPR_{si}}{A_{si}} \right)^{1-\sigma_s}}. \end{aligned}$$

---

[23] From the definition of Domar weights, we have $\overline{PY} = (\bar{\eta} \circ \bar{v})Y$, where $\overline{PY}$ is the vector of nominal value-added $P_{Ys}Y_s$. Since final output in this economy equals the sum of sectoral value-added, it is obvious that $(\bar{\eta} \circ \bar{v})\mathbf{1} = 1$.

[24] The normalizing constant Φ in the value-added model is given by $\exp\left(- \left(\bar{v}'\text{diag}\bar{\eta}\bar{v} + \sum_{s=1}^{S} \theta_s \log \theta_s \right) \right)$.

Similarly, for the value-added model, the same shares are given by

$$
\begin{aligned}
\frac{K_{si}}{K_s} &= \frac{\frac{1}{1+\tau_{Ksi}}\left(\frac{\widehat{TFPR}_{si}}{\widehat{A}_{si}}\right)^{1-\widehat{\sigma}_s}}{\sum_{i=1}^{N_s}\frac{1}{1+\tau_{Ksi}}\left(\frac{\widehat{TFPR}_{si}}{\widehat{A}_{si}}\right)^{1-\widehat{\sigma}_s}}, \\
\frac{L_{si}}{L_s} &= \frac{\frac{1}{1+\tau_{Lsi}}\left(\frac{\widehat{TFPR}_{si}}{\widehat{A}_{si}}\right)^{1-\widehat{\sigma}_s}}{\sum_{i=1}^{N_s}\frac{1}{1+\tau_{Lsi}}\left(\frac{\widehat{TFPR}_{si}}{\widehat{A}_{si}}\right)^{1-\widehat{\sigma}_s}}.
\end{aligned}
$$

From the definition of value productivity, we have $TFPR_{si} = \widehat{TFPR}_{si}^{\frac{1}{\eta_s}}$.[25] Combining this information and the definition of $\widehat{A}_{si}$ in step 1. It is easy to see that within sector allocation are the same in the two models. I next move to the between sector allocation. From the discussion above, it is obvious that $\mathcal{T}_{Ks} = \widehat{\mathcal{T}}_{Ks}$ and $\mathcal{T}_{Ls} = \widehat{\mathcal{T}}_{Ls}$. Also remember that $\widehat{\theta}_s = \eta_s v_s$ from step 2. Across sector allocation again can be shown to be identical from comparing the results in Section 4.2 and Section 4.3. This completes the proof. I summarize the results in the following proposition.

Proposition 7 *Let $\widehat{A}_{si} = A_{si}^{\frac{1}{\eta_s}}$, $\widehat{\sigma}_s = 1 + \eta_s(\sigma_s - 1)$, and $\widehat{\theta}_s = \eta_s v_s$, the two models are equivalent.*

The relationship between σ_s and $\widehat{\sigma}_s$ shows why it is necessary to specify the elasticity to be sector specific. As long as the value-added shares are not identical across sectors, it is not possible for both models to have identical elasticity of substitution across sectors. It also says the elasticity of substitution should be different in the two models, a point emphasized in Herrendorf et al. (2013). For practical considerations, since the elasticity is more likely to be estimated for goods instead of value-added, the value-added studies are more likely to have chosen wrong parameters. For example, Hsieh and Klenow (2009) cite Broda and Weinstein (2006) and Hendel and Nevo (2006) for their choice of elasticity while both studies estimate the parameter for goods instead of value-added. Hsieh and Klenow (2009) use 3 as a conservative measure of elasticity. If we take 3 as the elasticity for output, that for

[25] Remember $TFPR_{si} \propto (1+\tau_{Ksi})^{\alpha_s \eta_s}(1+\tau_{Lsi})^{(1-\alpha_s)\eta_s}$ and $\widehat{TFPR}_{si} \propto (1+\tau_{Ksi})^{\alpha_s}(1+\tau_{Lsi})^{(1-\alpha_s)}$.

value-added will be 2 if we assume $\eta_s = 0.5$ and only 1.56 if $\eta_s = 0.28$.[26] This will lower measured efficiency loss as the larger the elasticity, the higher the cost of misallocation. The next section shows that assigning incorrect values to the elasticity is the major reason why the value-added model might have overstated the cost of resource misallocation.

The scaling up of productivity in the value-added model has been discussed by Bruno (1978). It is just another manifestation of the magnification effect of sectoral linkages. To see this clearly, consider the following one sector example. Final output is produced by labor and intermediate input $Q = AL^{\alpha}X^{1-\alpha}$, which is used for consumption and intermediate inputs in production, $Q = C+X$. GDP in this economy then is simply given by $C = \alpha(1-\alpha)^{\frac{1-\alpha}{\alpha}} A^{\frac{1}{\alpha}} X$. If we derive the aggregate production from the output production function, we realize that the effect of a shock to productivity A is magnified through the input-output linkages. The exponent $\frac{1}{\alpha}$ is the Domar weight for this economy. In the full model,the Domar weights can be rewritten as $v_s = \frac{\widehat{\theta}_s}{\eta_s}$ such that it is clear to see the resemblance to the simple one-sector model. On the other hand, if we are only given data on the value-added and primary inputs, we will just scale up the size of the productivity shock. The value-added model hence is not wrong, it just implicitly includes the magnification effect explicitly presented in the output model.

4.4.2 Implementation

Even though the two models are theoretically equivalent in representing the underlying data, measured efficiency loss could still be different if the wedges and productivity are incorrectly estimated in one of the models. This subsection shows that this is not the case. Efficiency loss measured in the two models are identical. For this exercise, assume we observe data on output ($P_{si}Q_{si}$), capital (K_{si}), labor ($Ł_{si}$), and intermediate input use ($\sum_{q=1}^{S} P_q M_{qsi}$) at

[26] 0.5 is the rough value used in Jones (2011) and 0.28 is the average across industries found in the Chinese manufacturing data used below. Note the share of intermediate inputs is higher in manufacturing, see for example Donovan (2014).

the firm level.[27] Nominal value-added is defined as $P_{Ysi}Y_{si} = P_{si}Q_{si} - \sum_{q=1}^{S} P_q M_{qsi}$.[28] I also assume that we have assigned correct values to all the parameters.[29]

To refresh memory, I copy the efficiency loss measures from above. For the output model, it is given by

$$\left(\frac{Y^E}{Y}\right)^o = \prod_{s=1}^{S}\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s}\prod_{s=1}^{S}\left[\left(\frac{\beta_{Ks}^E}{\beta_{Ks}}\right)^{\alpha_s\eta_s}\left(\frac{\beta_{Ls}^E}{\beta_{Ls}}\right)^{(1-\alpha_s)\eta_s}\right]^{v_s}$$

For the value-added model, it is given by

$$\left(\frac{Y^E}{Y}\right)^o = \prod_{s=1}^{S}\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}\prod_{s=1}^{S}\left[\left(\frac{\widehat{\beta}_{Ks}^E}{\widehat{\beta}_{Ks}}\right)^{\alpha_s}\left(\frac{\widehat{\beta}_{Ls}^E}{\widehat{\beta}_{Ls}}\right)^{1-\alpha_s}\right]^{\widehat{\theta}_s}$$

Also remember the definition of sectoral TFP and the cross sector allocation rule, these measures are operative if we have measures of firm productivity and wedges in factor markets. I next discuss the identification of these objects in data.

For the output model, the wedges can be measured as

$$\begin{aligned}1+\tau_{Ksi} &= \frac{\alpha_s\eta_s P_{si}Q_{si}}{RK_{si}},\\ 1+\tau_{Lsi} &= \frac{(1-\alpha_s)\eta_s P_{si}Q_{si}}{WL_{si}}\end{aligned}$$

[27]Note that I have assumed that we only observe total intermediate input in nominal values, which is what most firm-level database can provide. Given our assumption of technology and the fact that there is no distortions in the intermediate markets, we can attribute total expenditure on intermediate inputs to different inputs, which can be further used to estimate the real quantity of intermediate inputs from different sectors given appropriate price indexes for sectoral output. I will later show that this however is unnecessary for the exercise here.

[28]Note that nominal value-added is measured net of wedges. See the discussion above.

[29]This might not be an easy job in particular for the elasticity of substitution. Previous studies have mostly made the simplifying assumption that the parameter is the same for all sectors, for example see Hsieh and Klenow (2009) and Oberfield (2013). See also the discussion in last subsection.

For the value-added model, they are given by

$$\begin{aligned}1+\tau_{Ksi} &= \frac{\alpha_s P_{Ysi}Y_{si}}{RK_{si}}, \\ 1+\tau_{Lsi} &= \frac{(1-\alpha_s)P_{Ysi}Y_{si}}{WL_{si}}\end{aligned}$$

Since we have $P_{Ysi}Y_{si} = \eta_s P_{si}Q_{si}$, it is obvious the wedges are identical in these two models if factor prices are assumed to be the same. We however need not to be worried about factor prices as they cancel out when the wedges are plugged into the efficiency loss measures.

Identifying firm's physical productivity is a bit more involved as there are no measures of output prices or value-added prices at the firm level. Following Hsieh and Klenow (2009), I calculate physical productivity in the output model as

$$A_{si} = P_s^{\frac{\sigma_s}{\sigma_s-1}} Q_s^{\frac{1}{\sigma_s-1}} \prod_{q=1}^{S}\left(\frac{P_q}{\lambda_{qs}}\right)^{\lambda_{qs}(1-\eta_s)} \frac{(P_{si}Q_{si})^{\frac{\sigma_s}{\sigma_s-1}}}{\left(K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}\right)^{\eta_s}\left(\sum_{q=1}^{S}P_qM_{qsi}\right)^{1-\eta_s}},$$

where I have used the allocation rule for intermediate inputs and the pricing function for the variety. The term before the fraction can be viewed as a constant as it will be canceled out in the process. Similarly, physical productivity measured in the value-added model is given by,

$$\widehat{A}_{si} = P_{Ys}^{\frac{\widehat{\sigma}_s}{\widehat{\sigma}_s-1}} Y_s^{\frac{1}{\widehat{\sigma}_s-1}} \frac{(P_{Ysi}Y_{si})^{\frac{\widehat{\sigma}_s}{\widehat{\sigma}_s-1}}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}.$$

To make a comparison between these two measures, notice that $\sum_{q=1}^{S}P_qM_{qsi} = (1-\eta_s)P_{si}Q_{si}$ and $P_{si}Q_{si} = \frac{1}{\eta_s}P_{Ysi}Y_{si}$. Plugging these into the output measure, we have

$$A_{si} = P_s^{\frac{\sigma_s}{\sigma_s-1}} Q_s^{\frac{1}{\sigma_s-1}} \prod_{q=1}^{S}\left(\frac{P_q}{\lambda_{qs}}\right)^{\lambda_{qs}(1-\eta_s)} \frac{\eta_s^{\frac{\eta_s\widehat{\sigma}_s}{1-\widehat{\sigma}_s}}}{(1-\eta_s)^{1-\eta_s}}\left(\frac{(P_{Ysi}Y_{si})^{\frac{\widehat{\sigma}_s}{\widehat{\sigma}_s-1}}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}\right)^{\eta_s}$$

Note that what matters for computing efficiency loss is the variation in physical productivity but not the absolute level, we can thus set both of the constants to be 1, then it's obvious $\widehat{A}_{si} = A_{si}^{\frac{1}{\eta_s}}$.[30] Given our discussion in step 3 of the equivalence proof, the relationship

[30]If instead we set the constants to be ϕ and $\widehat{\phi}$, then we have $\widehat{A}_{si} = \widehat{\phi}\phi^{-\frac{1}{\eta_s}}A_{si}^{\frac{1}{\eta_s}}$. Measured efficiency loss

in measured wedges and firm productivity between the two models implies that measured efficiency loss will be the same in the two models.

This result confirms that the value-added model is a correct representation of the underlying data generated from an output model. As discussed above, value-added productivity observed in data implicitly includes the magnification effect of intersectoral linkages, which is made explicit in the output model. The output model has magnification effect while measured productivity is less dispersed than in the value-added model. The two forces cancel out each other, leaving the efficiency measure unaffected.

Identical measured efficiency loss also relies on the fact that marginal revenue products and firm productivity can be correctly estimated in the value-added model. My definition of value-added assumes the same intermediate prices for all firms, we thus can also view them as a real value-added measure for the firms even though product prices are allowed to change across firms.[31] As shown by Bruno (1978), marginal revnue products and productivity for the real value-added production function can be correctly measured if the price of intermediate inputs are constant. This condition is satisfied when there are no distortions in the intermediate input markets. This assumption however is highly unrealistic. I proceed to the discuss the more realistic case with intermediate input market distortions in the next section.

4.5 Adding Intermediate Input Market Distortions

Having proved the equivalence results under the case with no distortions in intermediate input markets, I next add these distortions and discuss the measurement of misallocation in this more realistic case. There does not exist a correct value-added model with the newly

will still be the same in this case because the constants cancel out.

[31] This point becomes clearer when distortions in intermediate input use are introduced. Firms will then face different effective prices in the intermediate input markets while value-added is defined using the same market price.

added distortions.[32] I however continue to use the value-added model laid out in Section 4.2 as it is widely used in the literature.[33] The purpose is to show how the value-added model will bias the results. I first discuss the measurement of efficiency loss in the output model in next section. I then discuss how the implementation is affected in Section 4.5.2.

4.5.1 Measuring Misallocation in the Output Model

I first describe how adding distortion to the intermediate inputs markets changes the measurement of efficiency loss in the output model. The theoretical results in the value-added model is not affected by this but the identification of wedges and firm productivity will be affected, which I discuss in the next subsection.

Remember the firms try to maximize profit given distortions in factor markets from Section 4.2,

$$\pi_{si} = P_{si}Q_{si} - (1+\tau_{Ksi})RK_{si} - (1+\tau_{Lsi})WL_{si} - (1+\tau_{Msi})\sum_{q=1}^{S} P_q M_{qsi},$$

I now drop Assumption 2 and allow τ_{Msi} to be non-zero. Notice that I have assumed the same wedge for all the intermediate inputs, as the data does not allow us to distinguish distortions for inputs sourced from different sectors. This changes marginal revenue product for intermediate inputs

$$MRPM_{qsi} \equiv \lambda_{qs}(1-\eta_s)\frac{P_{si}Q_{si}}{M_{qsi}} = (1+\tau_{Msi})P_q. \tag{4.36}$$

Let $\mathcal{T}_{Ms} = MRPM_{qs}/P_q$ be the distortion on sectoral intermediate input use. Market clearing for sectoral output now reads as,

$$P_sQ_s = P_sC_s + \sum_{q=1}^{S}\frac{1}{\mathcal{T}_{Mq}}\lambda_{sq}(1-\eta_q)P_qQ_q,$$

[32] This is because there is no value-added production function at the sector level. That is, step 1 of the proof in Section 4.4.1 breaks down.

[33] See the review of literature in the Section 4.1.

which is now also distorted given the presence of $\mathcal{T}_{Ms}$. Let $\tilde{B} = B \circ \text{diag}(\mathcal{T}_M^{\circ -1})$ with $\mathcal{T}_M$ denoting the vector of $\mathcal{T}_{Ms}$. The vector of Domar weights $\tilde{V}$ can then be solved from,

$$\tilde{V} = \theta + \tilde{B}\tilde{V}.$$

Domar weights are now affected by market distortions, $\tilde{V} = (I - \tilde{B})^{-1}\theta$ with its element given by $\tilde{v}_s$. Let the sq^{th} element of $\tilde{B}$ be $\tilde{b}_{sq}$. We can then write the intermediate input demand of sector s for sector q's good as $M_{qs} = \tilde{b}_{qs}\frac{v_s}{v_q}Q_q$. This leads to the vector of log sectoral output

$$\bar{q} = (I - B')^{-1}(\bar{a} + \tilde{\omega}_q + \delta_K \log K + \delta_L \log L),$$

where the vector $\tilde{\omega}_q$ again summarizes the information on sectoral factor allocation.[34] This further leads to an aggregate production function summarized as follows.

Proposition 8 *The economy with intermediate input market distortions has a aggregate production function given as*

$$Y = \tilde{A}K^{\alpha}L^{1-\alpha}, \tag{4.37}$$

with total factor productivity is given by

$$\tilde{A} = \tilde{\gamma}\prod_{s=1}^{S}(TFP_s)^{\tilde{v}_s}\prod_{s=1}^{S}\left[(\beta_{Ks})^{\alpha_s\eta_s\tilde{v}_s}(\beta_{Ls})^{(1-\alpha_s)\eta_s\tilde{v}_s}\right], \tag{4.38}$$

where $\tilde{\gamma} = \prod_{s=1}^{S}\left(\frac{\theta_s}{v_s}\right)^{\theta_s}\prod_{s=1}^{S}\left[\prod_{q=1}^{S}\left(\tilde{b}_{qs}\frac{v_s}{v_q}\right)^{1-\eta_s}\lambda_{qs}\right]^{\tilde{v}_s}$ *and* α *is defined as above.*

Compared to Proposition 5, $\tilde{\gamma}$ is no longer a constant. The fact that the Domar weights are no longer constant however makes the measurement of efficiency loss much more involved. Notice that the aggregate output elasticity of capital and labor stays intact, such that the effect of misallocation still manifests itself as a shock to TFP. To make it more comparable to the existing literature, I go on to analyze the case with no misallocation across sectors by making the following assumption,

[34] Its element is given by $\alpha_s\eta_s\log\beta_{Ks} + (1-\alpha_s)\eta_s\log\beta_{Ls} + (1-\eta_s)\sum_{q=1}^{S}\lambda_{qs}\log\left(\tilde{b}_{qs}\frac{v_s}{v_q}\right)$.

Assumption 3 *$\mathcal{T}_{Js} = 1, \forall J \in \{K, L, M\}$ and $s \in \{1, 2, ...S\}$.*

This assumption directs attention to within sector allocation as in Hsieh and Klenow (2009). It also allows us to identify the factor shares in the production function using sectoral aggregates. Otherwise they can not be distinguished from the market distortions. With this assumption, efficient loss from resource misallocation in the output model reduces to

$$\left(\frac{Y^E}{Y}\right)^o = \prod_{s=1}^{S}\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s}, \tag{4.39}$$

while that in the value-added model is

$$\left(\frac{Y^E}{Y}\right)^v = \prod_{s=1}^{S}\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}. \tag{4.40}$$

With Assumption 3, the efficiency loss measures are identical in the two models if we have correct measures of revenue and physical productivity. This is surprising as the value-added model does not actually exist. Measured efficiency loss however will be different in the two models because marginal revenue products and firm productivity will be incorrectly inferred in the value-added model while the output model still leads to correct measures. The next subsection makes this point clear and discusses how the value-added model might measure efficiency loss in data wrong.

4.5.2 Implementation

This subsection performs the same exercise in Section 4.4.2 with hypothetical data on firm production and value-added computed as subtracting intermediate inputs from output. Importantly, intermediate inputs are assumed to be measured with the same market prices in the data, such that constructed value-added can be thought of a real value-added measure as firms actually face different prices in the intermediate input markets due to the presence of wedges.

Wedges and productivity in the output model will be correctly measured following the discussion in the previous section, but it is no longer the case for the value added model. Using the first-order condition for intermediate input use, value-added is now solved as

$$P_{Ysi}Y_{si} = \left(1 - \frac{1-\eta_s}{1+\tau_{Msi}}\right) P_{si}Q_{si}, \tag{4.41}$$

which when plugged into the first order condition for the value-added model leads to

$$\begin{aligned} 1+\tau_{Ksi} &= \frac{\alpha_s \left(1 - \frac{1-\eta_s}{1+\tau_{Msi}}\right) P_{si}Q_{si}}{RK_{si}}, \\ 1+\tau_{Lsi} &= \frac{(1-\alpha_s)\left(1 - \frac{1-\eta_s}{1+\tau_{Msi}}\right) P_{si}Q_{si}}{WL_{si}} \end{aligned}$$

It is obvious the wedges are mismeasured by a factor of $\frac{1}{\eta_s} - \frac{1-\eta_s}{\eta_s(1+\tau_{Msi})}$ as the share of intermediate inputs in total output is distorted. Following the same steps in section 4.4.2, firm productivity in the output model can be rewritten as

$$A_{si} = P_s^{\frac{\sigma_s}{\sigma_s-1}} Q_s^{\frac{1}{\sigma_s-1}} \prod_{q=1}^{S} \left(\frac{P_q}{\lambda_{qs}}\right)^{\lambda_{qs}(1-\eta_s)} \frac{\left(1 - \frac{1-\eta_s}{1+\tau_{Msi}}\right)^{\frac{\eta_s\widehat{\sigma}_s}{1-\widehat{\sigma}_s}} (1+\tau_{Msi})^{1-\eta_s}}{(1-\eta_s)^{1-\eta_s}} \left(\frac{(P_{Ysi}Y_{si})^{\frac{\widehat{\sigma}_s}{\widehat{\sigma}_s-1}}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}}\right)^{\eta_s}$$

Assume the constant terms are 1 as before, we now have firm productivity in the value-added model given by

$$\widehat{A}_{si} = \left(1 - \frac{1-\eta_s}{1+\tau_{Msi}}\right)^{\frac{\widehat{\sigma}_s}{\widehat{\sigma}_s-1}} (1+\tau_{Msi})^{\frac{\eta_s-1}{\eta_s}} A_{si}^{\frac{1}{\eta_s}}$$

Measured physical productivity hence is also distorted by the presence of distortions in the intermediate inputs markets.

I next proceed to express sectoral productivity in terms of inferred firm productivity and wedges. For the output model, it is given by

$$TFP_s = \left(\sum_{i=1}^{N_s} \left(\frac{A_{si}}{\left[(1+\tau_{Ksi})^{\alpha_s}(1+\tau_{Lsi})^{1-\alpha_s}\right]^{\eta_s}(1+\tau_{Msi})^{1-\eta_s}}\right)^{\sigma_s-1}\right)^{\frac{1}{\sigma_s-1}}, \tag{4.42}$$

while efficiency sectoral TFP in the output model is only determined by firm productivity as described above, $TFP_s^E = \left(\sum_{i=1}^{N_s} A_{si}^{\sigma_s-1}\right)^{\frac{1}{\sigma_s-1}}$. This equation breaks down revenue productivity and allows us to see the effect of different distortions more clearly. In the case of the value-added model, sectoral TFP is given by

$$\widehat{TFP}_s = \left(\sum_{i=1}^{N_s}\left(\frac{A_{si}\left(1-\frac{1-\eta_s}{1+\tau_{Msi}}\right)^{\frac{1}{\sigma_s-1}}}{\left[(1+\tau_{Ksi})^{\alpha_s}(1+\tau_{Lsi})^{1-\alpha_s}\right]^{\eta_s}(1+\tau_{Msi})^{1-\eta_s}}\right)^{\sigma_s-1}\right)^{\frac{1}{\eta_s(\sigma_s-1)}}. \tag{4.43}$$

This result comes from mismeasured wedges in capital and labor use and physical productivity, and also reflects the fact that distortions in the intermediate input markets are not taken into account in computing revenue productivity of firms. Given mismeasured firm physical productivity, efficient sectoral TFP will also be mismeasured as

$$\widehat{TFP}_s^E = \left(\sum_{i=1}^{N_s}\left(A_{si}\left(1-\frac{1-\eta_s}{1+\tau_{Msi}}\right)^{\frac{1+\eta_s(\sigma_s-1)}{\sigma_s-1}}(1+\tau_{Msi})^{\eta_s-1}\right)^{\sigma_s-1}\right)^{\frac{1}{\eta_s(\sigma_s-1)}} \tag{4.44}$$

Measurement errors at the firm level thus build up, such that both actual and efficient sectoral TFP will deviate from the true measures. These findings confirm the results of Bruno (1978): the real value-added production function will leads to incorrect measures of marginal products and productivity. Bruno (1984) has used this idea to explain the productivity slowdown and Basu and Fernald (1995) find it useful in understanding why productivity spillover across industries exists in value-added data but not in output data. I show here that the mismeasurement builds up to the sector level and further leads to mismeasured efficiency loss in the value-added model.

From the results above, it is not clear in which direction the value-added model will bias measured efficiency loss. To see this more clearly, I next approximate the measures by assuming A_{si}, $1+\tau_{Ysi}$,[35] and $(1+\tau_{Msi})$ jointly follow a log-normal distribution, with the log of the latter two having zero mean. I further assume $\eta_s = 0.5$ such that $1-\frac{1-\eta_s}{1+\tau_{Msi}}$ can be

[35] $1+\tau_{Ysi}$ is defined as $(1+\tau_{Ksi})^{\alpha_s}(1+\tau_{Lsi})^{1-\alpha_s}$, which summarizes the distortion in primary input use.

approximated by $\frac{1}{2}(1+\tau_{Msi})$.[36] Efficiency loss for a single sector in these two models can be approximated by[37]

$$\log\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} = \frac{v_s(1-\sigma_s)}{2}\left(\frac{\sigma_y^2}{4}+\frac{\sigma_m^2}{4}-\sigma_{ay}-\sigma_{am}+\frac{\sigma_{ym}}{2}\right),$$

where σ_x^2 is the variance of variable x and σ_{xz} is the covariance between variables x and z. Efficiency loss in the value-added model is approximated by

$$\log\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s} = \frac{v_s(1-\sigma_s)}{2}\left[\frac{\sigma_y^2}{4}+\left(\frac{1}{4}-\frac{1}{\sigma_s-1}\right)\sigma_m^2-\sigma_{ay}-\sigma_{am}+\left(\frac{1}{2}-\frac{1}{\sigma_s-1}\right)\sigma_{ym}\right]$$

Hence efficiency loss measured in these two model is differed by

$$\log\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} - \log\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s} = -\frac{v_s}{2}(\sigma_m^2+\sigma_{ym}) \tag{4.45}$$

From this comparison, efficiency loss in the value-added model can be biased from that in the output model in either direction. The sign and the size of the bias depends on how the wedges for intermediate inputs are distributed. I next go on to explore the bias empirically using Chinese data.

4.6 An Application to Chinese Data

In this section I take the models to data and evaluate the bias in the value-added efficiency loss measure quantitatively. The data I use is the Chinese Annual Survey of Industrial Production in 2005. Since the survey is only for the the industrial sector, it is not sufficient for measuring aggregate efficiency loss. Instead I will look at efficiency loss measures at

[36] The assumed value-added share in output is not so different from aggregate data. See Jones (2011) for the evidence. The share might be different across sectors, as shown in the Chinese data in the next section.

[37] Notice the measure is raised by the Domar weights. This expression actually gives the contribution of sector s to aggregate efficiency loss. It is the same case for the value-added model.

4-digit industry level using both models.[38] Despite that there are complicated input-output linkages between the sectors, the discussion above shows that what we need is only the sector's Domar weight if we ignore cross sector misallocation. The purpose in this section is to show how the use of the value-added model might distorts measured efficiency loss at the industry level, as distortions in the intermediate input use enters our measure of firm value-added.

The Survey of Industrial Production covers all non-state firms with more than 5 million yuan in revenue plus all state-owned firms in the industrial sector, which includes mining, manufacturing, utilities, and construction. I use the information on the firm's industry (at the four-digit level), wage payments,[39] employment, output, value-added, capital stock, and intermediate inputs. Capital stock is defined as the book value of fixed capital net of depreciation. Labor compensation in the data is systematically under-reported such that calculated industrial labor share is much smaller than those in the national accounts.[40] To correct for the under-reporting, Hsieh and Klenow (2009) raise the wage payments of all firms to make aggregate labor share calculated from the survey consistent with national accounts data. This however will create a labor share larger than 1 for some industries. To make the industry labor shares bounded by 1, I raise wage payments by a factor of 2.75 for firms in the industries with raw labor share less than 0.3 and by a factor of 1.26 for all other firms. This assumes that the industries with smaller labor share surfer more from the under-reporting of wage payments. Note that I only study misallocation within a sector, this adjustment should not bias the results too much. The procedure produces an aggregate labor share of 0.5. Finally, to allow for differences in worker human capital across firms, I use wage bills as my measurement of labor input instead of employment.

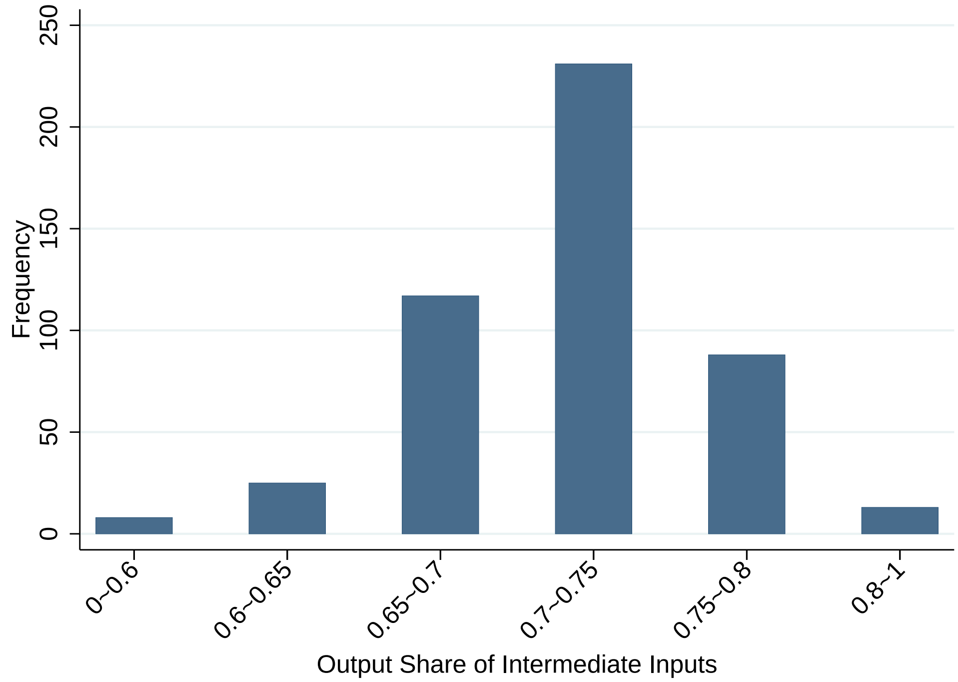
To measure efficiency loss, we need to assign values to parameters α_s, η_s, σ_s, and v_s.

[38]That is, $\log\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s}$ and $\log\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}$ as in last section.

[39]Wage payments include wages, retirement and unemployment insurance, health insurance, housing benefits, and employee supplementary benefits.

[40]It can be clearly seen in the data. For example, 59% of all firms report zero unemployment insurance payment, 38% of all firms report zero health insurance payment, while both of which are mandatory by the law.

Figure 4.1: Distribution of the Output Share of Intermediate Inputs

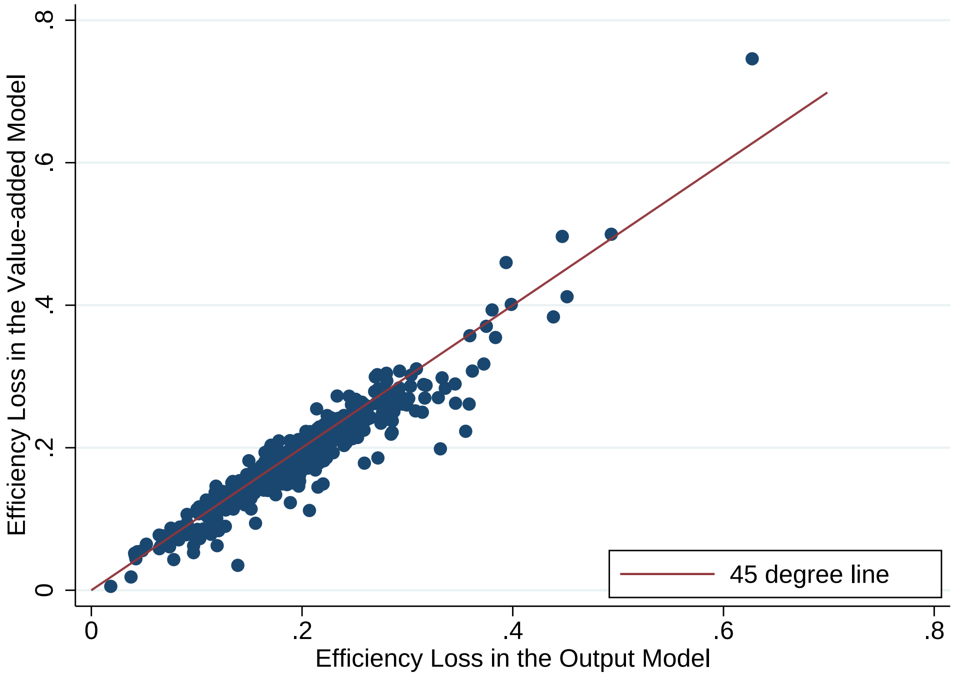


Note: The distribution of $1 - \eta_s$ at the industry level is plotted.

The output share parameters come from dividing factor payments by total output in an industry, thanks to the assumption that there is no misallocation across sectors. Capital income is derived by subtracting from total output the expenditure on intermediate input and the adjusted payment to labor. To relieve the burden of assign different elasticity of substitution for so many industries, I assume σ_s is 3 for all industries. This implies that $\widehat{\sigma}_s = 1 + \eta_s(\sigma_s - 1)$ is different across industries for the value-added model if η_s is different across industries. Figure 4.1 plots the distribution of the output share of intermediate inputs $(1 - \eta_s)$ for the 482 industries in the data. Most industries have relative high shares and there is some variation across industries. The average across industries is 0.72. Finally, as v_s affects both measures to the same degree, I simply normalize it to 1.

Figure 4.2 compares measured efficiency loss in the two models. The measures are not so different in the two models as most points center around the 45 degree line. The difference between the two measures is less than 5 percentage points for over 95% of the industries. There is no evidence that one model produces results substantially lower than the other, a

Figure 4.2: Efficiency Loss in Different Models



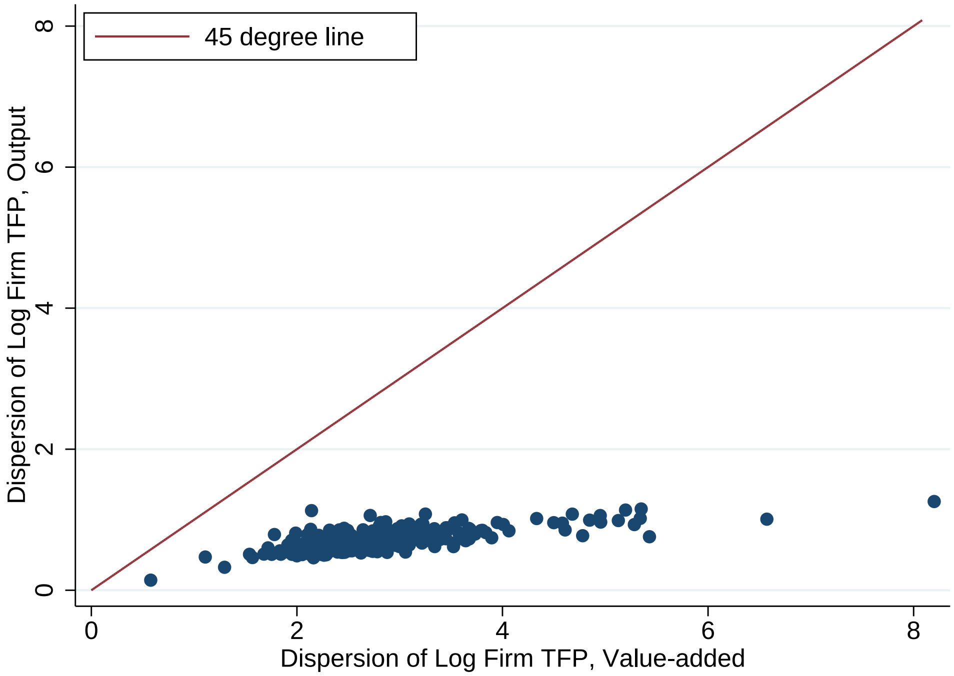
Note: What's plotted is $\log\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s}$ and $\log\left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}$, both computed by the author.

point my theoretical analysis emphasizes above. Despite the magnification effect of production networks, the reason that the efficiency loss in the output model is not substantially larger than that in the value-added model is lower dispersion of firm TFP measured in the output model. This can be clearly seen from Figure 4.3, where the standard deviation of log firm TFP in the output model is plotted against that in the value-added model. It is also interesting to point out that the dispersion measured in the output model are quite similar across industries, while it shows much more variation in the value-added model.[41] This suggests that the differences in the dispersion of firm TFP across industries also come from differences in intermediate input share across industries.

Efficiency loss measured in the value-added model will only deviate from that in the output model if there are distortions in intermediate input use. The fact that the two

[41]The coefficient of variation across industries is 0.25 in the value-added model and 0.18 in the output model.

Figure 4.3: Dispersion of Firm Productivity in Different Models



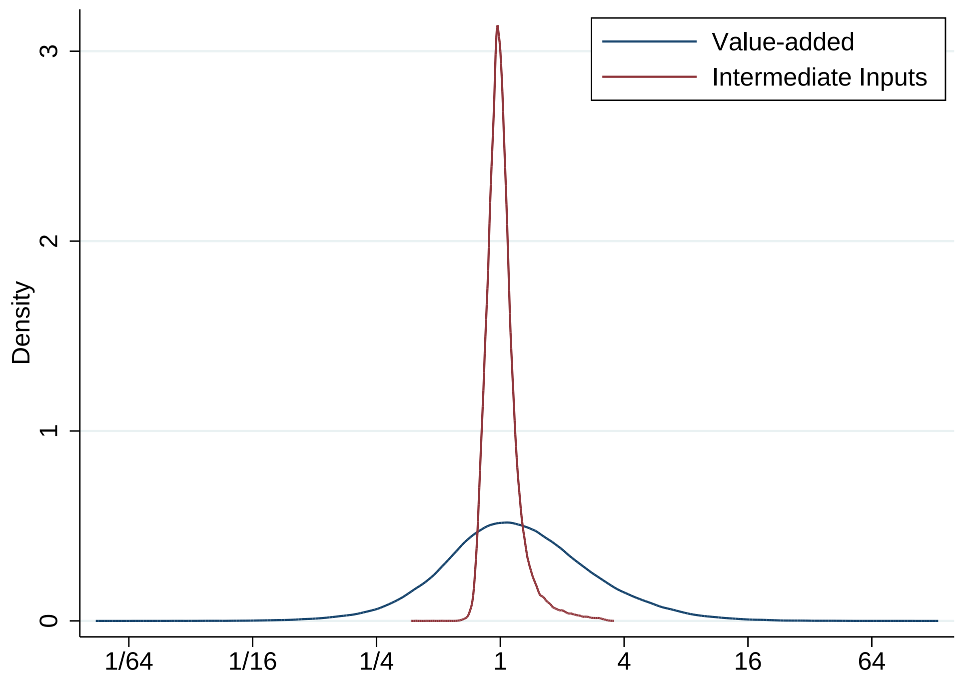
Note: Here dispersion is measured by standard deviation of log productivity.

measures stay close to each other suggest that distortions in intermediate input might be mild. Figure 4.4 confirms this in the data, where I plot the density of the marginal revenue product (in logs) of primary inputs and Intermediate inputs relative to their industry average respectively.[42] While the marginal revenue product of primary inputs has a very dispersed distribution with a standard deviation of 0.85, that of intermediate inputs centers around the industry average with a standard deviation of only 0.19. This confirms the finding of Krishna and Tang (2018).

The elasticity of substitution in the value-added model is assumed to be different across industries. If we instead use a constant elasticity by imposing the intermediate input share for the whole industrial sector on $\widehat{\sigma}_s = 1 + \eta_s(\sigma_s - 1)$. The resulted efficiency loss will not be so different. On the other hand, if we assume the elasticity is 3 as in Hsieh and Klenow (2009), the value-added model will produce much larger efficiency loss than the output model. This is because the true elasticity for the value-added model is only 1.56 on

[42]They are given by $MRPK_{si}^{\alpha_s} MRPL_{si}^{1-\alpha_s}$ and $\prod_{q=1}^{S} MRPM_{qsi}^{\lambda_{qs}}$.

Figure 4.4: Dispersion of Marginal Revenue Products



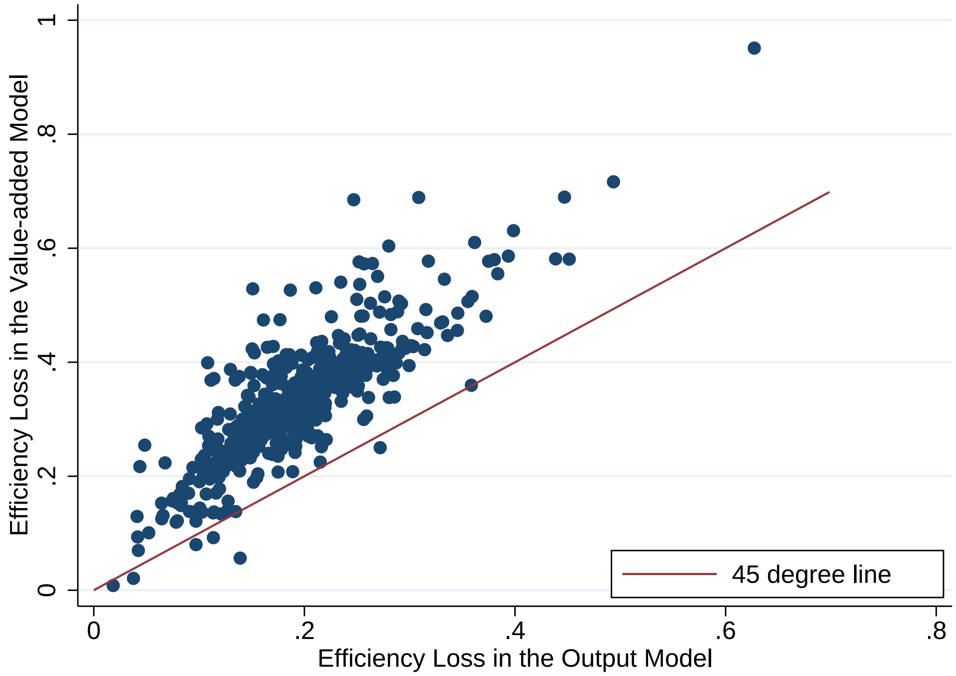
Note: What's plotted here is the logs of $MRPK_{si}^{\alpha_s} MRPL_{si}^{1-\alpha_s}$ and $\prod_{q=1}^{S} MRPM_{qsi}^{\lambda_{qs}}$, both relative to the industry average.

average given the transformation from the output model to the value-added model. This result is plotted in Figure 4.5. Measured efficiency loss is larger in the value-added model for over 95% of the industries and the difference between the two models is 14 percentage points on average. This finding suggests that the value-added model actually overstates efficiency loss due to incorrectly assigned elasticity of substitution. Production networks will not help in accounting for international income differences by raising measured efficiency loss from resource misallocation but tends to lower the estimated cost of misallocation once the differences between the output model and the value-added model are fully appreciated.

Finally, I examine how the approximation at the end of Section 4.5 compares to data. Under the log normality assumption, the approximation of efficiency loss says that

$$\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} > \left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s} \iff \sigma_m^2 + \sigma_{ym} < 0.$$

Figure 4.5: Efficiency Loss in Different Models with $\sigma_s = \widehat{\sigma}_s = 3$



Note: Here I have assigned 3 to $\widehat{\sigma}_s$ instead of using the theoretically correct value.

Table 4.1 presents the frequency counts of these incidents. If the approximation is accurate, we should observe larger numbers in the off diagonal cells. This is indeed the case in data. If $\sigma_m^2 + \sigma_{ym} < 0$, the probability of observing $\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} > \left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}$ is 0.70. On the other hand, if $\sigma_m^2 + \sigma_{ym} > 0$, the probability of observing $\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} < \left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}$ is 0.61. On average, the approximation will have the correct prediction on the relative size of efficiency loss measured in the two models at a probability a little over 2/3. A t-test that the prediction is random is rejected with a p-value less than 0.0001.

Table 4.1: Frequency Counts

	$\sigma_m^2 + \sigma_{ym} > 0$	$\sigma_m^2 + \sigma_{ym} < 0$
$\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} > \left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}$	28	286
$\left(\frac{TFP_s^E}{TFP_s}\right)^{v_s} < \left(\frac{\widehat{TFP}_s^E}{\widehat{TFP}_s}\right)^{\widehat{\theta}_s}$	44	124

Data source: Authors' own calculation.

4.7 Summary

This chapter shows that the widely-used value-added framework of Hsieh and Klenow (2009) does not necessarily understate the cost of resource misallocation. If there are no distortions in the intermediate input markets, a model with a production network *a la* Long and Plosser (1983) can be transformed into the value-added model and the two models produce the same efficiency loss. If there are distortions in the intermediate input markets, the value-added model produces biased results but the bias can go in either direction. The literature using the value-added model, if anything, might have overstated the efficiency loss due to the use of higher elasticity of substitution that is only suitable for the output model.

The findings in this chapter support the idea of Herrendorf et al. (2013) that we can either view sectors as categories of final expenditure or value-added. Both models can be correct representations of the same underlying data and they are connected through the input-output linkages. The findings also suggest that the models can be easily mis-specified if we don't distinguish between the two perspectives explicitly, a point also emphasized in Herrendorf et al. (2013). Given the recent surge in the study of production networks and multi-sector models in general, we should pay more attention to the distinction between final expenditure and value-added and how value-added models might be mis-specified.

Bibliography

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The Network Origins of Aggregate Fluctuations. *Econometrica 80*(5), 1977–2016.

Acemoglu, D. and V. Guerrieri (2008). Capital Deepening and Nonbalanced Economic Growth. *Journal of Political Economy 116*(3), 467–498.

Adamopoulos, T. (2011). Transportation Costs, Agricultural Productivity,and Cross-Country Income Differences. *International Economic Review 52*(2), 489–521.

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia (2015, September). Misallocation, Selection, and Productivity: A Quantitative Analysis with Panel Data from China. Unpublished.

Adamopoulos, T. and D. Restuccia (2014). The Size Distribution of Farms and International Productivity Differences. *American Economic Review 104*(6), 1667–97.

Adamopoulos, T. and D. Restuccia (2015, May). Land Reform and Productivity: A Quantitative Analysis with Micro Data. University of Toronto, Department of Economics Working Paper 540.

Alfaro, L., A. Charlton, and F. Kanczuk (2009). Plant Size Distribution and Cross-Country Income Differences. *NBER International Seminar on Macroeconomics 5*(1), 243–272.

Alvarez, J. A. (2017, January). The Agricultural Wage Gap: Evidence from Brazilian Microdata.

Alvarez-Cuadrado, F., N. V. Long, and M. Poschke (2013, May). Capital Labor Substitution, Structural Change, and Growth.

Alvarez-Cuadrado, F., N. V. Long, and M. Poschke (2014, January). Capital-labor Substitution, Structural Change and the Labor Income Share.

Asker, J., A. Collard-Wexler, and J. D. Loecker (2014). Dynamic input and resource (mis)allocation. *Journal of Political Economy 122*(5), 1013–1063.

Banerjee, A. V. and E. Duflo (2005). Growth Theory through the Lens of Development Economics. In *Handbook of Economic Growth*, Volume 1, Chapter 7, pp. 473–552. Elsevier.

Banerjee, A. V. and B. Moll (2010). Why does Misallocation Persist? *American Economic Journal: Macroeconomics 2*(1), 189–206.

Baqaee, D. R. (2016). Cascading Failures in Production Networks. Unpublished.

Baqaee, D. R. and E. Farhi (2017a). Productivity and Misallocation in General Equilibrium. Technical report, National Bureau of Economic Research.

Baqaee, D. R. and E. Farhi (2017b). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. Technical report, National Bureau of Economic Research.

Bartelme, D. and Y. Gorodnichenko (2015). Linkages and Economic Development. Technical report, National Bureau of Economic Research.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review 103*(1), 305–334.

Basu, S. and J. G. Fernald (1995). Are Apparent Productive Spillovers a Figment of Specification Error? *Journal of Monetary Economics 36*(1), 165–188.

Battistin, E., R. Blundell, and A. Lewbel (2009). Why Is Consumption More Log Normal than Income? Gibrat's Law Revisited. *Journal of Politcal Economy 117*(6), 1140–1154.

Baumol, W. J. (1967). Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. *American Economic Review 57*(3), 415–426.

Bigio, S. and J. LaO (2016). Financial Frictions in Production Networks. Technical report, National Bureau of Economic Research.

Black, D., N. Kolesnikova, and L. Taylor (2009). Earnings Functions When Wages and Prices Vary by Location. *Journal of Labor Economics 27*(1), 21–47.

Boppart, T. (2014). Structural Change and the Kaldor Facts in a Growth Model with Relative Price Effects and Non-Gorman Preferences. *Econometrica 82*(6), 2167–2196.

Brandt, L. and C. A. Holz (2006). Spatial Price Differences in China: Estimates and Implications. *Economic Development and Cultural Change 55*(1), 43–86.

Brandt, L., T. Tombe, and X. Zhu (2013). Factor market distortions across time, space, and sectors in china. *Review of Economic Dynamics 16*, 39–58.

Broda, C. and D. E. Weinstein (2006). Globalization and the Gains from Variety. *The Quarterly journal of economics 121*(2), 541–585.

Bruno, M. (1978). Duality, Intermediate Inputs and Value-added. In *Contributions to Economic Analysis*, Volume 2, pp. 3–16. Elsevier.

Bruno, M. (1984). Raw Materials, Profits, and the Productivity Slowdown. *The Quarterly Journal of Economics 99*(1), 1–29.

Buera, F. J. and Y. Shin (2013). Financial Frictions and the Persistence of History: A Quantitative Exploration. *Journal of Political Economy 121*(2), 221–272.

Bunkers, E. W. and W. W. Cochrane (1957). On the Income Elasticity of Food Services. *Review of Economics and Statistics 39*(2), 211–217.

Butzer, R., Y. Mundlak, and D. F. Larson (2010, November). Measures of Fixed Capital in Agriculture. World Bank Policy Research Working Paper 5472.

Caliendo, L., F. Parro, and A. Tsyvinski (2017). Distortions and the Structure of the World Economy. Technical report, National Bureau of Economic Research.

Carletto, C., S. Savastano, and A. Zezza (2013). Fact or Artifact: The Impact of Measurement Errors on the Farm Size-Productivity Relationship. *Journal of Development Economics 103*, 254–261.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1, Part A of *Handbook of Economic Growth*, pp. 679 – 741. Elsevier.

Caselli, F. and W. J. Coleman (2001). The U.S. Structural Transformation and Regional Convergence: A Reinterpretation. *Journal of Political Economy 109*(3), 584–616.

Caselli, F. and W. J. Coleman (2006). The World Technology Frontier. *American Economic Review 96*(3), 499–522.

Chen, C. (2016). Untitled Land, Occupational Choice, and Agricultural Productivity. *American Economic Journal: Macroeconomics*. forthcoming.

Chen, K. and A. Irarrazabal (2015). The Role of Allocative Efficiency in a Decade of Recovery. *Review of Economic Dynamics 18*(3), 523–550.

Chirinko, R. S. (2008). σ: The Long and Short of It. *Journal of Macroeconomics 30*, 671–686.

Ciccone, A. (2002). Input Chains and Industrialization. *The Review of Economic Studies 69*(3), 565–587.

Clarida, R. H. (1994). Cointegration, Aggregate Consumption, and the Demand for Imports: A Structural Econometric Investigation. *American Economic Review 84*(1), 298–308.

Comin, D., D. Lashkari, and M. Mestieri (2015, September). Structural Change with Long-run Income and Price Effects. NBER Working Paper No. 21595.

Costa, D. L. (2001). Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves. *Journal of Politcal Economy 109*(6), 1288–1310.

David, J. and V. Venkateswaran (2017). Capital Misallocation: Frictions or Distortions? Unpublished.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, Misallocation, and Aggregate Productivity. *The Quarterly Journal of Economics 131*(2), 943–1005.

De Magalhães, L. and R. Stantaeulàlia-Llopis (2015). The Consumption, Income, and Wealth of the Poorest: Cross-Sectional Facts of Rural and Urban Sub-Saharan Africa. Univeristy of Bristol, Department of Economics.

Deaton, A. (1974). The Analysis of Consumer Demand in the United Kingdom, 1900-1970. *Econometrica 42*(2), 341–367.

Deaton, A. (1988). Quality, Quantity, and Spatial Variation of Price. *American Economic Review 78*(3), 418–430.

Deaton, A. and O. Dupriez (2011, July). Spatial Price Differences within Large Countries. unpublished.

Dollar, D., T. Kleineberg, and A. Kraay (2016). Growth still is Good for the Poor. *European Economic Review* (81), 68–83.

Dollar, D. and A. Kraay (2002). Growth is Good for the Poor. *Journal of Economic Growth 7*, 195–225.

Dollar, D. and S.-J. Wei (2007). Das (Wasted) Kapital: Firm Ownership and Investment Efficiency in China.

Donovan, K. (2014, May). Agricultural Risk, Intermedaite Inputs, and Cross-Country Productivity Differences.

Fafchamps, M. and F. Shilpi (2003). The Spatial Division of Labour in Nepal. *Journal of Development Studies 39*(6), 23–66.

Fafchamps, M. and F. Shilpi (2005, April). Cities and Specialization: Evidence from South Asia. *Economic Journal 115*, 477–504.

Fajgelbaum, P. D. and A. K. Khandelwal (2016). Measuring the Unequal Gains from Trade. *Quarterly Journal of Economics 131*(3), 1113–1180.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. *American Economic Review 105*(10), 3150–3182.

Fuglie, K. O. (2010). Total Factor Productivity in the Global Agricultural Economy: Evidence from FAO Data. In J. M. Alston, B. A. Babcock, and P. G. Pardey (Eds.), *The Shifting Patterns of Agricultural Production and Productivity*, Chapter 4. The Midwest Agribusiness Trade Research and Information Center.

Gandhi, A., S. Navarro, and D. Rivers (2013, September). On the Identification of Production Functions; How Heterogenous is Productivity? Unpublished.

Gandhi, A., S. Navarro, and D. Rivers (2017, May). How Heterogeneous is Productivity? A Comparison of Gross Output and Value Added? Unpublished.

Gollin, D. (2002). Getting income shares right. *Journal of Political Economy 110*(2), pp. 458–474.

Gollin, D., M. Kirchberger, and D. Lagakos (2017, May). In Search of a Spatial Equilibrium in the Developing World. unpublished.

Gollin, D., D. Lagakos, and M. E. Waugh (2014). The Agricultural Productivity Gap. *Quarterly Journal of Economics 129*(2), 939–993.

Gollin, D., S. L. Parente, and R. Rogerson (2002). The Role of Agriculture in Development. *American Economic Review 92*(2), 160–164.

Gollin, D., S. L. Parente, and R. Rogerson (2004). Farm work, home work and international productivity differences. *Review of Economic Dynamics 7*(4), 827 – 850.

Gollin, D., S. L. Parente, and R. Rogerson (2007). The Food Problem and the Evolution of International Income Levels. *Journal of Monetary Economics 54*, 1230–1255.

Gollin, D. and R. Rogerson (2014). Productivity, Transport Costs and Subsistence Agriculture. *Journal of Development Economics 107*, 38–48.

Gopinath, G., Ş. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017). Capital Allocation and Productivity in South Europe. *The Quarterly Journal of Economics 132*(4), 1915–1967.

Gottlieb, C. and J. Grobovšek (2015, May). Communal Land and Agricultural Productivity.

Guner, N., G. Ventura, and Y. Xu (2008). Macroeconomic Implications of Size-dependent Policies. *Review of Economic Dynamics 11*(4), 721–744.

Hamilton, B. (2001). Using Engel's Law to Estimate CPI Bias. *American Economic Review 91*(3), 619–630.

Handbury, J. (2013). Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities. Kilts Booth Marketing series, Paper No. 1-054.

Harris, J. R. and M. P. Todaro (1970). Migration, Unemployment and Development: a Two-Sector Analysis. *The American Economic Review 60*(1), 126–142.

Heien, D. and C. R. Wesseils (1990). Demand Systems Estimation with Microdata: A Censored Regression Approach. *Journal of Business & Economic Statistics 8*(3), 365–371.

Hendel, I. and A. Nevo (2006). Measuring the Implications of Sales and Consumer Inventory Behavior. *Econometrica 74*(6), 1637–1673.

Herrendorf, B., C. Herrington, and A. Valentinyi (2015). Sectoral Technology and Structural Transformation. *American Economic Journal: Macroeconomics 7*(4), 104–133.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2013). Two Perspectives on Preferences and Structural Transformation. *American Economic Review 103*(7), 2752–2789.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2014). Growth and Structural Transformation. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 2, Chapter 6, pp. 855–941. Elsevier.

Herrendorf, B., J. A. Schmitz, JR, and A. Teixeira (2012). The Role of Transportation in U.S. Economic Development: 1840-1960. *International Economic Review 53*(3), 693–715.

Herrendorf, B. and T. Schoellman (2015). Why is Measured Productivity so Low in Agriculture? *Review of Economc Dynamics 18*, 1003–1022.

Hirschman, A. O. (1958). *The Strategy of Economic Development*. Yale University Press.

Horvath, M. (1998). Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks. *Review of Economic Dynamics 1*(4), 781–808.

Horvath, M. (2000). Sectoral Shocks and Aggregate Fluctuations. *Journal of Monetary Economics 45*(1), 69–106.

Houthakker, H. S. (1957). An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel's Law. *Econometrica 25*(4), 532–551.

Houthakker, H. S. (1960). Additive Preferences. *Econometrica 28*(2), 244–257.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics* (4), 1403–1448.

Hsieh, C.-T. and P. J. Klenow (2014). The Life Cycle of Plants in India and Mexico. *The Quarterly Journal of Economics 129*(3), 1035–1084.

Hulten, C. R. (1978). Growth Accounting with Intermediate Inputs. *The Review of Economic Studies 45*(3), 511–518.

Jacoby, H. G. (2000). Access to Marketsand the Benefits of Rural Roads. *Economic Journal 110*(465), 713–737.

Jacoby, H. G. and B. Minten (2009). On Measuring the Benefits of Lower Transport Costs. *Journal of Development Economics 89*, 28–38.

Jones, C. I. (2011). Intermediate Goods and Weak Links in the Theory of Economic Development. *American Economic Journal: Macroeconomics 3*.

Jones, C. I. (2013). Misallocation, Economic Growth, and Input-Output Economics. In *Advances in Economics and Econometrics, Tenth World Congress*, Volume 2. Cambridge University Press.

Kalemli-Özcan, S. and B. E. Sorensen (2012). Misallocation, Property Rights, and Access to Finance: Evidence from Within and Across Africa. Technical report, National Bureau of Economic Research.

Kislev, Y. and W. Peterson (1981). Induced Innovations and Farm Mechanization. *American Journal of Agricultural Economics 63*(3), 562–565.

Kislev, Y. and W. Peterson (1982). Prices, Technology, and Farm Size. *Journal of Political Economy 90*(3), 578–595.

Kongsamut, P., S. Rebelo, and D. Xie (2001). Beyong Balanced Growth. *Review of Economic Studies 68*(4), 869–882.

Krishna, P. and H. Tang (2018, April). Production Networks, Trade, and Misallocation. Unpublished.

Lagakos, D. and M. E. Waugh (2013). Selection, Agriculture, and Cross-Country Productivity Differences. *American Economic Review 103*(2), 948–80.

Lee, S. (2010). Ability Sortingand Consumer City. *Journal of Urban Economics 68*, 20–33.

León-Ledesma, M. A., P. McAdam, and A. Willman (2011, November). Aggregation, the Skill Premium, and the Two-level Production Function. European Central Bank Working Paper Series, No. 1400.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School 22*(2), 139–191.

Lianos, T. P. (1971). The Relative Share of Labor in United States Agriculture, 1949-1968. *American Journal of Agricultural Economics 53*(3), 411–422.

Liao, J. and W. Wang (2014, December). Financial Frictions and Agricultural Productivity Differences.

Liu, E. (2017). Industrial Policies and Economic Development. Unpublished.

Long, J. B. J. and C. I. Plosser (1983). Real Business Cycles. *Journal of Political Economy 91*(1), 39–69.

Manuelli, R. E. and A. Seshadri (2014). Human Capital and the Wealth of Nations. *American Economic Review 104*(9), 2736–2762.

Meyer, B. D. and J. X. Sullivan (2009, March). Five Decades of Consumption and Income Poverty. NBER Working Paper No. 14727.

Midrigan, V. and D. Y. Xu (2014). Finance and Misallocation: Evidence from Plant-level Data. *American economic review 104*(2), 422–58.

Moll, B. (2014). Productivity Losses from Financial Frictions: Can Self-financing undo Capital Misallocation? *American Economic Review 104*(10), 3186–3221.

Muhammad, A., J. L. Seale Jr., Birgit Meade, and A. Regmi (2011). International Evidence on Food Consumption Patterns: An Update Using 2005 International Comparison Program Data. Technical Bulletin Number 1929, United States Department of Agriculture.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Eco 46*(1), 69–85.

Munshi, K. and M. Rosenzweig (2016). Networks and Misallocation: Insurance, Migration,and the Rural-Urban Wage Gap. *American Economic Review 106*(1), 46–98.

Nakamura, E., J. Steinsson, and M. Liu (2016). Are Chinese Growth and Inflation Too Smooth? Evidence from Engel Curves. *American Economic Journal: Macroeconomics 8*(3), 113–144.

Nerlove, M. L. and E. Sadka (1991). Von Thünen's Model of the Dual Economy. *Journal of Economics 54*(2), 97–123.

Ngai, L. R. and B. Petrongolo (2017). Gender Gaps and the Rise of the Service Economy. *American Economic Journal: Macroeconomics 9*(4), 1–44.

Ngai, L. R. and C. A. Pissarides (2007). Structural Change in a Multisector Model of Growth. *American Economic Review 97*(1), 429–443.

Nigai, S. (2015). On Measuring the Welfare Gains from Trade under Consumer Heterogeneity. *Economic Journal 126*(593), 1193–1237.

Oberfield, E. (2013). Productivity and Misallocation during a Crisis: Evidence from the Chilean Crisis of 1982. *Review of Economic Dynamics 16*, 100–119.

Oberfield, E. and D. Raval (2014). Micro Data and Macro Technology. NBER Working Paper No. 20452.

Ogaki, M. (1992). Engel's Law and Cointegration. *Journal of Political Economy 100*(5), 1027–1046.

O'Mahony, M. and M. P. Timmer (2009, June). Output, Input and Productivity Measures at the Industry Level: the EU KLEMS Database. *The Economic Journal 119*, F374–F403.

Oostendorp, R. H. (2009). Globalization and the Gender Wage Gap. *The World Bank Economic Review 23*(1).

Pitt, M. M., M. R. Rosenzweig, and M. N. Hassan (2012). Human Capital Investment and the Gender of Labor in a Brawn-Based Economy. *American Economic Review 102*(7), 3531–3560.

Rao, P. (1993). Intercountry Comparisons of Agricultural Output and Productivity. Economic and social development paper, Food and Agriculture Organization of the United Nations (FAO).

Ravallion, M. and D. van de Walle (1991). Urban-rural Cost-of-Living Differentials in a Developing Economy. *Journal of Urban Economics 29*, 113–127.

Renkow, M., D. G. Hallstrom, and D. D. Karanja (2004). Rural Infrastructure, Transactions Costs and Market Participation in Kenya. *Journal of Development Economics 73*(1), 349–367.

Restuccia, D. and R. Rogerson (2008). Policy Distortions and Aggregate Productivity with Heterogenous Establishments. *Review of Economic Dynamics 11*, 707–720.

Restuccia, D. and R. Rogerson (2013). Misallocation and Productivity. *Review of Economic Dynamics 16*, 1–10.

Restuccia, D. and R. Stantaeulàlia-Llopis (2017). Land Misallocation and Productivity. NBER Working Paper 23128.

Restuccia, D., D. T. Yang, and X. Zhu (2008). Agriculture and Aggregate Productivity: A Quantitative Cross-country Analysis. *Journal of Monetary Economics 55*(2), 234 – 250.

Schultz, T. W. (1964). *Transforming Traditional Agriculture*. New Haven: Yale University Press.

Seale Jr., J., A. Regmi, and J. Bernstein (2003). International Evidence on Food Consumption Patterns. Technical Bulletin Number 1904, United States Department of Agriculture.

Shenoy, A. (2017). Market Failures and Misallocation. *Journal of Development Economics* (128), 65–80.

Święcki, T. (2014, September). Determinants of Structural Change.

Thirsk, W. (1974, 56). Factor Substitution in Colombian Agriculture. *American Journal of Agricultural Economics 56*(1), 73–84.

Tombe, T. (2015). The Missing Food Problem: Trade, Agriculture, and International Productivity Differences. *American Economic Journal: Macroeconomics 7*(3), 226–58.

Uy, T., K.-M. Yi, and J. Zhang (2013). Structural Change in an Open Economy. *Journal of Monetary Economics 60*(6), 667–682.

Valentinyi, A. and B. Herrendorf (2008). Measuring Factor Income Shares at the Sectoral Level. *Review of Economc Dynamics 11*, 820–835.

Vollrath, D. (2009). How Important are Dual Economy Effects for Aggregate Productivity? *Journal of Development Economics 88*, 325–334.

Vollrath, D. (2014). The Efficiency of Human Capital Allocations in Developing Countries. *Journal of Development Economics 108*, 106–118.

Welch, F. (1970). Education in Production. *Journal of Political Economy 78*(1), 1970.

Wingender, A. M. (2015). Skill complementarity and the dual economy. *European Economic Review 74*, 269 – 285.

World Bank (1994). *World Development Report: Infrastructure for Development*. Oxford University Press.

Xu, Y. (1999). Agricultural Productivity in China. *China Economic Review 10*, 108–121.

Yang, D. T. and X. Zhu (2013). Modernization of Agriculture and Long-Term Growth. *Journal of Monetary Economics 60*, 367–382.

Young, A. (2013). Inequality, the Urban-Rural Gap, and Migration. *Quarterly Journal of Economics*, 1727–1785.

Ziebarth, N. L. (2012). Misallocation and productivity during the great depression. Unpublished.

Ziebarth, N. L. (2013). Are China and India Backward? Evidence from the 19th Century U.S. Census of Manufactures. *Review of Economic Dynamics 16*, 86–99.

Zuleta, H. and A. T. Young (2007). Labor's Shares - Aggregate and Industry: Accounting for Both in a Model of Unbalanced Growth with Induced Innovation.