UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Resource Allocation and Efficiency in Developing Countries

Permalink
https://escholarship.org/uc/item/6458x094

Author
Chatterjee, Urmila

Publication Date
2011

Peer reviewed|Thesis/dissertation

Resource Allocation and Efficiency in Developing Countries

by

Urmila Chatterjee

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor in Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yuriy Gorodnichenko, Chair
Professor Charles I. Jones
Professor Steven Raphael

Fall 2011

Resource Allocation and Efficiency in Developing Countries

Abstract

by

Urmila Chatterjee

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Yuriy Gorodnichenko, Chair

This study fills some of the gaps in the literature on resource misallocation in Indian manufacturing. There are three key contributions in this study. The first is the proper measurement of misallocation in Indian manufacturing. This is achieved by incorporating data from both the formal and informal sectors and intermediate inputs in calculating misallocation. By using value added based measures of revenue productivity (TFPR) we find that there was greater misallocation in the formal sector compared to the informal sector in the period 1994-2005. We also find that resource misallcoation had increased in the formal sector during this period. Using a monopolistic competitive model with heterogeneity in firm productivity and including intermediate inputs in the production function we show that resource misallocation at the industry level can be amplified due to intermediate inputs. We show that the value added based measures of misallocation suffer from omitted variable bias and that the magnitude and sign of this bias is dependent on the covariance of the distortions to intermediate inputs and other factor inputs.

The second contribution is towards identifying the nature, magnitude and sources of misallocation. Analyzing distortions to factor inputs individually we find that the distortions to intermediate inputs to be the largest followed by the distortions to capital. We find that distortions to intermediate inputs are larger in the informal sector than in the formal sector. Using qualitative information on informal firms we find that misallocation is positively correlated with shortage of capital and poor access to intermediate inputs and negatively correlated with firm market share.

The third contribution is that we measure the magnitude of misallocation caused by a specific size dependent policy in India. A small scale firm in India can get up to 100% exemption from paying excise tax if its annual turnover is less than Rs. 10 million. Using data on formal manufacturing firms in India we show that such an exemption encourages firms to operate on a sub-optimal scale. We construct the counterfactual distribution of output without any tax distortion using maximum likelihood and non-linear least squares estimation methods. Using the empirical estimates of the scale parameter in the literature we find that there are output gains in the range of 13% to 230% under a revenue neutral flat tax regime. These gains are sensitive to the scale parameter of the production function and are explosive as the scale parameter approaches one.

This dissertation is dedicated to my grandmother, Veda Rao.

Contents

List of Figures

List of Tables

Acknowledgments

Chapter 1

Introduction

1.1 Motivation and Background

One of the most widely researched topics in economics has been the study of cross country income differences. Why are some countries richer than others and why do some grow faster than others? A review of the growth literature in the past few decades often points the finger towards factor accumulation, technology, productivity or geography for answers. Empirical evidence however shows that total factor productivity (TFP) accounts for bulk of the cross country per capita income differences[1]. TFP can be decomposed into two components. The first one is technology per se, that is how inputs are combined in production. The question of how good an economy is in producing new technology and ideas is perhaps more relevant for advanced countries. With increased openness and flow of ideas what seems more relevant for developing countries is the study of how technology is adopted. Or in other words, the efficiency with which inputs are used in production.

Some of the significant contribution to this literature is the work done by Parente and Prescott (2000) who look at effect of monopolistic structures on the use of technology. Howitt (2000) and Klenow and Rodriguez-Claire (2005) show how slow technology adoption can lead to large TFP differences. These papers are examples of within firm or within sector inefficiency. In recent years there has been a burgeoning interest in cross firm or cross sector inefficiency. The main argument that this strand of literature makes is that when firms or sectors face distortions which could be due to government policy, market structures or institutions they do not allocate resources optimally. This resource misallocation across firms or sectors will result in lower aggregate TFP. Examples of resource misallocation include public sector undertakings in China who have access to credit at subsidized interest rates or

---

[1]See Mankiw et al. (1992), Klenow and Rodriguez-Claire (1997), Hall and Jones (1999), Caselli (2005), Jones and Romer (2010)

Chapter 1

Introduction

1.1 Motivation and Background

One of the most widely researched topics in economics has been the study of cross country income differences. Why are some countries richer than others and why do some grow faster than others? A review of the growth literature in the past few decades often points the finger towards factor accumulation, technology, productivity or geography for answers. Empirical evidence however shows that total factor productivity (TFP) accounts for bulk of the cross country per capita income differences[1]. TFP can be decomposed into two components. The first one is technology per se, that is how inputs are combined in production. The question of how good an economy is in producing new technology and ideas is perhaps more relevant for advanced countries. With increased openness and flow of ideas what seems more relevant for developing countries is the study of how technology is adopted. Or in other words, the efficiency with which inputs are used in production.

Some of the significant contribution to this literature is the work done by Parente and Prescott (2000) who look at effect of monopolistic structures on the use of technology. Howitt (2000) and Klenow and Rodriguez-Claire (2005) show how slow technology adoption can lead to large TFP differences. These papers are examples of within firm or within sector inefficiency. In recent years there has been a burgeoning interest in cross firm or cross sector inefficiency. The main argument that this strand of literature makes is that when firms or sectors face distortions which could be due to government policy, market structures or institutions they do not allocate resources optimally. This resource misallocation across firms or sectors will result in lower aggregate TFP. Examples of resource misallocation include public sector undertakings in China who have access to credit at subsidized interest rates or

---

[1]See Mankiw et al. (1992), Klenow and Rodriguez-Claire (1997), Hall and Jones (1999), Caselli (2005), Jones and Romer (2010)

firms with more than 100 employees in India requiring government permission to fire workers. In the case of the former there would be more capital allocated by public sector firms than the amount prescribed by the efficient case, where public sector firms would pay the market interest rate just like the private firms. In the latter firms would use less labor than they would in the absence of distortionary government policy. Some of the key contributors to this line of work include Banerjee and Duflo (2005) who show that distortions in credit policy led to substantial gaps in the marginal product of capital and lower TFP in India. Restuccia and Rogerson (2008) use a model of heterogeneous production units to show that in the presence of idiosyncratic distortions firms face different prices which leads to misallocation of resources across firms within an industry. They calibrate their model to fit US data and show that firm level distortions in the form of taxes and subsidies can misallocate resources and result in 30-50% fall in TFP. Bartelsman et al. (2009) follow a similar approach using cross country evidence. Buera and Shin (2009) focus on financial frictions in the context of liberalization of capital accounts. Hsieh and Klenow (2009) use plant level manufacturing data from India and China and show that reallocation of resources within industries to the US efficiency levels can boost TFP in India by 40-60% and in China by 30-50%.

In chapter two, we extend the research undertaken in the above mentioned papers in measuring misallocation of resources. We make three innovations. The first innovation is the inclusion of the informal sector in measuring misallocation in Indian manufacturing. Previous work on misallocation in India by Hsieh and Klenow (2009) had incorporated data only on the formal manufacturing sector in Inda. There is an enormous presence of the informal sector in Indian manufacturing. There were 17 million informal firms in the manufacturing sector compared to around 0.13 million formal firms in 2004-05. Informal firms accounted for more than 82% of total employment in manufacturing. Given its sheer size any study of Indian manufacturing would be incomplete without incorporating the informal sector in it. The second innovation is the inclusion of intermediate inputs in measuring misallocation. Earlier studies on misallocation, including Hsieh and Klenow (2009), typically used value added measures and ignored the potential distortions to intermediate inputs. We develop a general equilibrium model of monopolistic competition where firms are heterogenous in productivity and face idiosyncratic distortions. We show that the inclusion of intermediate inputs is crucial for both accurately measuring the magnitude of misallocation and for identifying the sources of misallocation. The third innovation that we present in this chapter is the identification of the sources of distortion. We exploit qualitative information on informal firms in our dataset and identify some of the potential sources of misallocation.

In chapter three, we focus our attention on a particular source of resource misallocation. We quantify the extent of misallocation and output loss due to a specific size dependent policy of India. Small firms in India are exempt from paying excise tax if their output falls below a certain threshold level. Using graphical bunching evidence we show that such an exemption policy is distortionary as it encourages firms to operate at a sub-optimal scale in order to avoid paying taxes. We develop a model where firms are heterogeneous in productivity and

face decreasing returns to scale to describe this bunching behavior. In order to measure the potential output loss due to such a tax policy we construct a counterfactual distribution of undistorted level of output using maximum likelihood estimation and non-linear estimation techniques. We then calculate the welfare gains from moving to a flat tax regime where all firms pay a uniform tax and their output is unconstrained.

In the fourth and concluding chapter, we present a summary of our findings and directions for future work in the area of resource misallocation and efficiency.

Chapter 2

Misallocation in Indian Manufacturing: Role of the Informal Sector and Intermediate Goods

2.1 Introduction

This chapter looks at the misallocation of resources in Indian manufacturing using firm level data. It contributes to the growing interest in the literature on measuring and identifying misallocation in developing countries using firm level data and its effects on aggregate TFP. The paper that is most closely related to this study is the the seminal work of Hsieh and Klenow (2009). Hsieh and Klenow use plant level manufacturing data from India and China and show that reallocation of resources within industries to the US efficiency levels can boost TFP in India by 40-60% and in China by 30-50%. They show that when distortions are not symmetric across firms within an industry, the marginal revenue products are not equalized leading to inefficiency. Micro level misallocation will show up in aggregate productivity. The dispersion in marginal revenue products (TFPR) is then a useful measure of resource misallocation. Hsieh and Klenow try to relate misallocation to government reforms in India. They do not find much evidence to show that that reforms have reduced misallocation. Unfortunately their data set for India is from 1987 to 1994. Reforms are believed to have taken off in a big way in India after 1991. In fact for India, they report an increase in the gains from reallocation for the period 1987 to 1994, suggesting widening of the gap between actual and efficient income.

There are two key stand alone contributions that we present in this chapter. The first contribution is towards the proper measurement of misallocation in Indian manufacturing. We do this in two ways. First, we look at the misallocation for the entire manufacturing sector

in India by using data on both the formal and informal manufacturing sectors. Previous research by Hsieh and Klenow (2009) on resource misallocation in India had looked at only the formal manufacturing sector. There is an enormous presence of the informal sector in Indian manufacturing. There were 17 million informal firms in the manufacturing sector compared to around 0.13 million formal firms in 2004-05. Informal firms accounted for more than 82% of total employment in manufacturing. Table 2.1 provides more descriptive statistics on the formal and informal sectors in India. Given its sheer size, any study of Indian manufacturing would be incomplete without incorporating the informal sector in it. Using micro level data for the formal and informal firms for the year 2004-05 we find that the manufacturing sector in India as a whole could have realised TFP gains of nearly 111% by moving to an efficient allocation. Second, we incorporate intermediate goods in our calculation of resource misallocation. We develop a model where firms are heterogenous in productivity, face idiosyncratic distortions and use intermediate inputs in production. We show that the value added measure of distortion may suffer from omitted variable bias as it ignores both the potential distortions to intermediate inputs and the amplification of distortions across industries in the presence of intermediate inputs.

The second contribution is towards identifying the different sources of distortions. Recent work on resource misallocation has established that misallocation in developing countries is large and that the potential gains from removing these distortions is substantial. What however remains unsettled is the source and nature of these distortions. We show that decomposing the total factor revenue productivity (TFPR), which is a composite measure of misallocation, and looking at distortions to factor inputs individually is an important step towards identifying the nature and the source of distortions. Using qualitative data on the informal firms we find that the misallocation is positively correlated with the shortage of capital and weak access to intermediate inputs such as electricity and raw materials and negatively correlated with the market share of firms.

The rest of this chapter is organized as follows. Section 2.2 presents facts and trends in the formal and informal sectors. Section 2.3 presents the definition of informality and the different views on it in the literature. Section 2.4 describes a model of monopolistic competition with heterogeneous firm productivity and idiosyncratic distortions. Section 2.5 describes the data for the formal and informal sectors. Section 2.6 describes the identification strategy for distortions in the data. Section 2.7 presents the empirical results using the value added and gross output approaches to measure productivity. Section 2.8 contrasts the results of the value added and gross output approaches by analyzing the contribution of distortions to capital, output and intermediate inputs individually. Section 2.9 looks at the potential sources of distortion using informal sector data. Section 2.10 presents our conclusions.

Figure 2.1: Output in Formal and Informal Sectors 2004-05
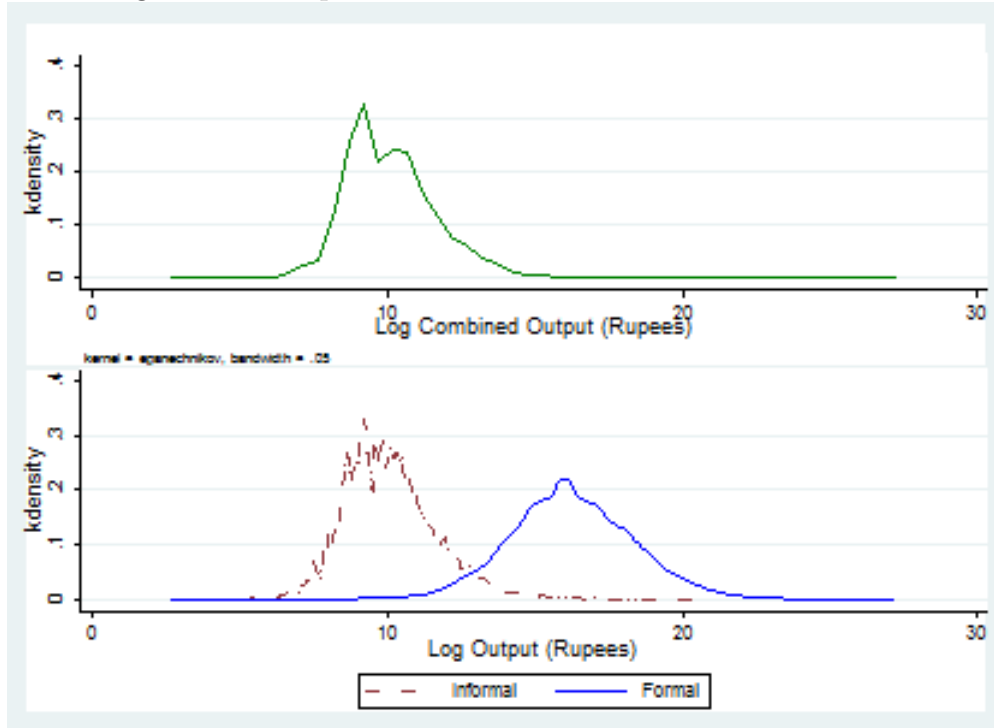
2.2 Formal and Informal Sectors: Facts and Trends

We begin this section by presenting some facts about the formal and informal sectors in India for the period 1994-95 to 2005-06. The informal sector has a huge presence in India. Figures 2.1, 2.2 and 2.3 plot the distribution of log output, log capital and log intermediate inputs for the entire manufacturing sector in India. The density plots indicate that the informal sector firms constitute a substantial portion of the entire manufacturing sector in India. The formal firms however dominate the right tail of the distributions. Table 2.1 provides a quick summary of the key trends in the formal and informal sectors in Indian manufacturing for the period 1994-95 to 2005-06. Here are the key findings:

1. According to the latest available estimates there were about 17 million informal firms compared to 0.13 million formal firms in the Indian manufacturing sector in the year 2005-06. The period 1995-2006 saw a 39% increase in informal firms compared to a 22% increase in formal firms. Also most of the increase happened in the first half of the period under study (we report the numbers for 1994-95 and 2005-06 only in Table 2.1)

2. The most important contribution of the informal sector is in job creation. The informal sector employed roughly 3.5 times more workers than the formal sector in 1994-95. This

Figure 2.2: Capital in Formal and Informal Sectors 2004-05
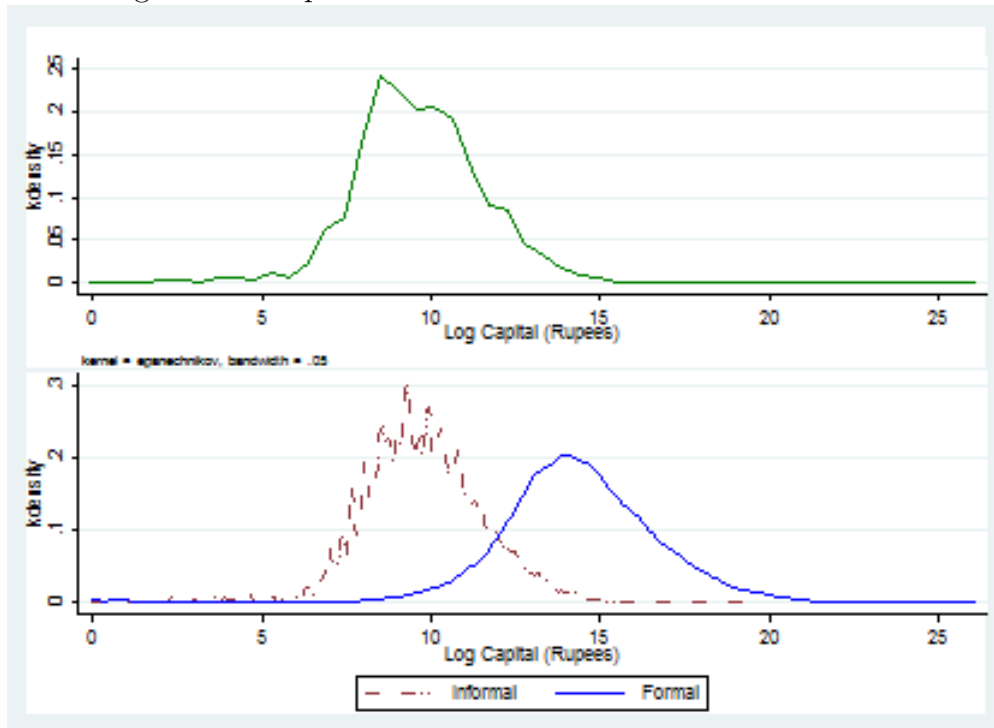
Figure 2.3: Intermediate Inputs in Formal and Informal Sectors 2004-05
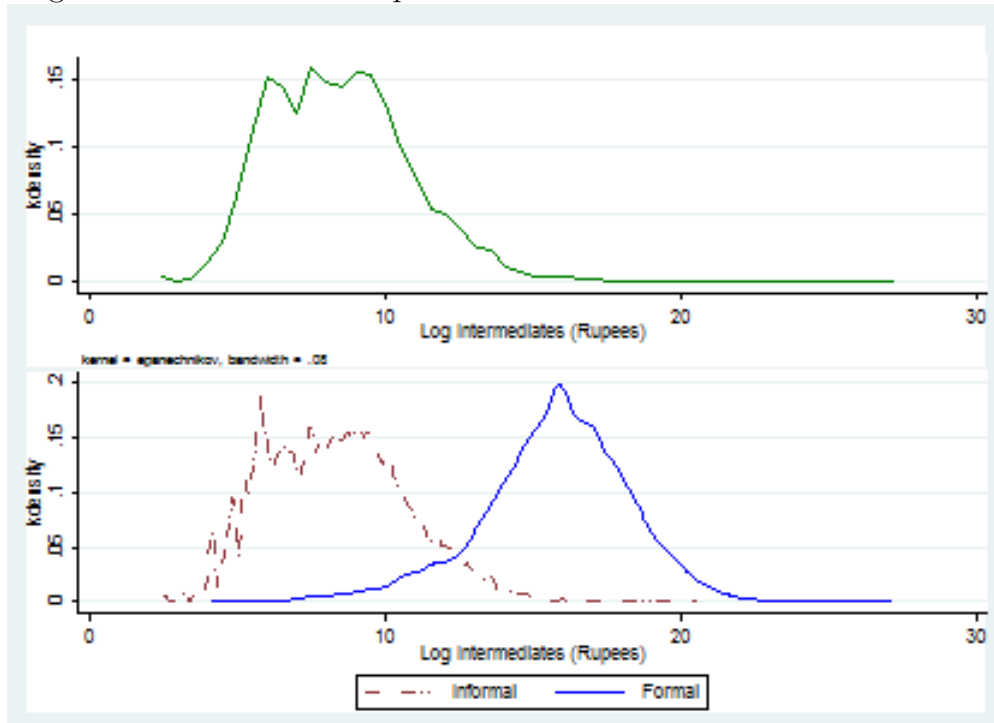
Table 2.1: Key Features of Formal and Informal Manufacturing in India

Informal Sector			
	1994-95	2005-06	Percentage Change
Number of Firms	12,200,000	17,000,000	39
Total Output (Rupees Billion)	708	2,670	277
Total Output/Firm (Median Rupees)	13,150	20,400	55
Total Employment	29,100,000	36,300,000	25
Employment/Firm (Median)	2	2	0
Fixed Capital (Rupees Billion)	430	1,450	237
Capital/Firm (Median Rupees)	6,860	15,700	129

Formal Sector			
	1994-95	2005-06	Percentage Change
Number of Firms	107,992	131,268	22
Total Output (Rupees Billion)	5,180	19,084	268
Total Output/Firm (Median Rupees)	-	5,037,172	-
Total Employment	8,256,712	7,987,780	-3
Employment/Firm (Median)	-	14	-
Fixed Capital (Rupees Billion)	1,069	3,996	274
Capital/Firm (Median)	-	809,715	-

Source: The source for informal sector is the National Sample Survey Organization of India

and the source for the formal sector is the Annual Survey of Industries

ratio went up to 4.5 in 2005-06. This increase is explained by a 25% increase in job creation in the informal sector and a 3% decline in job creation in formal sector. Given the growth rate of the number of firms and of total employment, and the fact that the median employment by a firm in the informal sector remained at 2 employees for 1994-95 and 2005-06, one could say that most of the job creation in the informal sector was on account of entry of new firms rather than due to the expansion of existing firms. Also, the informal sector is characterized by self employment. Adding more workers is similar to adding more firms. For the formal sector we have firm level data only for the year 2004-05. So we are not able to comment on whether the decline in job creation was on account of firm exit or due to contraction of existing firms. The median employment of a firm in the formal sector in 2004-05 was 14 employees. Most of the job creation happened in the first half of the 10 year period. In fact, the formal sector had positive job creation in this period, around 12%.

3. Figure 2.1 plots the distribution of log output. The top half of the figure plots the distribution of log total output which combines the output of the formal and informal sectors. The bottom half of the figure plots the distribution of log output for the formal and informal sectors separately. The large presence of the informal sector is once again highlighted in this figure. The distribution of log total output almost mirrors the distribution of log output of the informal sector except for the right tail. The formal sector log output makes up for a large part of the right tail of the distribution for total log output. Despite such a large presence of the informal sector it contributed to only 15% of the total output in 2004-05. The formal sector firms accounted for the remaining 85% of the total output in the manufacturing sector. The median value of output in the formal sector was 247 times more than the median output in the informal sector. Output increased for both sectors in the period under study but the formal sector output grew at a slightly higher rate than the informal sector output. Once again most of the increase in output happened in the first half of the 10 years in both the sectors.

4. Figure 2.2 plots the distribution of log capital. The formal sector as a whole uses 2.5 to 2.75 times more capital than the informal sector. When we compare the median values of capital employed in the two sectors we find that a firm in the formal sector employs 50 times more capital than a firm in the informal sector. The median value of capital employed in the formal sector was Rs. 809,715 in 2004-05 and Rs. 15,700 in the informal sector. It would be interesting to see the initial capital requirements in the two sectors. The per capita income for India in 2004-05 was between Rs. 25,000 and Rs. 30,000. For an average Indian entrepreneur, it was perhaps be easier to set up and run an informal enterprise. 63% of the formal firms had recieved institutional loans in 2004-05 as opposed to only 2.4% of the informal firms. Informal firms largely depend on personal savings and loans from money lenders and other informal sources of finance.

To summarize, the biggest contribution of the informal sector is in providing employment and livelihood to millions. The contribution of the informal sector towards gross output is very small compared to the formal sector. A median firm in the formal sector employs significantly higher amounts of capital, labor and intermediate inputs (see Figure 2.3) and produces significantly higher amount of output compared to a median firm in the informal sector. Given these findings, the interesting question one would ask is why is there such a large informal sector in India? How does the large informal sector affect India's total factor productivity? Informal firms are not just very small but also appear to be less productive than the formal firms. Only about 1.5% of the owners of the informal firms have an undergraduate degree. More than 99% of the firms in the informal sector are sole propreitorships compared to 27% of the firms in the formal sector. 37% of the formal sector firms were partnerships, 1% co-operative societies and the remaining 35% limited companies. Less than 6% of the informal firms surveyed maintained proper books of accounts. According to Bloom et al. (2011) the lack of delegation and professional management, which seems to be the main characteristic of informal firms, could lead to lower productivity. Productivity, however, is not just a function of ability but also a function of institutions and environment. Are firms informal because they are of low ability type or because of institutional factors that make formality costly? In the next section we present some of the views on the informal sector exisiting in the literature.

2.3 Informal Sector: Definition and Views

2.3.1 Definition of Informality

We begin this section by defining the term informality. There are several challenges when it comes to precisely defining what constitutes the informal sector. A wide range of terms such as unregistered, non-observed, unorganized, hidden, subterranean, unrecorded and black economy are used interchangeably with the term informal. The common thread is that these activities are not recorded or are imperfectly reflected in the official national accounting systems. The 15th International Conference of Labor Statisticians (ICLS) in 1993 attempted to correct the vagueness and plurality associated with the term "informal by adopting an international statistical definition of the informal sector. This definition was included in the System of National Accounts (SNA 1993) and is endorsed by most countries. The 15th ICLS defined the informal sector in terms of characteristics of the enterprises (production units) in which the activities take place, rather than in terms of the characteristics of the persons involved or of their jobs. The 15th ICLS recommended using one or more of the following three criteria:

- non-registration of the enterprise

- small size in terms of employment; and

- non registration of the employees of the enterprise

In the Indian context, the terms "formal" and "informal" have not been used in the official statistics. The terms used are the "organized" and "unorganized" sectors. The informal sector and the unorganized sector are quite close though not exactly the same. The organized sector comprises of enterprises for which the statistics are available regularly from the budget documents and the annual reports in the case of the public sector firms and through the Annual Survey of Industries (ASI) in case of registered manufacturing firms. On the other hand, the unorganized sector refers to those firms whose activities or collection of data is not regulated under any legal provision and / or which do not maintain any regular accounts. Non-availability of regular information has been the main criteria for treating the sector as unorganized. The National Sample Survey Organization (NSSO) of India collects data on the unorganized sector every five years. For the purpose of this study we treat the organized sector as the formal sector and a subset of the unorganized sector as the informal sector, the details of which are discussed in the section on Data Description.

2.3.2 Views on Informality

The informal sector is the subject of a vast literature and there are several views on it. One of the earliest studies on the informal sector was by Lewis (1954). Although he did not use the term informal, he introduced the concept of dual markets where informality could be equated with the low-wage, low-productivity segment of a dual market. Harris and Todaro (1970) extended this idea of dualism to a model with rural-urban migration. When the migration rate exceeds the rate at which urban jobs are created, many people end up taking up unproductive or under productive employment. Thus, the low-productive sector provides the necessary safety net for the poor and unemployed. The term informal sector was first used by Hart (1972) in an International Labor Organization (ILO) study of urban labor markets in Kenya. The findings of De Soto (1989) on the informal sector in Peru and other Latin American countries have been highly influential. According to De Soto, the informal sector is a rational response to excessive regulation. He argued that the entrepreneurial spirit of the informal sector needs to be fostered through deregulation and improvement in private property rights.

Following De Soto's pioneering work several studies have modeled the informal sector as a rational response to the legal and regulatory environment. This includes the work by Amaral and Quintin (2006), Azuma and Grossman (2008) and Paula and Scheinkman (2007). In these models, agents are heterogeneous in their managerial abilities. Agents with the lowest managerial ability become workers and the ones with the highest ability become formal managers, with the intermediate group running informal firms. Managers with greater abilities are able to run big firms and employ large amounts of capital. This makes them want to join the formal sector, where the marginal cost of capital is lower than the informal sector. The marginal firm trades off the cost of paying taxes in the formal sector versus the higher cost of capital and scale limitations in the informal sector. As a result, the marginal firm in the informal sector would use less capital and labor than it would if it joined the formal sector.

A less romantic view of the informal sector is associated with the studies published by Farrell (2004) and Bailey et al. (2005) of the McKinsey Global Institute. These studies view the informal sector as a collection of firms that remain hidden and small, and therefore unproductive, in order to avoid taxes and regulations, which in turn allows them charge lower prices. According to this view unleashing the informal firms will not lead to any development. What they propose instead is the displacement of the informal firms by formal and more efficiently run formal firms, a concept akin to the "Walmart effect. Another approach to modeling the informal sector is based on public choice theory. According to Loayza (1997), the state, as the institution that both monitors the regulatory and enforcement systems and administers public services, plays a crucial role in the formation of informal economies. If state officials, or interest groups related to them, profit in some way from the presence of the informal sector, they will create an environment that makes informality attractive or simply

unavoidable. Azuma and Grossman (2008) show that a large informal sector exists because productive endowments are unobservable and therefore the government cannot adjust the taxes it extracts from the formal sector producers according to each producers endowment. Given this constraint, if the distribution of endowments in not egalitarian, then a profit maximizing government would extract a large enough amount from the producers in the formal sector and that the poorly endowed producers would choose to work in the informal sector.

Although several views of the informal sector exist there is very little empirical evidence available to reject one view in favor of the other. Some attempts in this direction include the work of Schneider (2007) who estimates the size of the informal sector in 145 countries and shows that the high burden of taxation and social security laws are the main determinants of the size of the informal sector. Similar results are obtained by Loayza and Rigolini (2006). Pavnick and Goldberg (2005) find evidence that trade liberalization leads to greater informality when it is accompanied by labor reforms that allow labor markets to be more competitive in Brazil and Colombia. Trade liberalization increases competition from foreign firms and domestic formal firms cut labor costs by switching from permanent labor to part-time labor and sub-contracting. In the case of India, Besley and Burgess (2004) find that the states that passed pro-worker regulations saw a decline in output in the the formal manufacturing sector and an increase in output in the informal manufacturing sector.

La-Porta and Schleifer (2008) provide a helpful summary of the different views on the informal sector. They analyze data from the World Bank Informal and Micro Surveys for the poor and developing countries during the period 2003 to 2006. They find evidence that supports neither the romantic view endorsed by De Soto nor the parasitic view endorsed by the McKinsey Global Institute studies. Their study supports the Harris-Todaro's dual view of the informal sector as being the residual sector that comprises of unproductive labor that is unabsorbed by the formal sector. Using the evidence presented in section 2.2 of this chapter, we tend to agree with La Porta and Scheifer that the informal firms are typically of low productivity type. However one is not sure that moving firms from the informal sector to the formal sector would necessarily boost TFP as claimed by the authors, who compared small firms in the formal sector with the informal firms and found large productivity gaps.

There are two reasons why one would exercise caution before accepting the views of La-Porta and Schleifer (2008). One, the authors use value added per worker as a measure of labor productivity. There are several authors who have pointed out that differences in revenue productivity reflect differences in distortions rather than difference in true productivity (Foster et al. (2008), Jones (2008) and Hsieh and Klenow (2009)). Without firm level deflators we cannot accurately measure true productivity. Moreover, aggregate productivity is a combination of technology and the efficiency. A comparative study of the formal and informal sector needs to look at differences in both actual productivities and the efficiency with which resources are allocated. How efficiently firms allocate resources is a function of institutions, markets and government policy.

The empirical literature on informality has paid very little attention to the issue of allocation of resources. The non availability of micro data on the informal sector is perhaps the main reason for this. Given that there is a large data set on informal firms now available for India and that resource allocation has been studied for the formal sector by Hsieh and Klenow (2009) it is only logical for one to make a similar inquiry about the informal sector. In this chapter we do not focus much on why firms become informal but instead focus on how they allocate resources once they become informal. We are interested in knowing if there are any differences in resource misallocation between the formal and informal sectors and if the differences can help us identify the sources of misallocation. In the next section we build a model of monopolistic competition to measure the extent of resource misallocation in the economy.

2.4   Model

2.4.1   The Economic Environment

This is a static one period model of a closed economy with no uncertainty. There are S industries in this economy indexed by $s = 1, 2, ...S$. Within each industry there are N_s number of monopolistic competitive firms indexed by $i = 1, 2, ...N_s$. Firms are heterogeneous in productivity, A_{si}, which is a random draw from a distribution $F(a)$. Each firm combines labor, capital and intermediate inputs using Cobb Douglas technology to produce output Y_{si}.

$$Y_{si} = A_{si} K_{si}^{\alpha_s(1-\sigma_s)} H_{si}^{(1-\alpha_s)(1-\sigma_s)} M_{si}^{\sigma_s} \qquad (2.1)$$

K_{si} and H_{si} are quantities of physical and human capital used by firm i in industry s. To keep things simple we assume that all intermediate inputs are aggregated into one single intermediate good M. This method is similar to aggregating capital when capital is endogenous. Intermediate goods are in fact just like capital but with 100% depreciation. M_{si} is the quantity of inptermediate input used by firm i in industry s. We assume that the share of factor inputs, α_s and σ_s is the same for all firms within an industry and that $0 < \alpha_s < 1$ and $0 < \sigma_s < 1$. Each firm i in industry s produces output Y_{si} that can be used for final consumption by consumers or for intermediate usage by other firms. The firm i could be formal or informal. The production process is assumed to be the same for formal and informal firms. We are not interested in modeling firm entry and exit in this model. Instead we are interested in measuring the extent of misallocation of resources within an industry. C_{si} denotes the output of firm i in industry s that is used for final consumption and X_{si} denotes the output used for intermediate consumption.

$$Y_{si} = C_{si} + X_{si} \tag{2.2}$$

Since we are interested in looking at misallocation of firms within an industry we need an industry level aggregator. To keep things tractable and aggregation simple we assume that a representative firm for each industry s combines output of all N_s firms within that industry using CES technology with $\rho > 1$.

$$Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \tag{2.3}$$

The output Y_s of each industry s is sold either for intermediate use to the intermediate goods sector or for final consumption to the final goods sector at price P_s. Let X_s denote the output of industry s that is sold for intermediate use and C_s denote the output sold for final consumption. We express X_s and C_s using the indicator function in the following manner.

$$Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \mu(i) + \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} (1 - \mu(i)) \tag{2.4}$$

$$\text{where} \quad \mu(i) = 1 \ \text{ if } \ Y_{si} = C_{si}$$

$$\mu(i) = 0 \ \text{ if } \ Y_{si} = X_{si}$$

The above equation gives rise to the following constraint.

$$Y_s = C_s + X_s \tag{2.5}$$

The intermediate goods sector combines the intermediate inputs X_s from all S industries to produce the aggregate intermediate good, M, using Cobb Douglas technology. M is the aggregated intermediate input which is sold to individual firms. The final goods sector combines the output for final consumption from all industries, C_s to produce final good, C in a similar fashion.

$$M = \prod_{s=1}^{S} X_s^{\lambda_s}, \quad \sum_{s=1}^{S} \lambda_s = 1 \tag{2.6}$$

15

$$C = \prod_{s=1}^{S} C_s^{\theta_s}, \quad \sum_{s=1}^{S} \theta_s = 1 \tag{2.7}$$

The formulation of the equation for the intermediate good M warrants more explanation. λ_s is the share of each industry in the single intermediate good M. All firms in the economy use this single good M which combines intermediate output of different industries in fixed proportions. That is, all firms use the same combination of intermediate inputs and at the same time but differ in their amount of use. A more sophisticated approach to modeling intermediate inputs would be required if we were to study linkages between industries and use data from input-output tables. Input-output tables are however not available separately for the formal and informal sectors in India and we only have data on the amount of intermediate inputs expenditure made by firms.

We assume that all firms across all industries face the same wage rate w, interest rate r and intermediate good price q. Supply of human capital and capital is exogenous. The resource constraints in the economy are.

$$\sum_{s=1}^{S} \sum_{i=1}^{N_s} H_{si} = H \tag{2.8}$$

$$\sum_{s=1}^{S} \sum_{i=1}^{N_s} K_{si} = K \tag{2.9}$$

$$\sum_{s=1}^{S} \sum_{i=1}^{N_s} M_{si} = M \tag{2.10}$$

2.4.2 Optimization Problems

Final Good Problem : There is a single final good C produced by a representative firm in a perfectly competitive market for the final good. This firm combines the output C_s of S industries in the economy using Cobb Douglas technology and taking prices P_s as given. This final good is used for consumption only and is equal to the GDP of the economy. We normalize the price of the final good, P, to 1.

$$\text{Max } PC - \sum_{s=1}^{S} P_s C_s \tag{2.11}$$

$$\text{subject to} \quad C = \prod_{s=1}^{S} C_s^{\theta_s}$$

Intermediate Good Problem: There is a single intermediate good M produced by a representative firm in a perfectly competitive market for the the intermediate good. This firm combines the output X_s of S industries in the economy using Cobb Douglas technology and taking prices P_s as given.

$$\text{Max} \quad qM - \sum_{s=1}^{S} P_s X_s \tag{2.12}$$

$$\text{subject to} \quad M = \prod_{s=1}^{S} X_s^{\theta_s}$$

Industry S Problem: A representative firm solves the following maximization problem taking prices P_{si} as given

$$\text{Max} \quad P_s Y_s - \sum_{i=1}^{N_s} P_{si} Y_{si} \tag{2.13}$$

$$\text{subject to} \quad Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$$

Firm i problem : Firm i solves the following maximization problem taking wages, rental rates, price of intermediate input M, industry S demand and distortions as given

$$\text{Max} \quad \prod_{si} = P_{si} Y_{si}(1 - \tau_{ysi}) - w H_{si} - r(1 + \tau_{ksi})K_{si} - q(1 + \tau_{msi})M_{si} \tag{2.14}$$

$$\text{subject to} \quad Y_{si} = A_{si} K_{si}^{\alpha_s(1-\sigma_s)} H_{si}^{(1-\alpha_s)(1-\sigma_s)} M_{si}^{\sigma_s}$$

τ_{ysi} is the distortion that affects the marginal products of all inputs in the same proportion. τ_{ksi} denotes distortion that affects the marginal product of physical capital relative to human capital. τ_{msi} denotes distortion that affects the marginal product of intermediate goods relative to human capital. Size restrictions or output taxes are examples of distortions that affect the marginal product of all factors inputs. Obtaining subsidized credit or requirement of property as collateral for loans are examples of distortions to capital alone. An example of distortion to intermediate goods would be obtaining government clearances for importing goods beyond a certain level. Distortions to output in a particular industry s can show up as distortions to a firm in industry s' if firms in industry s' use output from industry s as intermediate inputs.

2.4.3 Defining Competitive Equilibrium

A competitive equilibrium in this environment consists of quantities Y, C, M, Y_s, C_s, X_s, Y_{si}, C_{si}, X_{si}, K_{si}, H_{si}, M_{si} and prices q, P_s, P_{si}, r, w for $s = $ 1,2,3,..S i $= 1, 2, 3, ... N_s$ such that

1. C_s solves final good sector problem

2. X_s solves intermediate good sector problem

3. C_{si} and X_{si} solves the industry s problem

4. K_{si}, H_{si} and M_{si} solve the firm i problem

5. Markets Clear:

   w clears the labor market,

   $$\sum_{s=1}^{S}\sum_{i=1}^{N_s}H_{si} = H$$

   r clears the capital market,

   $$\sum_{s=1}^{S}\sum_{i=1}^{N_s}K_{si} = K$$

   q clears the intermediate input market,

   $$\sum_{s=1}^{S}\sum_{i=1}^{N_s}M_{si} = M$$

   P_{si} clears the market for i variety,

   $$Y_{si} = C_{si} + X_{si}$$

   P_s clears the market for industry S output,

   $$Y_s = C_s + X_s$$

6. Others

   $$Y_{si} = A_{si}K_{si}^{\alpha_s(1-\sigma_s)}H_{si}^{(1-\alpha_s)(1-\sigma_s)}M_{si}^{\sigma_s}$$

$$Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}} \mu(i) + \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}} (1 - \mu(i))$$

$$\text{where} \quad \mu(i) = 1 \;\text{ if }\; Y_{si} = C_{si}$$

$$\mu(i) = 0 \;\text{ if }\; Y_{si} = X_{si}$$

$$M = \prod_{s=1}^{S} X_s^{\lambda_s}, \quad \sum_{s=1}^{S} \lambda_s = 1$$

$$C = \prod_{s=1}^{S} C_s^{\theta_s}, \quad \sum_{s=1}^{S} \theta_s = 1$$

$$Y = C$$

GDP, Y, in the economy is given by C. We have assumed that revenue from distortions are collected and rebated to households as a lump sum payment. We have 17 endogenous variables and 17 equations. We can now solve for equilibrium.

2.4.4 Solving for equilibrium

Solving the final goods problem yields,

$$P_s C_s = \theta_s PC, \quad P = \prod_{s=1}^{S} \left(\frac{P_s}{\theta_s}\right)^{\theta_s} \tag{2.15}$$

We normalize prices of final good C by setting them equal to 1. Solving the intermediate goods problem yields,

$$P_s X_s = \lambda_s q M, \quad q = \prod_{s=1}^{S} \left(\frac{P_s}{\lambda_s}\right)^{\lambda_s} \tag{2.16}$$

Solving industry S problem yields,

$$C_{si} = P_{si}^{-\rho} P_s^{\rho} C_s \qquad (2.17)$$

$$X_{si} = P_{si}^{-\rho} P_s^{\rho} X_s \qquad (2.18)$$

$$P_s = \left(\sum_{i=1}^{N_s} P_{si}^{1-\rho}\right)^{\frac{1}{1-\rho}} \qquad (2.19)$$

Solving the profit maximization of firm i in industry s yields,

$$K_{si} = \frac{\rho}{\rho-1} \frac{\alpha_s(1-\sigma_s)P_{si}Y_{si}(1-\tau_{ysi})}{r(1+\tau_{ksi})} \qquad (2.20)$$

$$H_{si} = \frac{\rho}{\rho-1} \frac{(1-\alpha_s)(1-\sigma_s)P_{si}Y_{si}(1-\tau_{ysi})}{w} \qquad (2.21)$$

$$M_{si} = \frac{\rho}{\rho-1} \frac{\sigma_s P_{si}Y_{si}(1-\tau_{ysi})}{q(1+\tau_{msi})} \qquad (2.22)$$

$$P_{si} = \frac{\rho}{\rho-1} \frac{mc}{\epsilon} \frac{(1+\tau_{ksi})^{\alpha_s(1-\sigma_s)}(1+\tau_{msi})^{\sigma_s}}{A_{si}(1-\tau_{ysi})} \qquad (2.23)$$

$$\text{where} \quad mc = r^{\alpha_s(1-\sigma_s)} w^{(1-\alpha_s)(1-\sigma_s)} q^{\sigma_s}$$

$$\epsilon = (\alpha_s(1-\sigma_s))^{\alpha_s(1-\sigma_s)}((1-\alpha_s)(1-\sigma_s))^{(1-\alpha_s)(1-\sigma_s)}(\sigma_s)^{\sigma_s}$$

The expression for price in equation (2.23) is obtained by combining the first order conditions for physical capital, human capital and intermediate goods with the expression for the production function. Price is a function of markup, marginal cost and distortions. When markups and marginal cost are constant and distortions are symmetric across all firms within an industry revenue productivity, which is given by $P_{si}A_{si}$, would be the same for all firms within an industry. We could then obtain measure of true productivity, A_{si}, from revenue productivity by using industry level deflators. However if distortions are non symmetric,

20

then in the absence of firm level deflators revenue productivity is a not a measure of true productivity but instead is a measure of resource misallocation.

We now define the terms $TFPR_{si}$ and $\overline{TFPR}_s$. $TFPR_{si}$ denotes the total factor revenue productivity of firm i in industry s. It is also a measure of firm level distortion as explained above. $\overline{TFPR}_s$ denotes the average level of distortion in industry s. It is a geometric weighted average of distortions to factor inputs, where the weights are the firm's share in total industry output.

$$TFPR_{si} = P_{si}A_{si} = \frac{\rho}{\rho-1}\frac{mc}{\epsilon}\frac{(1+\tau_{ksi})^{\alpha_s(1-\sigma_s)}(1+\tau_{msi})^{\sigma_s}}{(1-\tau_{ysi})} \tag{2.24}$$

$$\overline{TFPR}_s = \frac{\rho}{\rho-1}\frac{mc}{\epsilon}\left[\frac{1}{\sum_{i=1}^{N_s}\frac{(1-\tau_{ysi})}{(1+\tau_{ksi})}\frac{P_{si}Y_{si}}{P_sY_s}}\right]^A\left[\frac{1}{\sum_{i=1}^{N_s}(1-\tau_{ysi})\frac{P_{si}Y_{si}}{P_sY_s}}\right]^B\left[\frac{1}{\sum_{i=1}^{N_s}\frac{(1-\tau_{ysi})}{(1+\tau_{msi})}\frac{P_{si}Y_{si}}{P_sY_s}}\right]^C \tag{2.25}$$

$$\text{where,}\quad A = \alpha_s(1-\sigma_s)$$

$$B = (1-\alpha_s)(1-\sigma_s)$$

$$C = \sigma_s$$

We now combine the expression for P_s in equation (2.19), the expression for P_{si} in equation (2.23), the expressions for K_{si}, H_{si} and M_{si} in equations (2.20), (2.21) and (2.22) with the equations for $TFPR_{si}$ and $\overline{TFPR}_s$ defined above to derive the following expression for industry output Y_s

$$Y_s = TFP_s K_s^{\alpha_s(1-\sigma_s)} H_s^{(1-\alpha_s)(1-\sigma_s)} M_s^{\sigma_s} \tag{2.26}$$

$$\text{where}\quad TFP_s = \left[\sum_{i=1}^{N_s}\left(\frac{A_{si}\overline{TFPR}_s}{TFPR_{si}}\right)^{\rho-1}\right]^{\frac{1}{\rho-1}} \tag{2.27}$$

$$K_s = \sum_{i=1}^{N_s}K_{si} = \frac{\rho}{\rho-1}\frac{\alpha_s(1-\sigma_s)}{r}\sum_{i=1}^{N_s}\frac{(1-\tau_{ysi})P_{si}Y_{si}}{(1+\tau_{ksi})} \tag{2.28}$$

$$H_s = \sum_{i=1}^{N_s}H_{si} = \frac{\rho}{\rho-1}\frac{(1-\alpha_s)(1-\sigma_s)}{w}\sum_{i=1}^{N_s}(1-\tau_{ysi})P_{si}Y_{si} \tag{2.29}$$

21

$$M_s = \sum_{i=1}^{N_s} M_{si} = \frac{\rho}{\rho - 1} \frac{\sigma_s}{q} \sum_{i=1}^{N_s} \frac{(1 - \tau_{ysi})P_{si}Y_{si}}{(1 + \tau_{msi})} \qquad (2.30)$$

Equation (2.26) shows that when distortions are non symmetric inefficiency at the firm level will affect industry TFP. When distortions are symmetric across firms, marginal revenues would be equalized and TFP_s is equal to $\left[\sum_{i=1}^{N_s} A_{si}^{\rho-1} \right]^{\frac{1}{\rho-1}}$. To illustrate how micro level distortions can lower aggregate TFP let us assume that true productivity, $\log A_{si}$, and log(TFPR) are jointly normally distributed. log(TFPR) is the deviation of a firm's TFPR from the industry mean or the firm's inefficiency relative to other firms in the industry. It is equal to $\log \left(\frac{TFPR_{si}}{TFPR} \right)$. We can now express industry TFP as follows:

$$logTFPR_s = \frac{1}{\rho - 1} log \left(\sum_{i=s}^{N_s} A_{si}^{\rho-1} \right) - \frac{\rho}{2} var(logTFPR_{si}) \qquad (2.31)$$

Equation (2.31) allows us to characterize the negative effect of distortions by the variance of log TFPR. A larger variance of log TFPR would imply a larger degree of resource misallocation in an industry. We use the variance of log TFPR as the measure of misallocation in our empirical work. We now solve for the aggregate level of gross output and GDP.

We first substitute for M_s in equation (2.26) using equation (2.30) to get

$$Y_s = \left(\frac{\rho - 1}{\rho} \right)^{\frac{\sigma_s}{1-\sigma_s}} \left(\frac{P_s}{q} \right)^{\frac{\sigma_s}{1-\sigma_s}} \left(\sigma_s (1 - \overline{\tau}_{ms}) \right)^{\frac{\sigma_s}{1-\sigma_s}} TFP_s^{\frac{1}{1-\sigma_s}} K_s^{\alpha_s} H_s^{(1-\alpha_s)} \qquad (2.32)$$

$$\text{where} \quad (1 - \overline{\tau}_{ms}) = \left[\sum_{i=1}^{N_s} \frac{(1 - \tau_{ysi})}{(1 + \tau_{msi})} \frac{P_{si}Y_{si}}{P_s Y_s} \right]$$

According to equation (2.32) the output level of industry s would depend on the elasticity of substitution between firm outputs in industry s, the relative prices of final and intermediate good, intermediate goods share in industry s, the mean distortions to intermediate goods in industry s, TFP of industry s and finally physical and human capital allocated to industry s. We now further aggregate M_s over s industries to get:

$$qM = \left[\left(\frac{\rho - 1}{\rho} \right) \left(\sigma(1 - \overline{\tau}_m) \right) \right] \sum_{s}^{S} P_s Y_s \qquad (2.33)$$

$$\text{where} \quad \sigma(1 - \overline{\tau}_m) = \sum_s^S \sigma_s(1 - \overline{\tau}_{ms}) \frac{P_s Y_s}{\sum_s^S P_s Y_s}$$

The nominal value of output of industry s is given by $P_s Y_s = P_s C_s + P_s X_s$. Combining this expression with the demand for final and intermediate good in equations (2.15) and (2.16) and aggregating over S industries gives us

$$C = \sum_s^S P_s Y_s - qM \tag{2.34}$$

Combining equations (2.33) and (2.34) gives us expression for GDP, Y=C, as a function of total gross output in the economy.

$$Y = \left[1 - \left(\frac{\rho - 1}{\rho}\right)\left(\sigma(1 - \overline{\tau}_m)\right)\right] \sum_s^S P_s Y_s \tag{2.35}$$

$$\text{where} \quad \sum_s^S P_s Y_s = \sum_s^S \left[\left(\frac{\rho - 1}{\rho}\right)^{\frac{\sigma_s}{1-\sigma_s}} \left(\frac{P_s}{q^{\sigma_s}}\right)^{\frac{1}{1-\sigma_s}} \left(\sigma_s(1 - \overline{\tau}_{ms})\right)^{\frac{\sigma_s}{1-\sigma_s}} TFP_s^{\frac{1}{1-\sigma_s}} K_s^{\alpha_s} H_s^{(1-\alpha_s)}\right] \tag{2.36}$$

$$\text{and} \quad TFP_s = \left[\sum_{i=1}^{N_s} \left(\frac{A_{si}\overline{TFPR_s}}{TFPR_{si}}\right)^{\rho - 1}\right]^{\frac{1}{\rho - 1}}$$

The equations for gross output and GDP, equations (2.35) and (2.36), provide some key insights. First, in the case of symmetric distortions micro level misallocation does not affect aggregate productivity. Distortions will still lower output by affecting the allocation of physical and human capital. When there are non symmetric distortions within an industry, firm level misallocation will affect aggregate TFP besides affecting the allocation of physical and human capital. Second, the presence of intermediate goods can amplify the effect of distortions. This can happen in two ways. One, via TFP when non symmetric distortions are present. The elasticity of output with respect to TFP is given by $\frac{1}{(1-\sigma_s)}$. Suppose the share of intermediate inputs is 0.5. Any decrease in TFP due to distortion will be doubled via intermediate inputs. This is straightforward as intermediate goods provide the linkages for productivity or distortions to be carried across sectors. Another way in which intermediate inputs could affect GDP is if there are distortions to intermediate inputs. Even if distortions are symmetric across firms, any distortions to intermediate inputs will lead to lower gross output. This is captured by $(1 - \tau_{ms})$. The effect of distortions to intermediate on GDP is however ambiguous. One one hand gross output could fall as firms employ fewer intermediate inputs in production. But a decrease in the share of intermediate inputs will increase final consumption

as shown by equation (2.35). The share of intermediate goods is once again important in resolving this ambiguity. Thus, incorporating intermediate goods in crucial in correctly measuring the extent of resource misallocation.

We now turn our attention to empirical evidence on misallocation in manufacturing India. We first replicate Hsieh-Klenow (2009) methodology for measuring misallocation in the entire manufacturing sector by combining data on the formal and informal sectors. We are able to do this only for the year 2004-05 as the micro data for the formal sector is not available with us for the other years. We also present results on the formal and informal sectors individually. Hsieh and Klenow do not incorporate intermediate inputs in their calculations but instead rely on value added. But as highlighted above excluding intermediate goods in productivity calculations can be dangerous. So we recalculate misallocation by incorporating intermediate goods.

2.5   Data Description

Data for the formal sector is obtained from the Annual Survey of Industries (ASI) which is complied by the Central Statistical Organization (CSO) of India. The ASI covers manufacturing firms[1] registered under sections 2(m)(i) and 2(m)(ii) of the Factories Act, 1948, i.e. firms with 10 or more employees using power and firms with 20 or more employees not using power. The Survey also covers bidi and cigar manufacturing establishments registered under the Bidi Cigar Workers (Conditions of Employment) Act, 1966. This survey is conducted on an annual basis. For this study we use the data for the year 2004-05. The ASI for 2004-05 is a census of all registered manufacturing units with 100 or more and a random sample of one fifth of the remaining registered firms. We use sampling weights which are provided in the data set in all our calculations.

Data for the informal sector is obtained from the surveys of the unorganized manufacturing sector conducted by the National Sample Survey Organization (NSSO) of India. The NSSO conducts a survey of the unorganized manufacturing sector every five years. The unorganized manufacturing sector covers all manufacturing firms that are not covered by the ASI and all non-public sector firms in sectors like trade, transport, hotels restaurants, storage and warehousing, and services. In the unorganized sector, in addition to the unincorporated proprietary or partnership firms, firms run by cooperative societies, trusts, private and public limited companies (Non ASI) are also covered. The informal sector is therefore a subset of the unorganized sector and only includes unincorporated proprietary or partnership firms. We do not include the cooperative societies, trusts, private and public limited companies not covered under the ASI in the formal sector data in order to keep our results comparable to the results in Hsieh and Klenow (2009). Hsieh-Klenow only look at the manufacturing firms covered under the ASI. For this study we use data on the informal sector for the years 1994-95, 2000-01 and 2005-06[2].

---

[1]The ASI covers all regions in the country except for the States of Arunachal Pradesh, Mizoram,Sikkim and Union Territory of Lakshadweep.

[2]The NSSO collects data for the year starting in July 1st and ending in June 30th . The accounting year for ASI is from Arpil 1st to March 31st.

For our empirical analysis we use 4 digit industry level data for all years. For replicating Hsieh-Klenow statistics we use the same definitions of gross value added, human capital and physical capital stock as in their paper. Labor compensation is used to measure human capital and is defined as the sum of wages, benefits and bonuses. Capital stock is defined as average net book value of the firm's machinery, equipments and structures at the beginning and at the end of the year. Industry capital and labor shares are set equal to the corresponding US industry shares. This data is obtained from the NBER productivity database at the 4 digit industry level. We set the factor shares equal to the US shares in both the formal and informal sectors in order to identify the distortions in marginal products. We elaborate on our identification strategy in the next section. We lose a few industries because of the matching and end up with 174 different industries. Data on labor compensation in the NBER database which is based on the Census and ASM excludes fringe benefits and social security contributions. The labor share in the NBER database is around two thirds of what it would be if non-wage forms of compensation were included. As in Hsieh and Klenow we augment labor shares by a scaling factor of $3/2$. For the empirical analysis of our model which incorporates intermediate inputs, in addition to the labor input and capital stock defined above, we use total gross output and intermediate inputs. Gross output is the sum of the nominal value of production, trade income and other income such as rent or commission received and intermediate inputs are the sum of materials, energy and business services. We once again set the factor shares equal to the corresponding US industry shares[3].

2.6   Identification Strategy

In this section we discuss how distortions are identified in the data. The scale parameters in the production function namely, α_s and σ_s, are assumed to be the same for all firms within the industry s. These scale parameters are typically measured using data on factor shares. However, in the presence of distortions resources are not allocated efficiently and therefore revenue factor shares do not give correct values of the scale parameters. We can see this by rewriting the first order conditions of the firm's maximization problem given by equations (2.20), (2.21) and (2.22).

$$\frac{rK_{si}}{P_{si}Y_{si}} = \frac{\rho}{\rho-1} \frac{\alpha_s(1-\sigma_s)(1-\tau_{ysi})}{(1+\tau_{ksi})} \tag{2.37}$$

$$\frac{wH_{si}}{P_{si}Y_{si}} = \frac{\rho}{\rho-1}(1-\alpha_s)(1-\sigma_s)(1-\tau_{ysi}) \tag{2.38}$$

$$\frac{qM_{si}}{P_{si}Y_{si}} = \frac{\rho}{\rho-1} \frac{\sigma_s(1-\tau_{ysi})}{(1+\tau_{msi})} \tag{2.39}$$

---

[3]We use value added shares of capital and labor for the replication Hsieh-Klenow statistics. In the case of our extended model we use the total output shares of capital, labor and intermediate inputs.

Figure 2.4: Labour Share in Output: US V/s India
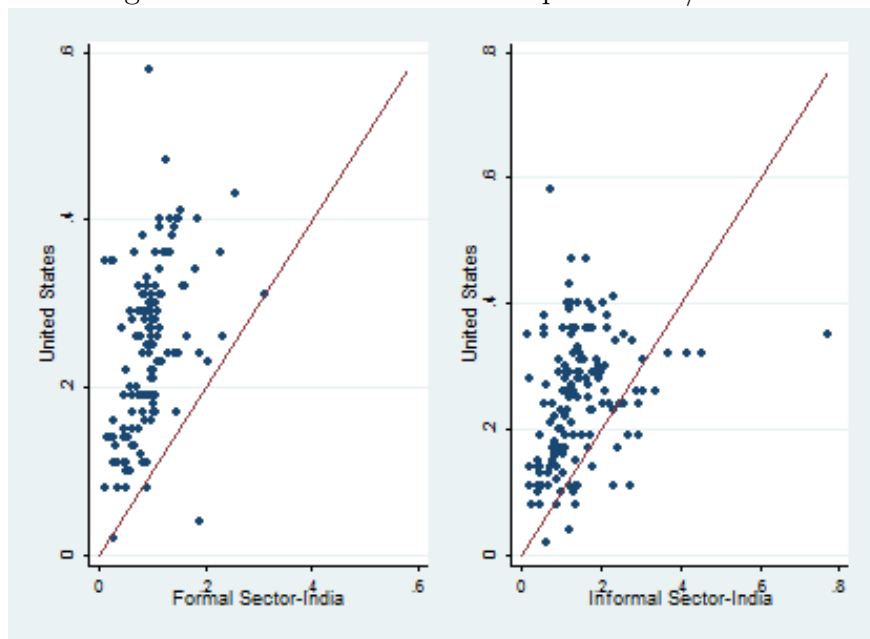
In order to identify the distortions in the data we set the factor shares equal to the equivalent US industry factor shares. Figures 2.4, 2.5 and 2.6 show the scatter plot of US factor shares and Indian factor shares (both formal and informal). Capital share is calculated as a residual after subtracting labor and intermediate inputs shares for both the US and India. The distance from the 45 degree line tells us the extent to which factor inputs in India are misallocated relative to the US. For the purpose of this identification we make the following assumptions. One, the US is the efficiency benchmark when it comes to resource allocation. Two, the production technologies used by the US firms and the Indian firms, both formal and informal firms, is the same. The first assumption is common in the literature. The second assumption may be questionable.

The production technologies of the US firms and the formal firms in India may not be very different. However the production technologies used by the informal sector firms in India may be very different from that of the US firms. Matching industries at the narrowest disaggregation possible is one way to address this issue. We are able to match industries only at the four digit level. The problem of different production technologies in the formal and informal sector may persist even after matching industries. Moreover we also assume that factor markets are perfect and factor prices being the same for both formal and informal firms. Firms in the informal sector typically use more labor than their formal counterparts. For example, an informal firm in the tailoring business may hire 5 workers to do a job whereas a formal firm may hire one worker and invest in a sewing machine to do the same job. The capital share of the formal firm would be higher than that of the informal firm holding everything else constant and perhaps closer the US counterpart. Applying the US shares to both the formal and informal firms in this case would estimate misallocation to be higher in the informal firm. We acknowledge that some of the assumptions we make to achieve identification may be strong and hope to address them more seriously in our future work.

Figure 2.5: Intermediate Goods Share in Output: US V/s India
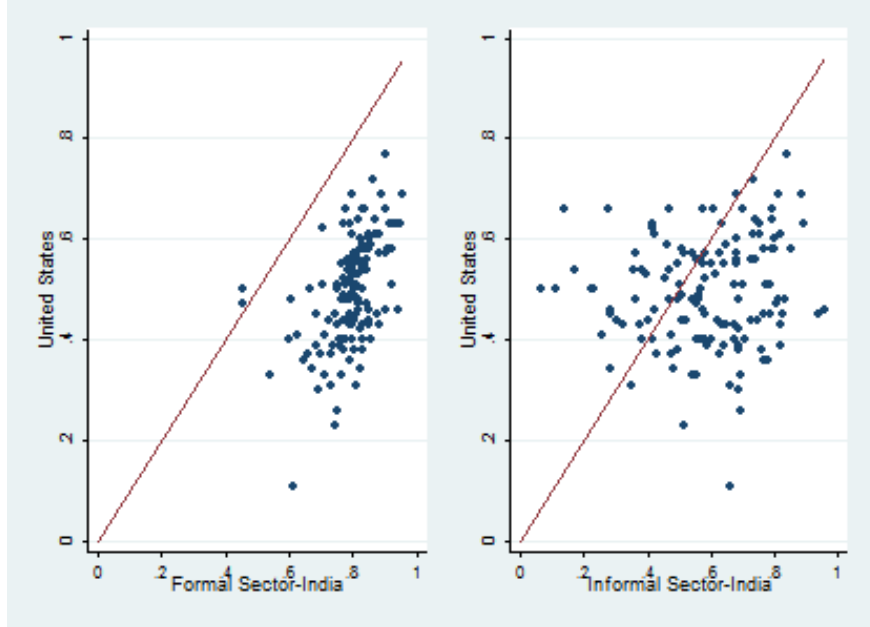
Figure 2.6: Capital Share in Output: US V/s India
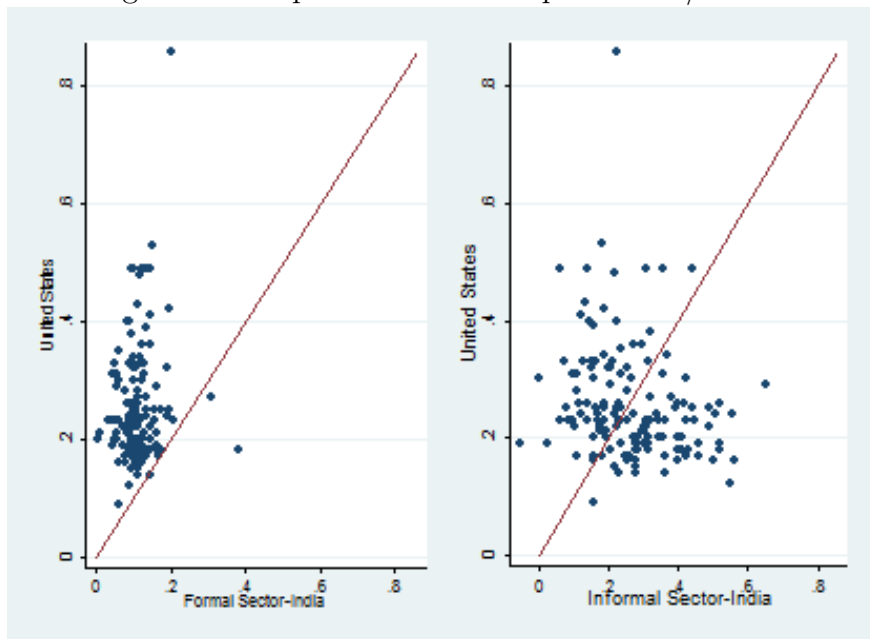
2.7 Empirical Results

We first measure the extent of misallocation using log TFPR. For our empirical estimation we infer distortions and productivity from the data in the following manner.

$$TFPR_{si} = P_{si}A_{si} = \frac{P_{si}Y_{si}}{K_{si}{}^{\alpha_s(1-\sigma_s)}H_{si}{}^{(1-\alpha_s)(1-\sigma_s)}M_{si}{}^{(\sigma_s)}} \tag{2.40}$$

$$\frac{(1-\tau_{ysi})}{(1+\tau_{ksi})} = \frac{\rho}{\rho-1}\frac{1}{\alpha_s(1-\sigma s)}\frac{rK_{si}}{P_{si}Y_{si}} \tag{2.41}$$

$$(1-\tau_{ysi}) = \frac{\rho}{\rho-1}\frac{1}{(1-\alpha_s)(1-\sigma s)}\frac{wL_{si}}{P_{si}Y_{si}} \tag{2.42}$$

$$\frac{(1-\tau_{ysi})}{(1+\tau_{msi})} = \frac{\rho}{\rho-1}\frac{1}{\sigma s}\frac{qM_{si}}{P_{si}Y_{si}} \tag{2.43}$$

$$A_{si} = \kappa_s \frac{P_{si}Y_{si}{}^{\frac{\rho}{\rho-1}}}{K_{si}{}^{\alpha_s(1-\sigma_s)}H_{si}{}^{(1-\alpha_s)(1-\sigma_s)}M_{si}{}^{(\sigma_s)}} \tag{2.44}$$

Since we do not observe true productivity, A_{si}, in the data, we use the elasticity of substitution to back out true productivity from revenue productivity in equation (2.44). Firms that have higher true productivity will be able to charge lower prices in our model and this translates into higher demand. We try to replicate Hsieh Klenow methodology as closely as possible and therefore our choice of parameter values are the same as the values chosen by Hsieh and Klenow. We set our elasticity of substitution parameter ρ equal to 3 and the rental rate to 10%. The term κ_s can be thought of as an industry deflator. Since relative productivities are unaffected by κ we set its value to 1. As in Hsieh and Klenow we first trim 1% of the tails of log TFPR to reduce the influence of outliers and then recalculate log TFPR and gains from reallocation.

2.7.1 Value Added Approach

We first replicate the methodology used by Hsieh and Klenow (2009) to calculate misalloation in the informal sector in India. Hsieh-Klenow do not incorporate intermediate goods in their model and measures of distortion are based on value added. They calculate misallocation in the formal manufacturing sector in India between the years 1987-88 and 1994-95. We calculate log TFPR and gains from reallocation for the entire manufacturing sector (both formal and informal) for the year 2004-05. We also calculate log TFPR and gains from reallocation for the years 1994-95 and 2000-01 for the informal manufacuring sector in India. One of the shortcomings in the paper by Hsieh and Klenow is that they are not able to measure the effect of reforms on misallocation successfully as

Figure 2.7: TFPR using Value Added: Formal V/s Informal 2004-05
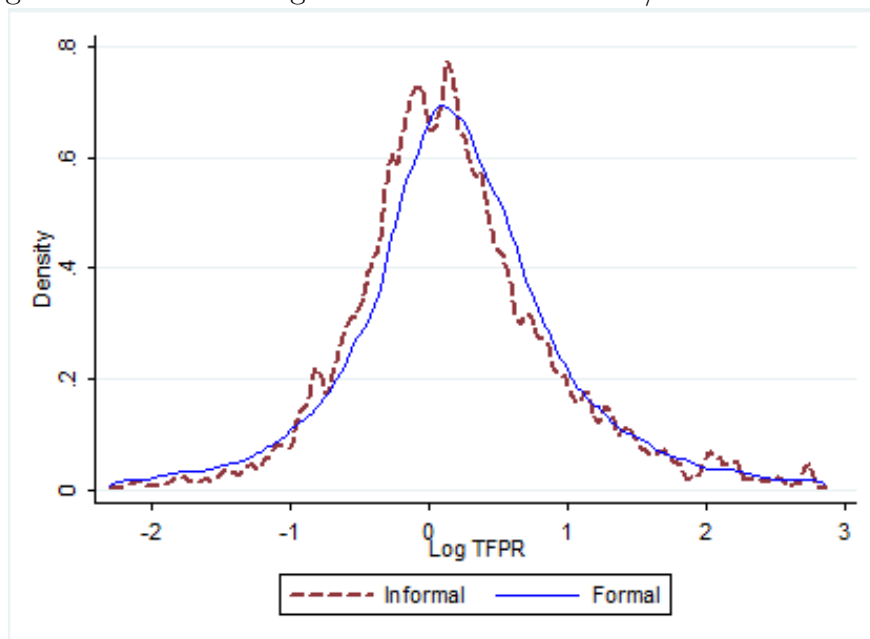
their data set ends in 1994-95. Although some reforms with regard to licensing were initiated in the 1980s in India bulk of the reforms took place after 1991, especially during the latter half of the 90s. By combining the results of Hsieh-Klenow for the formal sector for 1994-95, our results for the formal sector in 2004-05 and the informal sector for the years 1994-95, 2000-01 and 2005-06 we are able to analyze the effect of reforms on misallocation for the period 1994-2005.

The entry and exit of firms into the informal sector may be dependent on government policy and regulation. But once a firm becomes informal government policies such as taxation, soical security requirements etc do not apply to them. If distortions other than those caused due to government policy are the same for both the formal and informal sectors then the differences in log TFPR for the two sectors should give us an estimate of the magnitude of distortion due to government

Table 2.2: Dispersion of log TFPR using Value Added: 2004-05

	All	Formal	Informal
stdev	0.75	0.80	0.65
75-25	0.83	0.82	0.75
90-10	1.81	1.83	1.60

29

Figure 2.8: TFPR using Value Added: Informal
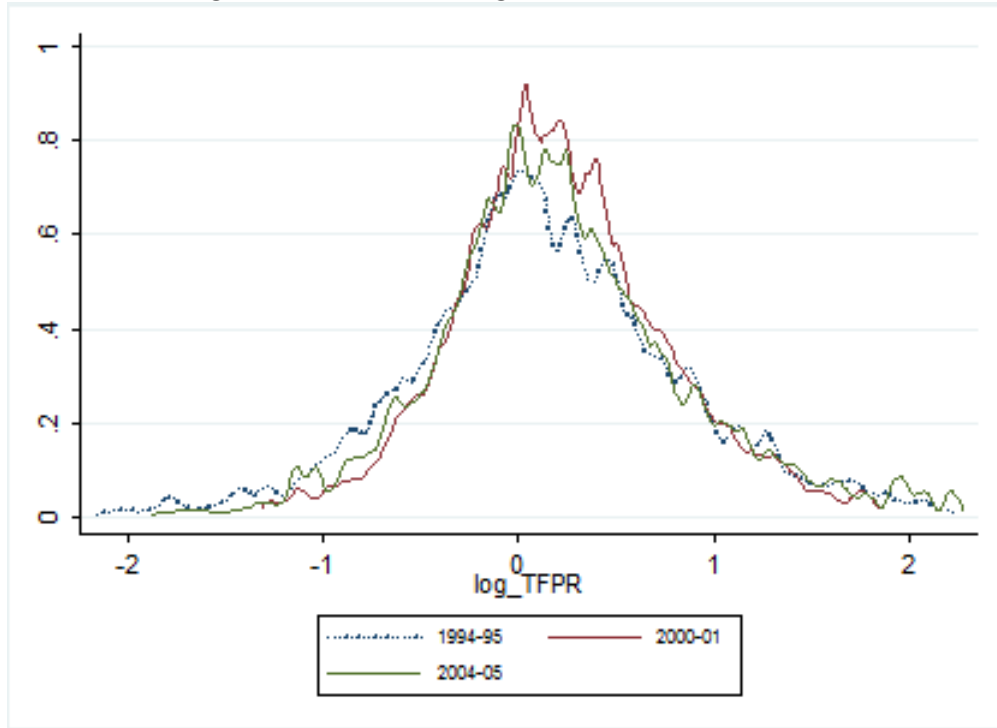
Table 2.3: Dispersion of log TFPR using Value Added: Formal

	1994-95(HK)	2004-05
stdev	0.67	0.80
75-25	0.81	0.82
90-10	1.60	1.83

Table 2.4: Dispersion of log TFPR using Value Added: Informal

	1994-95	2000-01	2004-05
stdev	0.69	0.54	0.65
75-25	0.83	0.66	0.75
90-10	1.74	1.35	1.60

policy. Figure 2.7 plots the distribution of log TFPR using value added for the formal and informal sectors for the year 2004-05. Figure 2.8 plots the distribution of log TFPR using value added for the informal for the years 1994-95, 2000-01 and 2004-05. Tables 2.2, 2.3 and 2.4 show the standard deviation of log TFPR, the difference between the 75th and 25th percentiles and the difference between the 90th and 10th percentiles for the formal and informal manufacturing sector in India using value added. Table 2.2 presents the dispersion in log TFPR for the entire manufacturing sector (formal and informal combined) and for the the formal and informal sectors individually. Table 2.3 presents the results for the formal sector alone for the years 1994-95 and 2004-05. The numbers for the years 1994-95 for the formal sector are taken from Hsieh and Klenow (2009). Table 2.4 presents the results for the informal sector alone for the years 1994-94, 2000-01 and 2004-05. We make the following observations:

1. We find that resource misallocation has worsened in the formal manufacturing sector in India during the period 1994-95 to 2004-05. The standard deviation of log TFPR has increased from 0.67 to 0.80. Most of this increase can be attributed to the action at the tails. The ratio of the 75th to 25th percentiles was 2.3 in level terms for both 1994-95 and 2004-05. The ratio of the 90th to 10th percentile however increased from 5.0 to 6.2 in level terms.

2. When we compare log TFPR for the two sectors we find that the misallocation in the formal sector was slightly lower than the misallocation in the informal sector in 1994-95. Most of this difference was at the tails. The ratio of the 90th to 10th percentile for the formal sector in levels was 5.7 compared 5.0 in the formal sector. This however changed in 2005-06. The standard deviation of log TFPR was 1.2 times higher in the formal sector compared to the informal sector. As mentioned earlier the source of increased misallocation is at the tails.

3. We find that for the informal sector resource misallocation has reduced marginally during the period 1994-95 to 2005-06. If we divide the period into two halves we find that resource misallocation reduced quite a bit in the first half but went up again in the second half. We had observed a similar trend earlier while we were describing facts about the informal sector. Looking at aggregate data we found that both the formal and informal sector performed much better in terms of output and job creation in the first of the the period under study.

Table 2.5: TFP Gains from Reallocation using Value Added (%)

	1987-88	1994-95	2000-01	2004-05
All	-			110.9
Formal	100.4(HK)	127.5(HK)	-	142.3
Informal	-	97.2	58.4	86.2

One possible explanation for the increased misallocation reported in Table 2.3 is due to the linkages between the formal and informal sectors, the labor market linkages in particular. In Table 2.1 we reported that the formal sector had a negative 3% growth in employment in the period 1994-95 to 2004-05. The informal sector saw 25% increase in employment in the same period. If we break up the period into two halfs we find that the formal sector had a 12% increase in the first half and 13. 2% negative growth in the second half. The informal sector on the other hand saw 24% increase in employment in the first half and 1% increase in employment in the second half. If we look at the increase in the number of informal firms we find that there was a 1.2% increase in the second half of the period under study. Because the informal sector is characterized by small firms and self employment it quite is possible that small firms or workers who lost jobs in the formal sector moved into the informal sector during this period. This combined with the fact that most of the distortionary policies of the government are at targeted towards large firms could perhaps explain the increase in the ratio of the 90th and 10th percentiles in the formal sector in 2004-05. Some of these distortionary policies are excise tax credit to small firms, reservation of products for small firms, compulsory providend fund contributions if a firm has more than 20 workers, etc. A dyanamic open economy model with linkages between the formal and informal sectors could perhaps better explain the trends that we find in our results.

We now look at the gains to reallocation by eliminating non symmetric distortions within an industry. We compare current output to the scenario where there are no firm level distortions. That is we calculate $\left(\frac{Y_{efficient}}{Y} - 1 \right) 100$. In the absence of intermediate goods aggregation is easy and we have the following simple expression for Y.

$$Y = \prod_{s=1}^{S} \left(TFP_s K_s^{\alpha_s} H_s^{1-\alpha_s} \right)^{\theta_s}$$

Note that in the absence of non symmetric distortions (efficiency case) the industry TPP is not affected by misallocation at the firm level. TFP_s in the efficient case is therefore equal to $\left[\sum_{i=1}^{N_s} A_{si}^{\rho-1} \right]^{\frac{1}{\rho-1}}$.

Gains from equalizing TFPR of firms with an industry are given in Table 2.5. For the entire manufacturing sector elimination idiosyncratic distortions would have boosted TFP by 110.9%. For the formal sector the reallocation gains for 1987-88 and 1994-95 are from Hsieh-Klenow. We calculate the gains for the year 2004-05. According to the estimates for the latest years in our data eliminating firm level distortions would boost (manufacturing) output by 86.2% in the informal sector and 142.3% in the formal sector. The statistics reported by Hsieh-Klenow for the formal manufacturing sector are 100.4% for 1987-88 and 127.5% for 1994-95. From our results it appears that reforms have not reduced misallocation in the formal sector at all. On the contrary the gap between actual and efficient output has increased from 100.4% in 1987-88 to 142.3% in 2004-05. Hsieh and Klenow had found a similar trend for the period 1987-88 to 1994-95. Misallocation in the informal sector on the other hand has reduced slightly. The informal sector alloccations are however are far from being efficient.

If government policies affect TFPR then we should have seen a decrease rather than an increase in TFPR dispersion during the reforms period. There were several reforms initiated in India during the period 1994-95 to 2004-05. There were reductions in import tariffs, rationalization of excise taxes and custom duties, delicensing of industries, dereservation of products which were previously reserved for the small scale industry etc. All of these policies should have reduced idiosyncratic distortions and therefore reduced resource misallocation. There are three possible explanations for why we do not see any reduction in misallocation measured by dispersion in TFPR. One, misallocation due to other sources besides government policy such as financial frictions or changes in demand structures increased. Two, reforms may not have been of sufficient degree. The best way to describe the reforms process in India would be piecemeal and gradual. The reforms that we mention above were not all implemented at the same time. A bunch of reforms especially related to the external economy were initiated in 1991. Subsequently reforms were initiated at a gradual pace. Moreover several of the distortionary policies especially with regard to labor and small firms still persist. According to Arnold et al. (2010) reforms in manufacturing have to go hand in hand with reforms in manufacturing to improve manufacturing productivity. Aghion et al. (2008) find the delicensing policy had unequal effects in different states in India. Delicensing had positive effects on output in states that initiated labor reforms. The third reason could be that our model is misspecified. One of the drawbacks of using the value added approach to calculate misallocation at the industry level is that it ignores the possible distortions due to intermediate inputs. Jorgenson and Stiroh (2000) have emphasized the role of intermediate inputs in correctly measuring TFP. Jones (2011a) and Jones (2011b) show how distortions to intermediate inputs can have an amplifying effect depending on the input-output structure of industries. We have already shown how the inclusion of intermediate goods can affect allocations and total output in our model. In the next section we recalculate the measures of log TFPR and gains from reallocation with the inclusion of intermediate goods. We find that including intermediate goods is quite crucial in the measurement of resource misallocation.

Table 2.6: Dispersion of log TFPR using Gross Output: 2004-05

	All	Formal	Informal
stdev	0.49	0.46	0.43
75-25	0.56	0.50	0.51
90-10	1.19	1.06	1.07

Table 2.7: Dispersion of log TFPR using Gross Output: Informal

	1994-95	2000-01	2004-05
stdev	0.53	0.40	0.43
75-25	0.60	0.48	0.51
90-10	1.28	0.97	1.07

Table 2.8: TFP Gains from Reallocation using Gross Output (%)

	1994-95	2000-01	2004-05
All	-	-	28.7
Formal	-	-	25.0
Informal	35.5	27.3	35.6

2.7.2   Gross Output Approach

We now repeat the exercise that we carried out in section 2.7.2 using the gross output approach described by our model. Figures 2.9 and 2.10 plot the distribution of log TFPR calculated using gross output for the formal and informal sectors. Table 2.6 shows the dispersion in log TFPR for the entire manufacturing sector and for the formal and informal sectors using gross output measures for 2004-05. Table 2.7 shows the dispersion in log TFPR for the informal sector for the years 1994-95, 2000-01 and 2004-05. Table 2.8 presents the gains from reallocation to efficiency case for the year 2004-05 for the entire manufacturing sector and for the years 1994-95, 2000-01 and 2005-06 in the informal sector . The main findings are noted below.

1. The amount of misallocation when measured using gross output is much lower than the value added measures. This is much more so in the formal sector. The standard deviation of log TFPR is 40% lower in 2004-05 when intermediate inputs are included on TFPR calculations. The ratio of the 75th-25th percentiles is 1.7 in level terms and the ratio of the 90th-10th percentile is 2.9 in level terms in the formal sector. Misallocation in the formal sector seems only slightly higher than the informal sector.

2. Misallocation in the informal sector seems to behave in the same way as in the case of value added. There is a slight reduction in misallocation during the period 1994-05 to 2005-06. The trend of misallocation decreasing in the first half of the period and then increasing in the second half is seen even here.

3. The gains from reallocation for the entire manufacturing sector are in the range of 25%-35%. In fact although the formal and informal sector display similar degrees of misallocation, the gains from reallocation are much higher for the informal sector.

In the next section we try to reconcile the results that we obtained using the value added approach as in Hsieh and Klenow(2009) and the gross output approach.

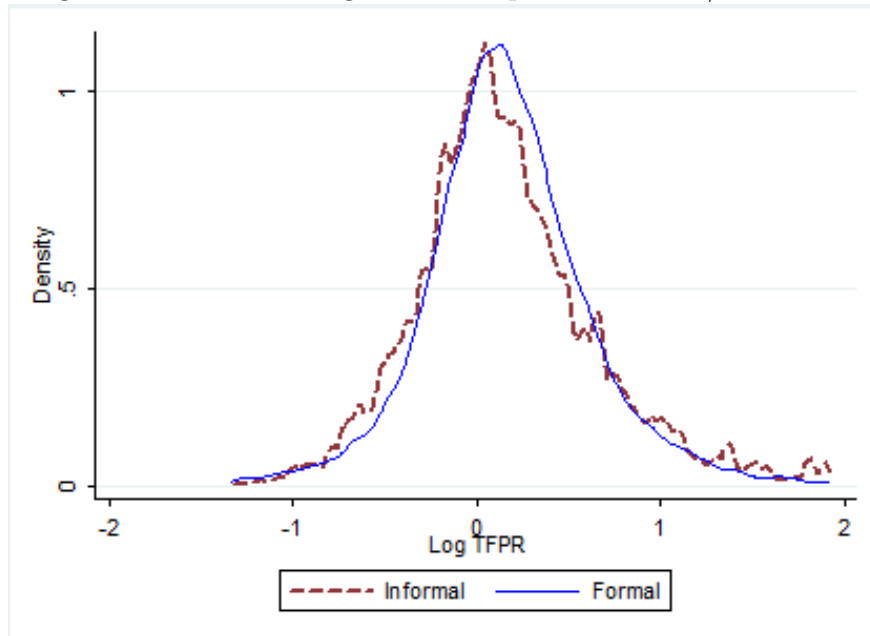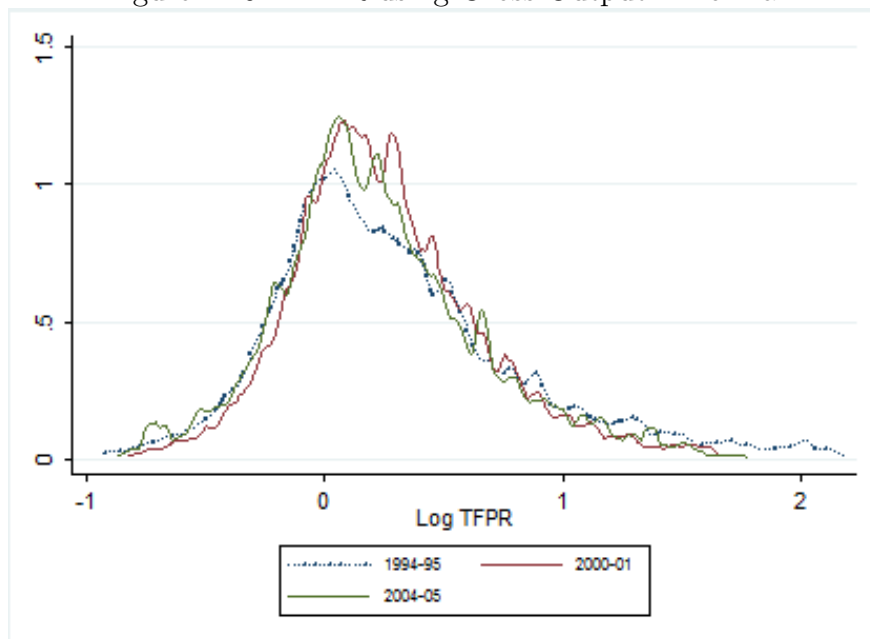Figure 2.9: TFPR using Gross Output: Formal V/s Informal

Figure 2.10: TFPR using Gross Output: Informal

2.8 Value Added V/s Gross Output

The value added approach of measuring misallocation yielded much higher estimates of misallocation than the gross output approach. The standard deviation of log TFPR for the entire manufacturing sector in 2004-05 using gross output was 35% lower than the standard deviation calculated using value added. The misallocation in the formal sector relative to the informal sector appears to be much smaller when gross output is used. The gains from reallocation appear to much smaller in the formal sector despite misallocation being slightly bigger than the informal sector. What is the explanation behind the differences in magnitude of misallocation that we obtain under the two approaches?

Value added, while useful is calculating GDP ,is not always the best way to measure productivity. The most attractive feature of value added is that when aggregated across firms and industries it is equal to GDP. It is however not useful to study productivity as it assumes technical change to be the same across all industries. According to Jorgenson and Stiroh (2000), by correctly accounting for intermediate inputs the gross output approach allows TFP gains to be better allocated across industries. According to Jones (2011a), intermediate goods matter because the chain is as strong as its weakest link. The chain here refers to the production process and the links to intermediate inputs. Including intermediate inputs can affect misallocation in two ways. One, intermediate inputs may themselves be subject to distortions. Two, intermediate goods serve as a conduit to transport distortions across industries. According to Basu and Fernald (1997) when used as a measure of output to calculate productivity the value added approach suffers from omitted variable bias unless there is perfect competition and the elasticity of substitution between intermediate inputs and other inputs in production is zero. Empirical studies have rejected the notion of perfect competition. Broda and Weinstein (2006) find that the estimates of elasticity of substitution are typically in the range of three to ten, thus, indicating positive profits. Basu and Fernald show that the omitted variable bias that occurs when one uses value added to measure productivity in an imperfect competition setting is dependent on the intermediate inputs intensity and its covariance with other inputs. Value added measures of productivity are unbiased if the ratio of intermediate inputs to output is constant and is orthogonal to other inputs usage.

In order to fully understand the role of intermediate inputs we need to look at distortions to factor inputs individually and the variance-covariance matrix of distortions. The measure of misallocation in Hsieh and Klenow (2009) is a composite of output and capital distortions. Expressing log TFPR in value added terms and taking logs yields the following expression.

$$logTFPR_{si}^{VA} = log(constant_s^{VA}) + \alpha_s log(1 + \tau_{ksi}) - log(1 - \tau_{ysi}) \qquad (2.45)$$

Similarly, the expression for log TFPR using gross output is

$$logTFPR_{si}^{GO} = log(constant_s^{GO}) + \alpha_s(1 - \sigma_s)log(1 + \tau_{ksi}) + \sigma_s log(1 + \tau_{msi}) - log(1 - \tau_{ysi}) \qquad (2.46)$$

Figure 2.11: Distortions to Factor Inputs using Value Added
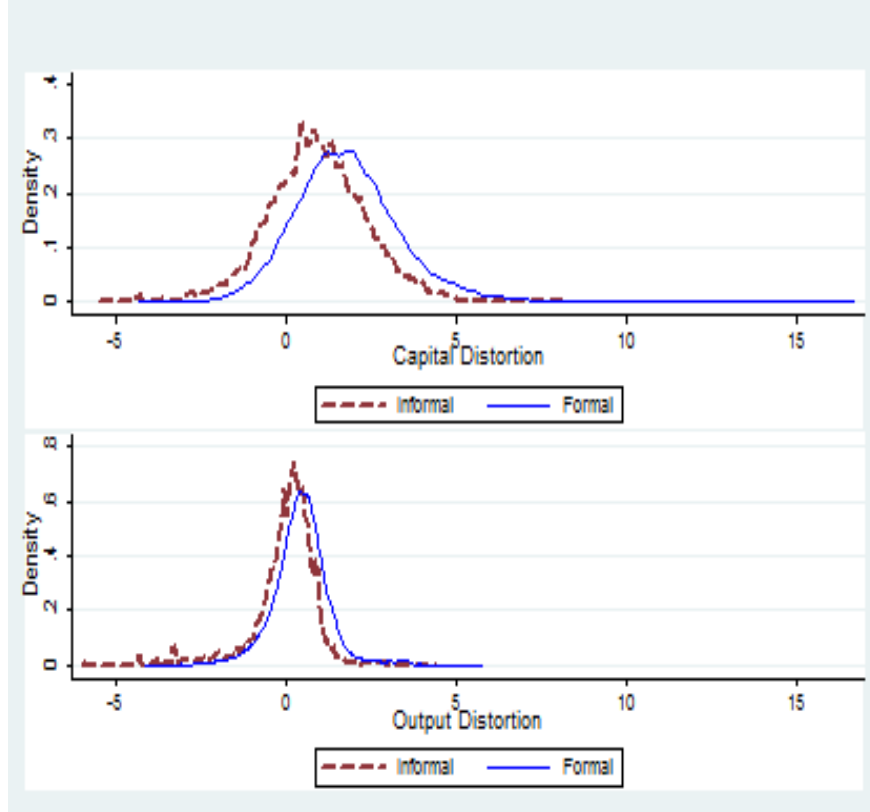
Equations (2.45) and (2.46) measure the log of revenue productivity using value added and gross output approaches respectively. In these equations the slope parameter that measures the contribution of individual factor input distortions to overall distortions is equal to the corresponding input factor share. Notice that the inclusion of intermediate inputs suppresses the share of capital distortions in log TFPR from α_s to $\alpha_s(1 - \sigma_s)$. The median value of σ_s is the NBER Productivity Database is around 0.5 and the value of α_s is around 0.25. Therefore, with the inclusion of intermediate inputs the weight on capital distortions reduces from 0.25 to 0.125. Figures 2.11 and 2.12 plot the distortions to factor inputs for the formal and informal sectors. Figure 2.11 plots the distortions to capital and output in the value added approach and Figure 2.12 plots the distortions to capital, output and intermediate inputs in the gross output approach. The expressions for the distortions, $(1 + t_{ksi})$, $(1 + t_{msi})$ and $(1 - t_{ysi})$, are derived from the first order conditions of the firm's profit maximization problem given by equations (2.20)-(2.22).

Figure 2.12: Distortions to Factor Inputs using Gross Output
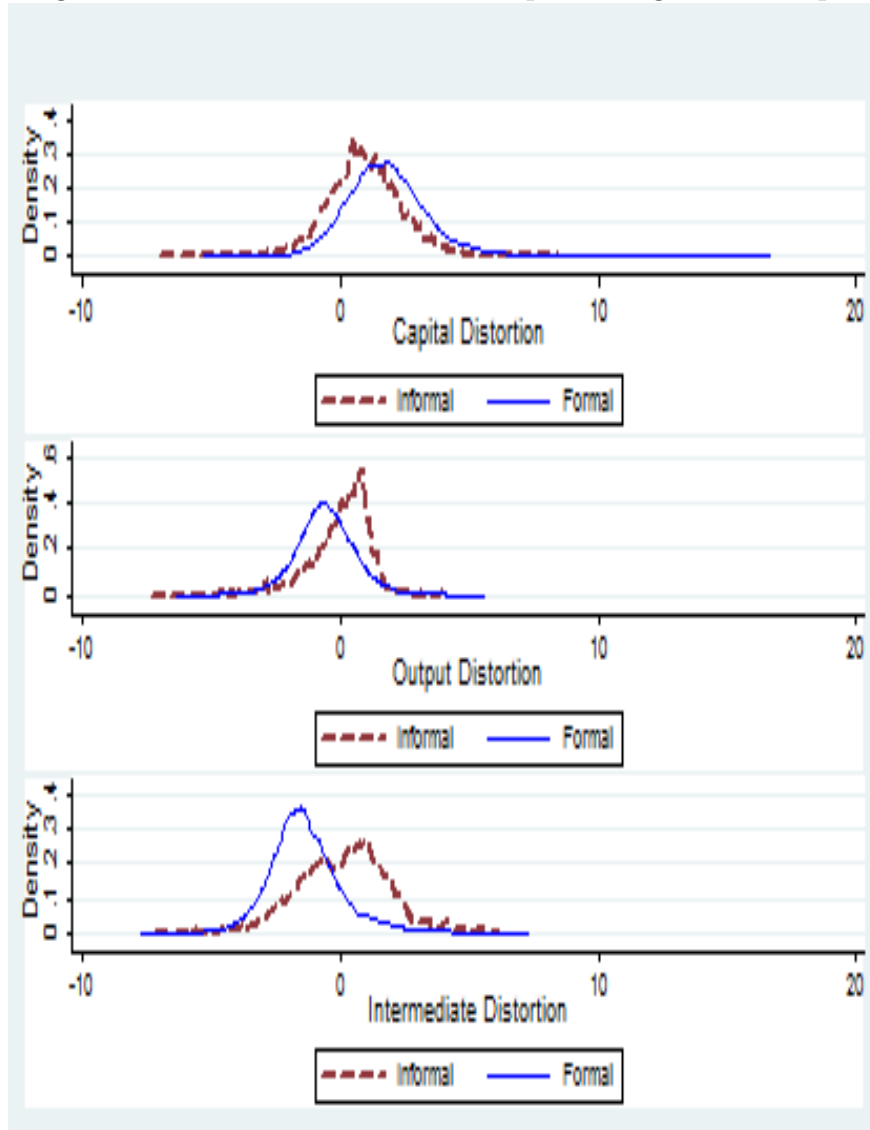
Table 2.9: Factor Input Distortions 2004-05: Value Added Approach

	All	Formal	Informal
var(1-τ_{ysi})	0.77	0.64	0.77
var(1+τ_{ksi})	2.25	2.56	2.22
Covar($(1+\tau_{ksi}),(1-\tau_{ysi})$)	0.81	0.81	0.71

For the purpose of measuring misallocation however we need to look at the dispersion in log TFPR. In other words, we need to look at the variance and covariance of individual distortions. Equation (2.48) below gives the expression for the variance of log TFPR using value added. Whereas equation (2.50) gives the expression for the variance of log TFPR using gross output. Table 2.9 shows the variances and covariances of individual distortions in the formal and informal sectors using value added for the year 2004-05. Table 2.10 does the same using gross output. Figure 2.13 shows the covariation of individual distortion using the gross output. We summarize the main findings below.

$$
\begin{aligned}
Var(logTFPR_{si}^{VA}) \;=\; & \alpha_s{}^2 Var(log(1+\tau_{ksi})) + Var(log(1-\tau_{ysi})) - \\
& 2\alpha_s Cov((1+\tau_{ksi}),(1-\tau_{ysi}))
\end{aligned}
$$

$$(2.48)$$

$$
\begin{aligned}
Var(logTFPR_{si}^{GO}) \;=\; & \alpha_s{}^2(1-\sigma_s)^2 Var(log(1+\tau_{ksi})) + \sigma_s{}^2 Var(log(1+\tau_{msi})) + \\
& Var(log(1-\tau_{ysi})) + 2\alpha s(1-\sigma_s)\sigma_s Cov((1+\tau_{ksi}),(1+\tau_{msi})) - \\
& 2\alpha_s(1-\sigma_s)Cov((1+\tau_{ksi}),(1-\tau_{ysi})) - 2\sigma_s Cov((1+\tau_{ksi}),(1-\tau_{ysi}))
\end{aligned}
$$

$$(2.50)$$

Figure 2.13: Covariance of Distortions to Factor Inputs
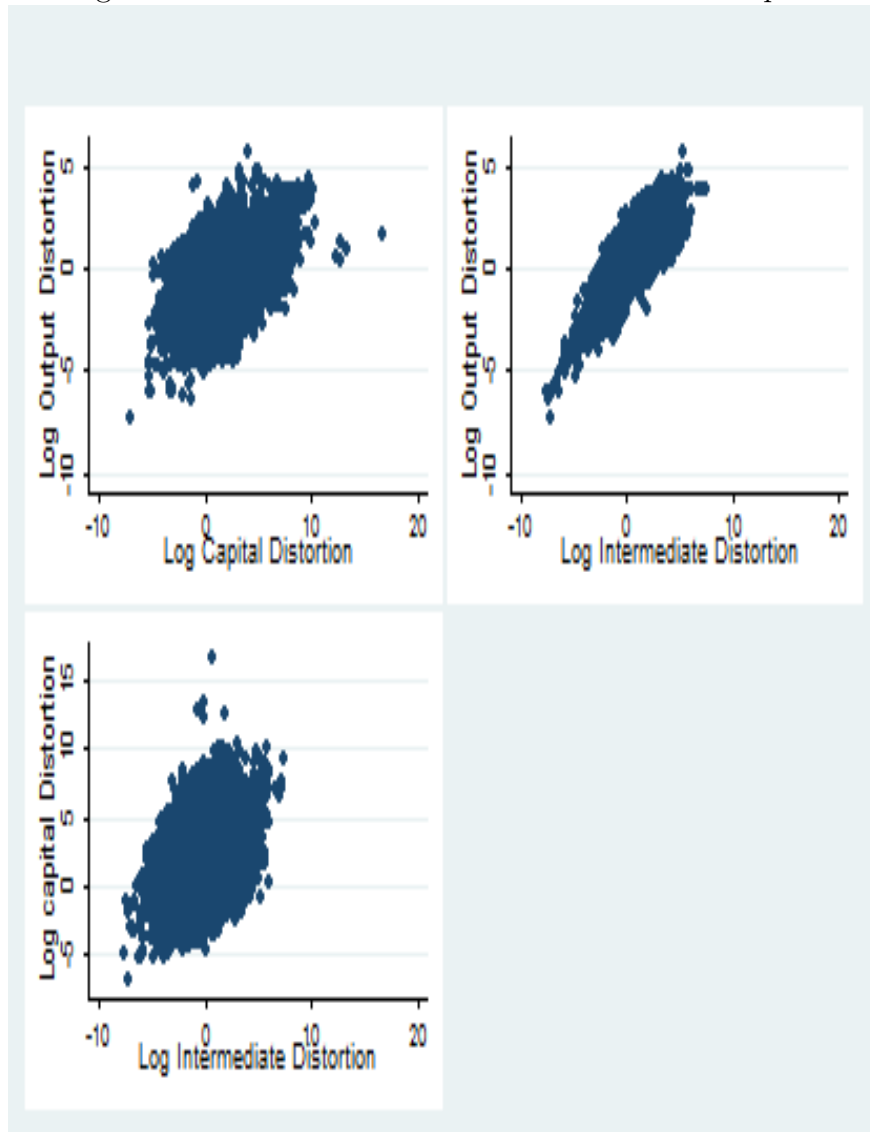
Table 2.10: Factor Input Distortions 2004-05: Gross Output Approach

	All	Formal	Informal
var($1-\tau_{ysi}$)	1.12	1.23	1.11
var($1+\tau_{ksi}$)	2.22	2.62	2.19
var($1+\tau_{msi}$)	2.79	1.86	2.72
Covar($(1-\tau_{ysi})),(1+\tau_{msi})$	1.56	1.43	1.52
Covar($(1+\tau_{ksi}),(1+\tau_{msi})$)	0.75	0.97	0.80
Covar($(1+\tau_{ksi}),(1-\tau_{ysi})$)	0.72	0.90	0.74

The distortions to intermediate inputs is the largest followed by distortions to capital and output in both the formal and informal sectors. This is measured by the variance of the individual distortions. The distortions to output reflect distortions that affect the marginal products of all factors in the same way. This explains the negative sign for the covariances between distortions and distortions to capital and intermediate inputs. The covariance between distortions to output and intermediate goods is quite high compared to the other covariances. A high covariance between distortions to output and intermediate goods can be explained by the amplifying effect that the inclusion of intermediate goods typically produces. Distortions to output affect resource allocation directly and indirectly via intermediate inputs. Distortions to output are carried across firms and industries via intermediate inputs. The negative sign in the covariance is precisely to avoid double counting of these distortions. A large covariance between output and intermediate distortions would therefore suppress the effect of large intermediate distortions in the variance of log TFPR.

Distortions to capital are also quite high in both formal and informal sectors. Capital distortions are however more in the formal sector compared to the informal sector. A firm in the informal sector typically borrows from the money lender. The money lender charges a high interest rate due to monopoly power but he does so for most of his borrowers. Formal firms however can obtain finance from different sources at different rates. They can avail bank or other institutional loans, raise money in equity and debt markets etc. Banks in India lend to small firms, state owned firms and priority sectors at a concessional rate. The formal credit market in India appears more segmented than the informal credit market thus suggesting greater degree of idiosyncrasy in distortions to formal sector capital. This is a plausible explanation for distortions to capital to be bigger in the formal sector. Distortions to capital have a lower weight in the gross output approach compared to the value added approach for reasons explained earlier.

The size of the distortions to individual factor inputs, their interaction with other distortions and the shares of factor inputs are all important in correctly identifying the source and magnitude of overall misallocation. The model based on the value added approach fails to properly account for the role played by intermediate inputs and the distortions to intermediate inputs and is therefore misspecified. The model based on gross output although superior in its specification could lead us to the wrong conclusions about misallocation if we do not decompose the overall misallocation and study the distortions individually as in Tables 2.9 and 2.10. To summarize, a model of misallocation should incorporate intermediate inputs and should study the role of distortions to capital, output and intermediate goods individually for proper measurement and identification of the source of overall misallocation. Equipped with this knowledge we look at some of the potential sources of distortion in the next section.

2.9 Sources of Resource Misallocation

The existing literature on misallocation suggests substantial gains from reallocation of resources an efficient level. Our study confirms this finding. Using the gross output approach to measure misallocation we find that there can be TFP gains between 25% to 35% by just removing the indiosyncratic distortions to firms within an industry. The difficult question that is not fully answered yet in the literature is with regard to the source of distortions. In the previous section we decomposed our composite measure of misallocation, log TFPR, to look at individual distortions. We found distortions to capital and intermediate inputs to be quite high. Our endevour in this section is to exploit the rich data on the informal sector to throw some light on the sources of these distortions. The informal sector survey for India asks firms questions on the shortage of capital, infrastructure bottlenecks and marketing problems. Table 2.11 shows the correlations betwen log TFPR and the above mentioned qualitative variables. The qualitative variable is a dummy that records 1 if a firm answers yes to the question whether it faces a particular problem and 0 if the answer is no. The first column of the table shows the percentage of firms facing a particular problem. The second column gives you the regression coefficient of regression log TFPR, which is calculated using the gross output approach, on a particular problem. All the regression results are statistically significant.

According to our regression results, the shortage of capital is positively correlated with log TFPR. The shortage of capital can be viewed as a distortion to capital. Banerjee and Duflo (2005) have shown that distortions in credit policy led to substantial gaps in marginal product of capital and lower TFP in India. Lack of an electricity connection and non availability of raw materials can be viewed as distortions to intermediate inputs. lack of an electricity connection and non availability of raw materials are also positively related to log TFPR. Lack of proper electrification forces many entrepreneurs in India to buy power at a premium rate from private electricity providers or invest in a in house power generator Bhide (2008). Some state run companies get electricity at subsidized rates. Whereas farmers in some states get electricity for free.

Log TFPR is negatively correlated with competition from larger units and marketing problems. This could imply that smaller firms that find it harder to market their output face fewer distortions

Table 2.11: Sources of Distortion in Informal Sector: 2004-05

Problem faced	Firms facing problems (%)	Regression coefficient
Shortage of Capital	46.4	0.08
No Electricity Connection	6.3	0.11
No Raw Material Availability	9.6	0.01
Competition from larger units	24.3	-0.02
Marketing Problems	17.4	-0.003

compared to larger firms. This is line with the findings of Hsieh and Klenow (2009) that misallocation is positively correlated with firm size in India. We could also view marketing problems and competition from larger units as proxies for markups. Markup is the inverse of elasticity of substitution in our model and assumed to be the same for all firms. Given that the barriers to entry are low in the informal sector and the enormous size of the informal sector, it might be the case that the elasticity of substitution for the informal firms is much higher than the formal firms. A higher elasticity of substitution would imply a lower markup and lower log TFPR. Although the regression coefficients are small in magnitude they are useful in identifying the potential sources of resource misallocation.

2.10 Conclusion

In this chapter we try to fill some of the gaps in the study of resource misallocation in manufacturing in India. We first measure misallocation for the entire manufacturing sector in India by incorporating the informal sector in our analysis. Earlier work by Hsieh and Klenow (2009) included the study of the formal manufacturing sector in India alone. The presence of the informal sector in manufacturing in India is huge, when viewed especially in terms of employment generation. By incorporating the informal sector in our study we are able to correctly measure misallocation for the entire manufacturing sector in India. We build a model of monopolistic competition with intermediate goods to measure misallocation. We show that in order to accurately estimate productivity or misallocation, especially at the industry level, we need to include intermediate goods. Using data on the informal manufacturing sector in India for the years 1994-95, 2000-01 and 2005-06 and data on the formal manufacturing sector for 2004-05 we first replicate the methodology adopted by Hsieh and Klenow to measure distortions. Using value added estimates we find that misallocation had increased in the formal sector for the period 1994-95 to 2004-05 most of which came from the increase in the 90/10 percentiles of log TFPR. Efficient reallocation would result in nearly 111% TFP gains. We also find that the misallocation in the formal sector to be larger compared to the informal sector for the period under study suggesting that the formal firms face larger distortions than informal firms. This result however is not very robust and is sensitive to the methodology used to measure productivity. Using gross output estimates we find misallocation to be much lower. We also find very little difference in misallocation in the formal and informal sectors. In fact we find that the formal sector has lesser efficiency gains from reallocation than the formal sector.

In order to understand the differences in the value added and gross output measures of misallocation we look at the decomposition of our misallocation measure, log TFPR. We find that the value added measures suffer from omitted variable bias as they ignore the distortions to intermediate inputs. We show that measuring the distortions to factors individually and looking at the variance-covariance matrix of distortions is important to understand the source and magnitude of this bias. Finally, we try to identify the sources of misallocation exploiting qualitative information about the informal firms. We find misallocation to be positively correlated with the shortage of capital and non availability of intermediate inputs like electricity and raw materials. We also find that misallocation is negatively correlated with the market size of the firm, indicating that larger firms face more distortions. Future research in the area needs to look at some of these sources of misallocation in more depth and quantify the amount of distortion by caused by each of them.

Chapter 3

Tax Threshold and Bunching: Evidence from Indian Firms

3.1   Introduction

The existence of a large number of small firms is a characteristic of both developed and developing countries. What is however striking is the proportion of small firms in developing countries. Table 3.1 provides the size distribution (employment) of firms in the US and India. 95.2% of the firms in the manufacturing sector in India have less than five employees compared to 47.7% of the firms in the US[1]. This raises two interesting questions. One, why are there so many small firms in India? Two, what are the macroeconomic implications of having so many small firms? What determines the size of a firm has been a subject of wide research. One of the earliest contributions in answering this question was by Viner (1932) who emphasized on the shape of the production function and predicted that a singular firm size distribution would occur in every industry. This view has been challenged by empirical studies that find dispersion in cross-sectional distribution in firm size. Most modern theories attribute firm size to efficiency. A seminal paper by Lucas (1978) showed how both small and large firms can co exist. According to Lucas the size of a firm is determined by the ability of the entrepreneur, with more able entrepreneurs optimally choosing a larger scale of operation. Differences in innate ability may explain differences in firm size distribution within an economy but cannot explain the differences in the firm size distribution across economies. There is no reason to believe that the distribution of innate abilities is different in India and the US.

A more plausible explanation for cross country differences in firm size distribution lies in understanding the role played by institutions and government policy. Efficiency is not just a function of innate ability but also a function of the environment that comprises of institutions such as legal

---

[1]US figures are for all firms and is from the US Census Bureau for 2008. India figures for formal manufacturing from Annual Survey of Industries (2005) and for informal manufacturing from National Sample Survey Organization (2005)

Table 3.1: Firm Size Distribution in US and India

Firm Size (Employment)	US Firms Share(%)	India Firms Share(%)
1-4	47.7	95.2
5-9	13.9	3.5
10-19	8.8	0.9
20-99	9.3	0.4
100-500	4.7	0.1
500-	15.6	0.03

systems, credit and labor markets, regulation and government policy. In recent years there has been a tremendous interest in studying the role of misallocation of resources due to distortionary government policies leading to inefficient firm size distribution and low TFP levels. Studies by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) show that there are substantial TFP gains by eliminating these distortions and moving to a more efficient firm size distribution. We present the findings of these and related papers in detail in the literature review section of this chapter.

In this chapter we study the potential gains from eliminating a specific size dependent policy in India. A firm in India can get up to 100% exemption from paying excise tax, which is the tax on production, if its annual turnover is less than Rs. 10 million.[2] We believe that this exemption encourages firms to operate at a sub-optimal scale. Besides discouraging growth, such a policy also increases the burden of taxation on larger and more productive firms. Small firms in India are granted all sorts of concessions, from paying taxes to subsidized credit. In an interesting study Bhide (2008) compares high growth small firms in the US with their counterparts in India. He finds that the proportion of expanding small firms in India is much lower compared to the US. Interviews with 100 successful entreprenuers in Bangalore, the city known as the Silicon Valley of Asia, revealed that only half of the firms anticipated future growth but more than 80% planned to start or own another business. Many of these businesses were in unrelated fields. For example, one of the entrepreneurs who was in road construction wanted to start a business to sell tomato ketchup. When asked why he would not want to expand his business in road construction itself he replied that he did not want to deal with tax officials and that growing would mean hiring more workers and providing worker welfare. The ability to evade taxes and avoid paying bribes to tax officials was a common response from many of the entrepreneurs. It appeared that avoiding the payment

---

[2]A small scale firm with an annual turnover of less than Rs. 10 million gets 100% exemption from excise tax if it does not claim credit for tax paid on inputs and gets 40% exemption if it claims tax credit on inputs.

of excise tax, which is around 16% on average, allowed smaller firms to compete with larger, tax paying firms that enjoyed economies of scale in similar businesses. The cost of paying the excise tax was not just lower profit margins but also the cost inolved in dealing with tax officials, complicated tax laws and tax filing procedures.

There are two main contributions in this chapter. First, we present strong evidence of bunching of firms near the threshold of Rs. 10 million. That is, evidence that firms are operating at a sub-optimal scale due to the presence of the excise tax exemption. Two, we measure the magnitude of distortion caused by such a tax exemption policy and measure the potential gains from moving to a flat tax regime where all firms pay a uniform tax. Using Annual Survey of Industries data from India for 2004-05 we find strong graphical evidence of bunching around the threshold output level of Rs.10 million for small scale firms. We believe that the excise tax exemption makes firms at the margin of Rs 10 million restrict their output to avail the tax exemption. We develop a model where firms are heterogeneous in productivity and face decreasing returns to scale to describe this bunching behavior. We determine the extent of bunching by using a methodology similar to Saez (2010) who analyzed bunching at kink points of the US income tax schedule generated by jumps in marginal tax rates by looking at graphical evidence. We use maximum likelihood estimation to estimate the parameters of a counterfactual distribution of output without any tax distortion. We estimate the number of bunching firms, the undistorted counterfactual output levels and a flat tax rate that is revenue neutral.

We find the relationship between the scale parameter and output gains to be non linear, with the output gains exploding when the scale parameter approaches one. Moving to a flat tax regime, where are all firms are taxed at a uniform rate, can lead to welfare gains in the range of 13% to 230% using the values of the scale parameter in the literature. We then estimate the scale parameter using the output of the highest bunching firm. We find that the amount of bunching in the data is far less compared to the amount of bunching predicted by our model using values of the scale parameter found in the literature. As a robustness check we repeat the above exercise of constructing the counterfactual distribution of undistorted output by using non-linear least squares estimation and find similar results.

The rest of this chapter is organized as follows. In section 3.2 we present a review of related literature. In section 3.3 we describe our data set and present descriptive statistics. In section 3.4 we provide a brief description of the excise tax policy in India. In section 3.5 we present graphical evidence for bunching. In section 3.6 we describe bunching behavior using a model with firm level heterogeneity. In section 3.7 we construct our counterfactual distribution of output using maximum likelihood estimation and calculate the potential gains from moving to a flat tax regime. In Section 3.8 we present results from using non-linear least squares estimation method. In section 3.9 we study the relationship between welfare gains and the returns to scale parameter. We conclude with section 3.10.

3.2 Literature Review

Our work is related to the burgeoning literature on the role of distortions in explaining income differences across countries. Some of the recent research in this area includes the work done by Restuccia and Rogerson (2008), who calibrate their model to fit US data and show that firm level distortions in the form of taxes and subsidies can misallocate resources and result in 30-50% fall in TFP. Bartelsman et al. (2009) follow a similar approach using cross country evidence. Buera and Shin (2009) focus on financial frictions in the context of liberalization of capital accounts. Hsieh and Klenow (2009) use plant level manufacturing data from India and China and show that reallocation of resources within industries to the US efficiency levels can boost TFP in India by 40-60% and in China by 30-50%. Our work is perhaps closest to Onji (2009), Guner et al. (2008) and Gollin (2005), who look at policies that distort the distribution of firm size. Onji documents the effects of a tax threshold on firm behavior in the context of the Japanese value added tax (VAT). He finds bunching evidence of large firms masquerading as many small firms to avoid the payment of tax. Onji however does not measure the output loss as a result of the tax threshold. Guner et al calibrate their benchmark economy to the US and show that when firm size is reduced by 20% aggregate output falls by 8.1% and number of small firms increases by 23.5%. They also show that while different size dependent policies (tax versus labor regulation) can quantitatively have similar effects on output the welfare implications may be quite different with higher welfare costs associated with tax on capital versus restrictions on labor use. Gollin (2005) calibrates his model to match frm level data in Ghana and looks at the differential tax treatment of small versus large manufacturing firms and finds substantial effects on output.

The above mentioned chapters provide strong quantitative evidence that resource misallocation due to distortions can be large. They also discuss the possible sources of these distortions such as taxes, adjustment costs, labor laws or financial frictions. The area where evidence is weak is in the quantitative assessment of these distortions. If eliminating distortions can lead to substantial TFP gains in developing countries then it would be important to identify the nature and magnitude of specific distortions. Some recent attempts in this area includes the work of Hsieh and Klenow (2009), who try to relate misallocation measured as dispersion in revenue productivity to the licensing and size related policies of the Indian government. They find very little evidence. Midrigan and Xu (2010) employ a methodology similar to Hsieh and Klenow to plants in Korea to quantitatively assess the role of three distortions, namely a) non-convex capital adjustment costs b) borrowing frictions and c) limited insurance for investment risk. They find that these three distortions account for a very small fraction (10%) of the total misallocation in the data. Banerjee and Duflo (2005) measure misallocation of capital by looking at the specific policy of priority lending by banks at subsidized interest rates to small scale firms in India using data from a public sector bank and 253 firms. They exploit the change in policy in 1998 with regard to the investment limit of small scale firms to be eligible for subsidized credit and show that newly eligible (larger firms) on average received more working capital compared to smaller firms. They estimate that the return to capital of these firms was as high as 94%. Gollin (2005) finds that a tax policy disproportionately penalizing large firms can reduce output by nearly one-half, compared with an alternative policy regime in which all firms face the same taxes and regulatory costs in Ghana.

In the past decade several countries, especially in Eastern Europe and some of the OECD countries, have moved closer to a flat tax regime. Gorodnichenko et al. (2007) show that a flat tax regime led to lower tax evasion and higher tax revenues in Russia. Mankiw (2009) find that a flat tax regime is close to optimal when productivity is distributed lognormally. There have been several reforms to excise tax that have taken place in India in the past decade. In 1999 there were as many as 11 different rates of excise tax, in addition to special duties and exemptions. By 2005 a convergence to a uniform tax rate of 16% was sought in order to increase tax compliance and reduce administrative costs. The exemption of excise tax for small firms has nevertheless continued. In this chapter we show that the excise tax exemption policy is distortionary as it encourages firms to remain small, leading to output losses and increasing the tax burden of larger and more productive firms

Our empirical strategy is similar to the one developed in Saez (2010). The approach uses kinks in the marginal tax rate schedule to identify behavioral responses to taxation. Saez's analysis first demonstrates in a stylized setting that given smooth and convex preferences, no individual should have a taxable income level immediately above a point at which the marginal tax rate jumps. The model is then extended to account for the fact that we do observe taxpayers reporting incomes immediately above kink points. His empirical analysis reveals substantial bunching at the first kink in the US marginal tax rate schedule for self-reported income, but no evidence of bunching at higher kinks. The measured elasticities implied by the bunching behavior are somewhat lower than those found by other natural experiment approaches investigating the effects on taxes on labor supply. Chetty et al. (2009) also investigate bunching behavior at various kink points of the Danish individual income tax schedule. They find that bunching behavior increases with the size of the jump in marginal tax rate. To explain this finding they propose an equilibrium model where workers face jobs with specified hours requirements and search frictions. They conclude that the underlying structural elasticity can be higher than observed behavioral responses. This insight could potentially explain the small return to scale implied by the bunching behavior of Indians firms with respect to the excise threshold we find in this study.

3.3 Data and Descriptive Statistics

The Indian manufacturing sector is made up of two kinds of firms, formal and informal. Formal firms are firms that are registered under the Factories Act, 1948,[3]. These firms are required to comply with several government rules and regulations with regard to paying taxes, maintaining books of accounts, labor management practices and other aspects of the production process. The informal firms on the other hand are firms that are not registered under any authority. They are essentially very small firms that escape government supervision and regulation. Although the size of the informal sector is huge in Indian manufacturing for the purpose of this study we use data on the formal sector only. The excise tax exemption that we study in this chapter is available only to registered firms. Informal firms by definition are unregistered firms that do not pay any taxes and therefore do not qualify for any tax exemption. This is confirmed by the data as well. Figure 3.1

---

[3]Under the Factories Act 1948, all firms with 10 or more employees with power or 20 or more employees without power need to be registered.

Figure 3.1: Log Output in Formal and Informal Sectors 2004-05
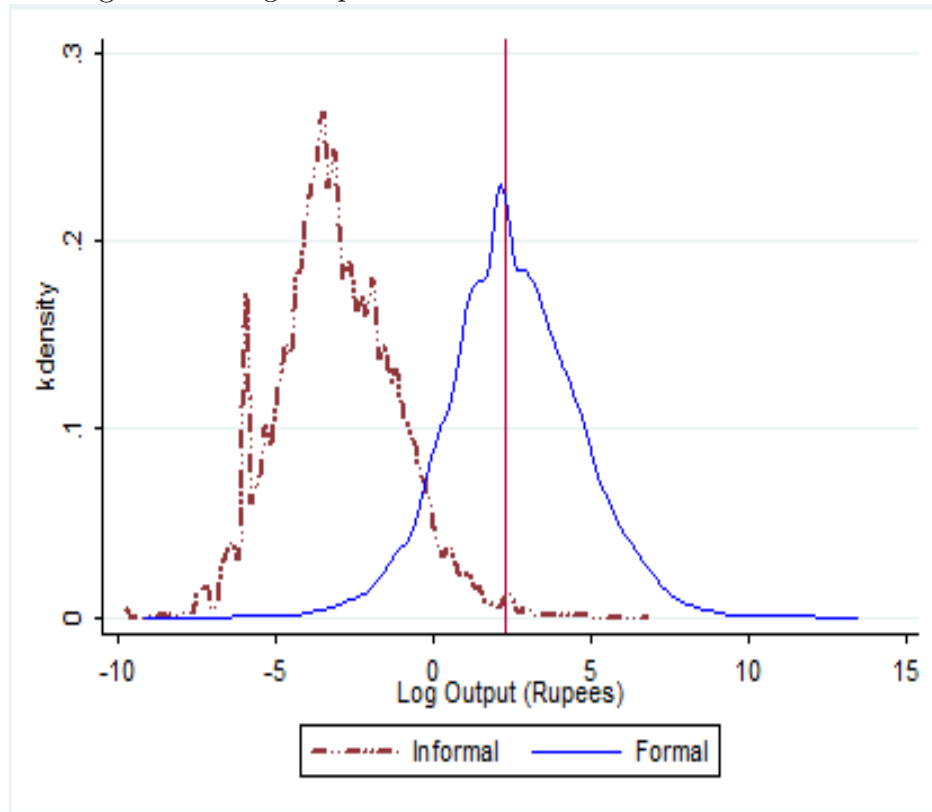
plots the distribution of pre-tax value of log output for the formal and informal sectors. We find evidence of bunching at the threshold of log of Rs. 10 million, indicated by the vertical line in the graph, for the formal sector only.

Our data for the formal manufacturing sector in India is obtained form the Annual Survey of Industries (ASI) compiled by the Central Statistical Organization of India. We use data for the year 2004-05. The data has a sample of 49,340 manufacturing firms and firm level information at the 4-digit industry level. The sample design is a census of all firms with 100 or more workers and a random sample of one-fifth of the remaining registered firms (less than 100 workers). We use sampling weight in our calculations to reflect population estimates. The population estimate of the number of formal firms in manufacturing is 151,523. Output information however is available only for 95,279 firms. All our calculations are based on the population estimate of firms. For the purpose of this chapter we define output as the gross sale value (pre-tax) of total production. This excludes other sources of income such as interest earned or royalties received.

Table 3.2 shows the distribution of firms and output on the basis of total employment in the formal manufacturing sector in India for 2004-05. 23% of the firms have less than 10 employees and account for less than 1% of the output. 13% of the firms employ 100 or more employees but account for almost 80% of the total manufacturing output. Table 3.3 shows the distribution of firms on the basis of output. Firms are either small or big. More than 40% of the firms in Indian manufacturing

Table 3.2: Distribution of Firms based on Employment 2004-05

Employment Size	Firms (no.)	Firms Share(%)	Employment Share(%)	Output Share(%)
0-9	61,613	40.7	2.4	1.0
10-19	32,443	21.4	5.6	3.4
20-49	29,357	19.4	11.3	7.8
50-99	12901	8.5	11.1	8.3
100-499	12809	8.5	32.1	27.0
500-	23400	1.6	37.7	52.5

have output larger than Rs. 500 million and contribute to 75.6% of the total manufacturing output and to 46.3% of the total employment. Firms producing less than Rs. 10 million contribute to only 1% of the total manufacturing output but account for nearly 30% of the total employment in Indian manufacturing.

Small scale industry (SSI) status is granted to a firm on the basis of the original value of plant and machinery of the firm, which need to be less Rs. 10 million. Our data set does not contain information on a firm's SSI status nor does it contain the original value of plant and machinery. We only have information about the value of plant and machinery only for the year 2004-05. Table 3.4 shows the distribution of firms, employment and output on the basis of investment in plant and machinery. We define investment in plant and machinery as the net closing book value of plant and machinery.[4]. According to Table 3.4 there were almost 82% firms in Indian manufacturing whose value of plant and machinery was less than Rs. 10 million. These firms accounted for 45% of total employment and about 19% of total output. Large firms (investment in plant and machinery greater than Rs 10 million), although few in number were important in terms of contribution to total output and total employment as well.

To Summarize, we have looked at the firm size distribution in the formal Indian manufacturing sector on the basis of output, employment and capital. We find that the formal manufacturing sector in India is characterized by the missing middle. Firms are either very small or very large. When measured in terms of capital and labor the majority of the firms in the Indian manufacturing are small. The contribution of small firms to total employment is substantial. Most of these firms are characterized by self employment. However their contribution to total output is very small.

---

[4]The ASI dataset has information on the gross value of plant and machinery, additions and revaluations. This information is however available for only one third of the observations in the sample.We therefore use the closing net value of plant and machinery which is available for 39,417 observations in our sample.

Table 3.3: Distribution of Firms based on Output 2004-05

Output	Firms	Firms	Employment	Output
(Rs.mn)	(no.)	Share(%)	Share(%)	Share(%)
0-9	44,882	29.6	10.2	1.1
10-19	12,159	8.0	4.7	1.1
20-49	14,477	9.6	8.1	3.0
50-99	8,623	5.7	8.3	4.0
100-499	10,931	7.2	22.3	15.2
500-	60,393	39.9	46.3	75.6

Table 3.4: Distribution of Firms, Output and Employment based on Investment 2004-05

Plant	Firms	Firms	Output	Employment
(Rs. mn)	(no.)	Share(%)	Share(%)	Share(%)
0-5	87,769	75.4	14.1	37.2
5-10	7,504	6.4	4.7	7.7
10-15	3,028	2.6	2.7	4.7
15-20	1,982	1.7	2.4	3.0
20-25	1,305	1.1	2.0	2.7
25-50	3,103	2.7	6.4	7.6
50-	11,714	10.1	67.8	37.1

Large firms although fewer in number than small firms are important for their contribution to both total output and total employment.

3.4   Excise Tax Policy for Small Firms

Excise tax was the largest source of tax revenue for the government at 33% of total tax revenue in 2004-05. The excise tax revenue was Rs.1,085 billion in 2004-05 and was equal to 3.3% of India's GDP. This tax burden was borne by only 25% of the firms in the manufacturing sector in India. In this chapter we focus on the concessions that small scale firms are eligible for under the Central Excise Tax, hereafter referred to as the excise tax. A firm normally attains small scale industry (SSI) status if the investment in plant and machinery is within a certain threshold limit.[5]. In the context of excise tax the eligibility for SSI status is decided on the basis of annual turnover rather than investment in plant and machinery. A firm in India pays a basic excise tax of 16% on the value of gross output. All firms irrespective of their investment or number of employees are eligible for up to 100% excise tax exemption in the current year as long the turnover was less than Rs. 30 million in the previous financial year. [6]. There are two types of excise tax exemption schemes available for eligible firms:

- SSI without CENVAT: An SSI firm that does not claim CENVAT (credit for excise tax paid on inputs purchased) can get 100% exemption if annual value of clearances does not exceed Rs.10 million[7]

- SSI with CENVAT: An SSI firm that claims CENVAT pays only 60% of the normal tax rate if annual value of clearances does not exceed Rs. 10 million.

In a nutshell, a firm with annual turnover less than Rs. 10 million can get from 40-100% exemption from paying excise tax. That is, in 2004-05 a firm availing tax exemption paid a basic excise tax rate ranging between 0% and 9.6%. All firms with annual turnover above the limit of Rs. 10 million paid the basic excise tax rate of 16%. In 2004-05, 75% of the firms in Indian manufacturing paid zero excise tax. The median output level of these firms was Rs. 8 million.

---

[5]The threshold limit for investment in plant and machinery,whether held on ownership terms or on lease or on hire purchase, in determining the SSI status of a firm has changed over the years. It was set at Rs.0.5 million in the 1950s and has gradually increased since then. The threshold limit in 2004-05, which is the period under consideration for this study, was Rs. 10 million. This limit has been increased to Rs. 50 million since September 2006. Till 1983 the per-requisite for eligibility for excise concessions was that the SSI unit should be registered with the State Directorate of Industries. In 1994 this requirement was done with and there was no distinction between registered and non-registered SSI unit and the eligibility for excise concessions was henceforth determined by the firm's annual turnover rather than investment

[6]The threshold limit was increased to Rs. 15 million in April 2007

[7]Annual clearances are wholesale prices at the factory gate, exclusive of taxes.

3.5   Bunching Evidence

In order to obtain the excise concession the pre-tax gross value of output (output from now on) of a firm should not exceed Rs 10 million. Figure 3.2 plots the histogram and kernel density of output of of firms whose output does not exceed Rs. 50 million. 75% of the firms in our dataset falls within this range. We do this to eliminate the effect of very large firms and to zoom in on the bunching area. One can clearly see bunching around the threshold of Rs 10 million as indicated by the vertical line. This bunching is more clear when we plot the log of output. Figure 3.3 plots the log of output of all firms. We once again find bunching at log(10). We also find that except for the bunching firms, the log of output is normally distributed. In Figure 3.3 we also find some bunching at log(4). The extent of bunching is much lower at this threshold compared to the threshold of log(10). We believe that this bunching is a result of the policy that propreitory and partnership firms whose output is below Rs. 4 million do not require a statutory audit for the purposes of paying income tax. In addition, firms with output less than Rs. 4 million are not required to maintain proper book of accounts. It is harder for us to measure the effect of such a policy as we do not have adequate information to clearly identify the taxable amount of profit income of such firms.

It is also possible that firms under report the actual level of output. In reality firms may not be restricting output but just under reporting. One way to resolve the problem of under reporting is to check for bunching in inputs, especially intermediate inputs, which are more divisible than labor or capital. Figure 3.4 plots the histogram for intermediate inputs. We do not find any bunching for intermediate inputs near the threshold of log(4) but find significant bunching at the threshold of log(4). We therefore focus only on the bunching due to the excise tax exemption in this chapter. In the next section we present a simple model to explain this bunching behavior by firms.

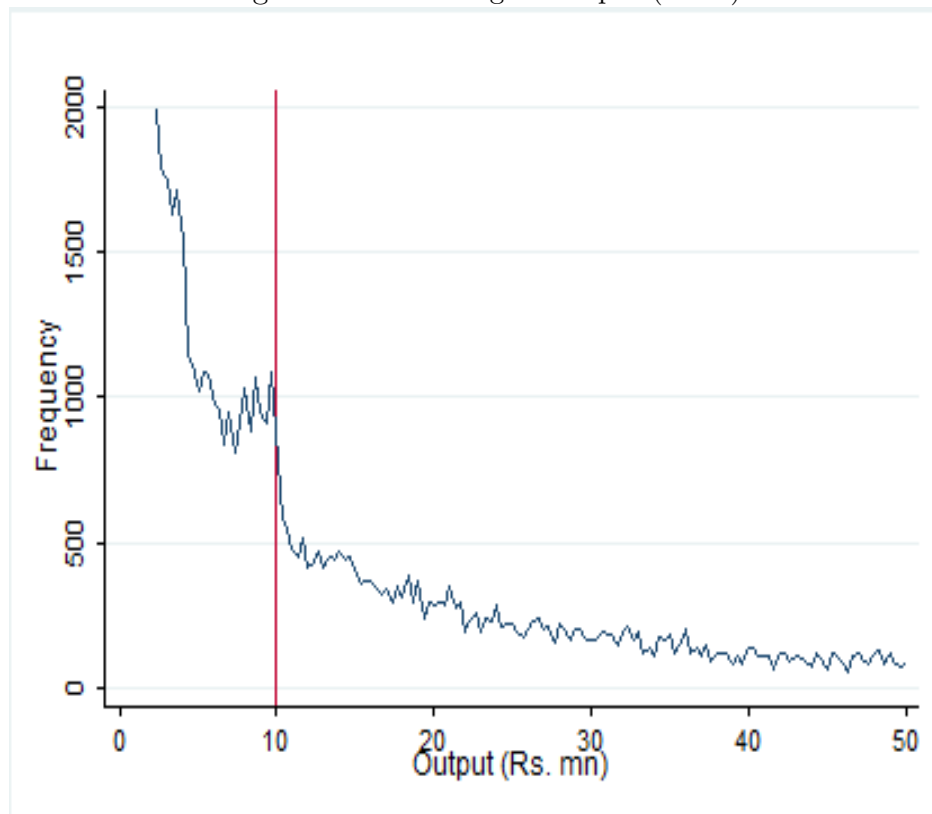Figure 3.2: Bunching in Output (levels)
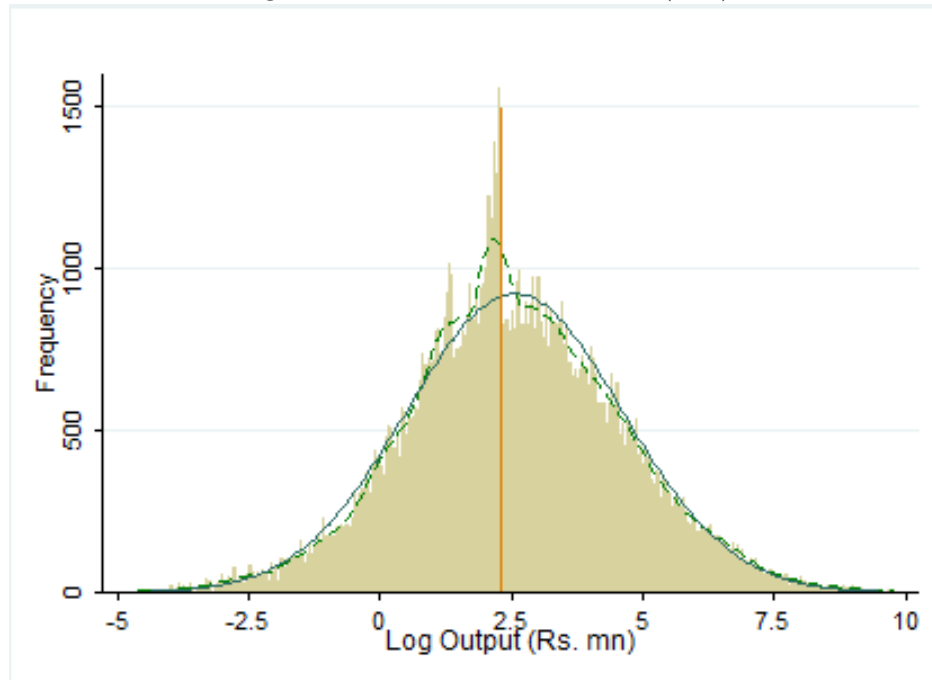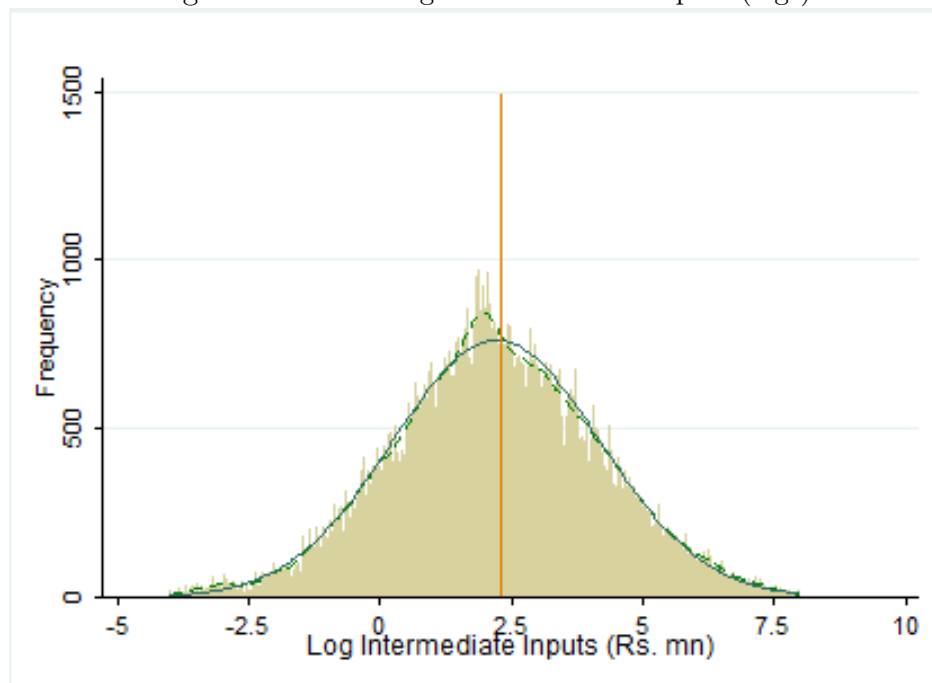
Figure 3.3: Bunching in Output (logs)

Figure 3.4: Bunching in Intermediate Inputs (logs)

3.6   A Model with Firm Level Heterogeneity

Consider a market with firms of different productivity and size which all produce a common good y. Labor and capital are used in production and output for each firm is given by $y = \theta(k^\alpha l^\beta)^\gamma$. Heterogeneity is captured by the parameter θ, which denotes individual managers' quality as entrepreneurs and is distributed according to a cumulative distribution $H(\theta)$ and a density function $h(\theta)$. We assume firms are price takers on both inputs and the output markets. The prices for labor and capital are exogenously given by w and r respectively and the price of y is normalized to one. Finally we assume decreasing returns to scale in the Cobb Douglas production function, i.e. $\alpha + \beta = 1$ and $\gamma < 1$, so that firms with different levels of productivity can coexist in a competitive equilibrium.

To reflect the particular environment faced by firms in our data, we include in the model a size-dependent linear tax on output. Specifically, firms with output higher than a threshold $\overline{y}$ must pay an *ad valorem* tax τ on total receipts. Firms that produce less than $\overline{y}$ do not have to pay the tax. This gives rise to the following profit maximization problem.

$$\prod(0; \theta) = \max_{k,l} \theta(k^\alpha l^\beta)^\gamma - rk - wl \tag{3.1}$$

if a firm's output is lower or equal to the threshold $\overline{y}$ and

$$\prod(1 - \tau; \theta) = \max_{k,l}(1 - \tau)\theta(k^\alpha l^\beta)^\gamma - rk - wl \tag{3.2}$$

if output is higher than $\overline{y}$

The usual first-order conditions for labor and capital are not sufficient to describe the optimal levels of inputs (and corresponding output) since the presence of the tax threshold will lead to a corner solution for a range of productivity levels. For the constrained firms, the total profit given the first-order conditions for labor and capital is lower than the profit they will make if they limit output to $\overline{y}$ and don't have to pay any tax. For unconstrained firms, the concavity of the production function ensures that the interior solution maximizes profits. We also know from cost minimization that the constrained firms will maintain the same capital-to-labor ratio given by $\frac{k}{l} = \frac{w}{r}\frac{\alpha}{\beta}$ and will display the same conditional input demand functions as the unconstrained firms. This is because all firms face the same prices for labor and capital and that the production function is homogeneous. With this information we can now characterize the relationship between output, profit and productivity θ for all firms.

Proposition 1. Given the productivity level θ and prices r and w there are three sets of firms:

a) untaxed and unconstrained firms for which the optimal output falls below the threshold $\bar{y}$ and whose output and profit function is given y

$$\prod_a(0;\theta) = (1-\gamma)\gamma^{\frac{\gamma}{1-\gamma}}\left(\frac{\alpha}{r}\right)^{\frac{\alpha\gamma}{1-\gamma}}\left(\frac{\beta}{w}\right)^{\frac{\beta\gamma}{1-\gamma}}\theta^{\frac{1}{1-\gamma}} \tag{3.3}$$

b)untaxed and constrained firms that would have chosen some level of output $y(0,\theta)^* \geq \bar{y}$ absent the tax, but instead choose to limit production to $\bar{y}$ so as to not incur the tax payments and whose profit function is

$$\prod_b(0;\theta)|_{y=\bar{y}} = \bar{y} - \left(\frac{\alpha}{r}\right)^{-\alpha}\left(\frac{\beta}{w}\right)^{-\beta}\bar{y}^{\frac{1}{\gamma}}\theta^{-\frac{1}{\gamma}} \tag{3.4}$$

c)taxed and unconstrained firms that choose not to limit production and pay the tax because forgone profits due to limiting production are higher than their tax liabilities and whose profit function is

$$\prod_c((1-\tau);\theta) = (1-\gamma)\gamma^{\frac{\gamma}{1-\gamma}}\left(\frac{\alpha}{r}\right)^{\frac{\alpha\gamma}{1-\gamma}}\left(\frac{\beta}{w}\right)^{\frac{\beta\gamma}{1-\gamma}}(1-\tau)^{\frac{1}{1-\gamma}}\theta^{\frac{1}{1-\gamma}} \tag{3.5}$$

Since the firms' profits are increasing and monotone in θ, we know that there exists a unique θ^* such that $\prod(0;\theta)|_{y=\bar{y}} = \prod(1-\tau;\theta)$ and $y(1-\tau;\theta^*) \geq \bar{y}$, namely that the firm's profit when output is constrained is equal to its unconstrained profit when taxed. Moreover, all firms with $\theta \geq \theta^*$ will produce output $y^* = y(1-\tau;\theta^*)$ and pay taxes. Figure 3.5 shows how a firm's profit varies with productivity θ, tax status and whether a it limits output or not according to Proposition 1. The first vertical line gives the productivity level of the firm whose interior solution to the maximization problem with no tax coincides with the tax threshold $\bar{y}$. The second vertical line denotes the firm with the highest productivity level that will constrain its output at y^*. Note that all firms are located along the full line.

Finally, this model predicts a gap in output levels so that there will be no firm that produces output levels between $\bar{y}$ and y^*. This is shown in Figure 3.6. The histograms plot the distribution of firm output (pre-tax) with a tax threshold and the line graphs plot the counterfactual density when firms face no size dependent tax. The top half of the figure depicts perfect bunching as predicted by our model. In reality however there is noise in the data due to reasons like incorrect reporting, lack of control over incomes,etc. By adding noise to the data we can reconstruct a graph that looks similar to the behavior of firms in the actual data, given by Figure 3.1. This is depicted in Figure 3.6. In the next section we discuss how we derive the counterfactual distribution for the actual data.
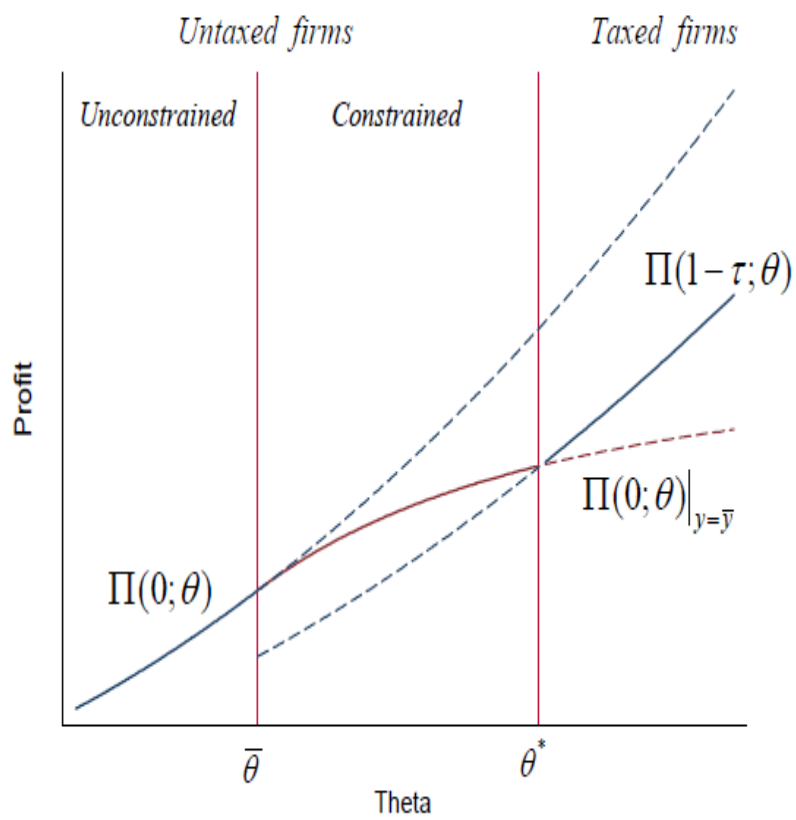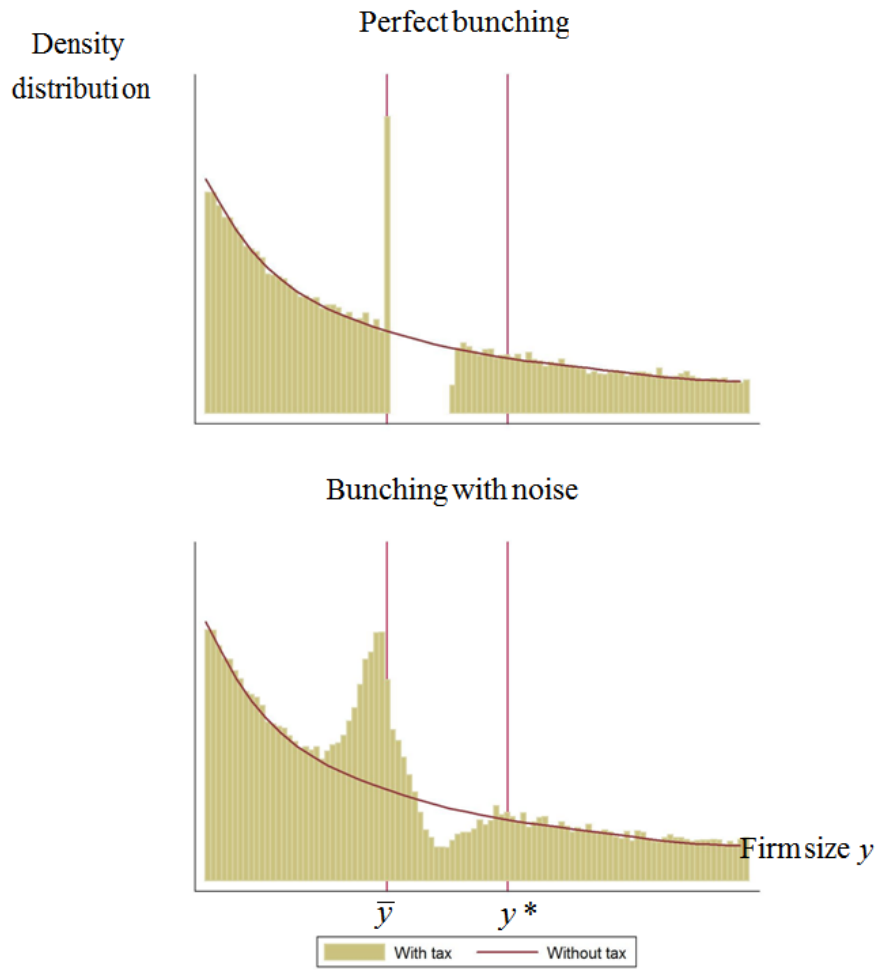
59

Figure 3.5: Productivity and Profits in the Model

Figure 3.6: Bunching Behavior of Firms in the Model

3.7 Counterfactual Output using Maximum Likelihood Estimation

In this section we construct the smooth counterfactual distribution of undistorted output using maximum likelihood estimation (MLE). We carry out the estimation procedure by excluding data on the bunching firms. We first determine the limits of the bunching region in the distribution. We assume that the log of undistorted output is normally distributed, i.e. $logy \sim N(\mu, \sigma)$. Figure 3.7 shows the distribution of actual log output and the normally distributed counterfactual log output. Output levels are rescaled to millions before taking logs. The vertical line in the figure is the tax threshold limit at Rs. 10 million, equal to Rs. 2.3 million on the log scale. The area that lies above the normal distribution curve in the neighborhood of the tax threshold is the bunching area. By visual inspection we determine the bunching limits to be Rs. 1.7 and 2.7 million. In level terms these limits are Rs. 6 million and Rs. 13.5 million.As explained earlier we do not get perfect bunching at the threshold either because firms cannot control output perfectly or due to inaccurate reporting. We now estimate the counterfactual distribution of the undistorted level of output by excluding the bunching region determined by the limits described above. With our assumption of normality of log output and the limits of the bunching region we determine the log likelihood function as follows.

Let the lower limit of the bunching region be denoted by y_l and the upper limit by y_u. $\overline{y}$ is the threshold level of output. In the absence of noise there should be a gap between $\overline{y}$ and y_u. These are the bunching firms that restrict output to the threshold level, $\overline{y}$. We however find that there is noise in the data and that this noise is symmetric around the threshold. The probability density function of the truncated normal distribution that excludes the bunching region is given by equation (3.6).

$$f(y; \mu, \sigma, y_l, y_h) = \frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{\left[1 - \Phi(\frac{y_u-\mu}{\sigma}) + \Phi(\frac{y_l-\mu}{\sigma})\right]} \qquad (3.6)$$

The log-likelihood of the truncated distribution of y is given by equation (3.7).

$$L = \Sigma_i^n ln\left[\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})\right] - ln\left[1 - \Phi(\frac{y_u-\mu}{\sigma}) + \Phi(\frac{y_l-\mu}{\sigma})\right] \qquad (3.7)$$

We first carry out maximum likelihood estimation for the benchmark case where there are no distortions and no taxes. All firms in the economy are untaxed and unconstrained. In order to estimate the parameters of the undistorted counterfactual distribution we need to feed in data on undistorted output of all firms excluding the firms in the bunching region in the log likelihood function given in equation (3.7). We however do not observe the undistorted output of all firms in our dataset. Our model in section 3.6 predicted that there would be three sets of firms:

62

Figure 3.7: Counterfactual Distribution - Maximum Likelihood Estimation
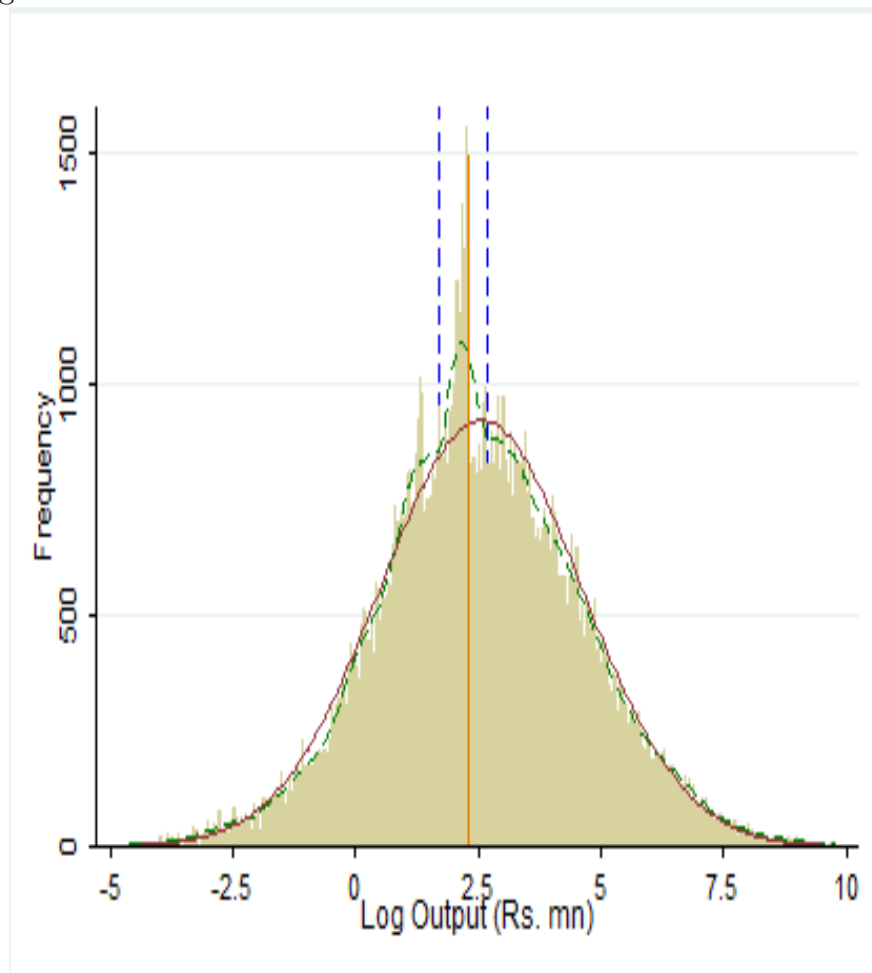
- Untaxed and unconstrained firms whose output $y_a < \overline{y}$

- Untaxed and constrained firms whose output $y_b = \overline{y}$

- Taxed and unconstrained firms whose output is $y_c > \overline{y}$.

The profit maximizing output of each of these three sets of firms is given by equations (3.8)-(3.10).

$$y_a^* = \gamma^{\frac{\gamma}{1-\gamma}} \left(\frac{\alpha}{r}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta\gamma}{1-\gamma}} \theta^{\frac{1}{1-\gamma}} \tag{3.8}$$

$$y_b^* = \overline{y} \tag{3.9}$$

$$y_c^* = \gamma^{\frac{\gamma}{1-\gamma}} \left(\frac{\alpha}{r}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta\gamma}{1-\gamma}} (1-\tau)^{1-\gamma} \theta^{\frac{1}{1-\gamma}} \tag{3.10}$$

Only the output of firms below the threshold $\overline{y}$, denoted by y_a^*, is undistorted (unconstrained and untaxed). Equations (3.8)-(3.10) in our model however allow us to determine the undistorted output levels of firms with output above y_u, denoted by y_c^*, for a given level productivity θ, the tax rate τ and the scale parameter of the production function, $\gamma = \alpha + \beta$.

$$y_c^* = (1-\tau)^{\frac{\gamma}{1-\gamma}} y_a^* \tag{3.11}$$

In order to calculate the the untaxed and unconstrained output of a firm whose output $y_c > \overline{y}$ we need to divide y_c^* by $(1-\tau)^{\frac{\gamma}{1-\gamma}}$. The basic excise tax rate in 2004-05 was 16% on average in India. We therefore set τ equal to 0.16. For the value of the scale parameter γ we appeal to the literature. Empirical studies of the returns to scale parameter such as Basu and Fernald (1997), Nguyen and Reznek (1990), Atkeson and Kehoe (2005) and Gorodnichenko (2007) find that the returns to scale parameter is typically between 0.8 and 0.9. Gorodnichenko (2007) finds that the profit share in revenue is a robust non-parametric economic diagnostic for estimates of returns to scale. The median revenue profit margins of firms that lie below the bunching region is around 4%. The median revenue profit margins of the bunching firms is around 8% and those above the bunching region is around 15%. The value of the scale parameter suggested by these profit margins is in line with the findings of empirical studies on the scale parameter. We carry out the estimation by setting the scale parameter equal to 0.70, 0.75, 0.80, 0.85, 0.90 and 0.95. We feed the untaxed and unconstrained output in our log-likelihood function to estimate the mean, μ, and standard deviation , σ, of the counterfactual distribution of undistorted output.

The estimated parameters, μ and σ, allow us to calculate the undistorted counterfactual level of output both within and outside the bunching region, number of bunching firms and the undistorted level of output. The number of bunching firms is estimated as follows. Because firms bunch at the threshold only from above, the underlying distribution immediately to the left of $\overline{y}$ is identified

by the lognormal assumption. Firms located immediately to the right that fill the gap between $\bar{y}$ and y_u come from a mixed distribution of bunching firms as well as unconstrained firms that cannot perfectly control their desired output levels. Selecting only the lower portion of bunching firms will give us an unbiased estimate of the total number of bunching firms if the output noise is symmetrically distributed around zero.

The estimated parameters of the undistorted counterfactual distribution also allow us to calculate the revenue neutral flat tax rate. A revenue neutral flat tax rate is the tax rate that is uniform across all firms and that yields the same amount of tax revenue as in the current tax regime. In this economy all firms behave like the taxed and unconstrained firms whose profit function is given by equation (3.5) and equilibrium output by equation (3.10).

Let TR denote total tax revenue in the existing regime. TR is the revenue generated by taxing firms with output more than Rs. 10 million at 16%. If there are a total number of N firms in the economy and each firm is indexed by i, the revenue neutral flat tax rate is given by τ_{flat} such that,

$$\sum_i^N \tau_{flat} y_c^* = TR \tag{3.12}$$

$y*_c$ is the equilibrium level of output in the taxed and unconstrained case. Using equation (3.11) we can rewrite equation (3.12) such that the flat tax rate is a function of the total revenue and total undistorted level of output.

$$\tau_{flat}(1 - \tau_{flat})^{\frac{\gamma}{1-\gamma}} = \frac{TR}{\sum_1^N y_a^*} \tag{3.13}$$

y_a^* in equation (3.13) is the equilibrium level of output in the undistorted counterfactual case and τ_{flat} is the flat tax rate. We calculate TR and $\sum_1^N y_a^*$ using the estimated parameters of the undistorted counterfactual distribution, μ and σ, and the expression for the moments of the truncated lognormal distribution in the equations given below by equations (3.14) and (3.15).

$$TR = 0.16 N \Phi(-a) EV(y_c^*) \frac{\Phi(\sigma - a)}{\Phi(-a)} \tag{3.14}$$

$$\text{where} \quad EV(y_c^*) = e^{\frac{\gamma}{1-\gamma} ln(1-0.16) + \mu + \frac{\sigma^2}{2}}$$

$$a = \frac{y_u - \mu}{\sigma}$$

$$\sum_i^N y_a^* = N * EV(y_a^*) \tag{3.15}$$

65

$$\text{where} \quad EV(y_a^*) = e^{\mu + \frac{\sigma^2}{2}}$$

Given total tax revenue, total undistorted output and the scale parameter we estimate the flat tax rate, τ_{flat} , in equation (3.13) using numerical methods. The parameter estimates, mean counterfactual output level before and after flat tax, the number of bunching firms, the output of the highest bunching firms and the revenue neutral flat tax rate for different values of the scale parameter are provided in Table 3.5. The following are the key observations:

1. Our model does a good job of estimating the parameters of the counterfactual distribution of output. The estimated mean μ and σ are significant for all values of the scale parameter. The mean and standard deviation of log output are 2.56 and 2.08 for all firms in our dataset. The mean and standard deviation excluding the bunching area are 2.66 and 2.08.

2. The results are sensitive to the choice of the scale parameter, σ. The results appear less sensitive to smaller values of σ. This can be explained by the relationship between distorted and undistorted output given by equation (3.11). When the returns to scale parameter is large the reduction in equilibrium output due to the presence of a tax is more costly than in the case where the returns to scale parameter is small.

3. The revenue neutral flat tax rate is close to the existing tax rate of 16% for most values of the scale parameter. The flat tax rate converges to 15.6% as the scale parameter moves closer to 0. The revenue flat tax rate falls exponentially as the scale parameter approaches 1. A similar non linear relationship is also observed between the scale parameter and the number of bunching firms.

Table 3.6 tells us the output gains or losses when we move from the current tax regime to a flat tax regime. So far we have used the undistorted counterfactual level of output in the economy where there are no taxes as our benchmark. A more accurate benchmark to measure the effects of the excise tax exemption below a certain threshold would be the counterfactual level of output under a revenue neutral flat tax regime. The expression for total and expected value of counterfactual output under the flat tax regime is given in equation (3.16).

$$\sum_i^N y_a^* = N * EV(y_c^*) \tag{3.16}$$

$$\text{where} \quad EV(y_c^*) = e^{\frac{\gamma}{1-\gamma} ln(1-\tau_{flat}) + \mu + \frac{\sigma^2}{2}}$$

The following are the key insights emerging from the results presented in Table 3.6.

1. The gains to total output are always positive. However the gains under flat tax regime are not positive for all firms. In order to see this we divide the distribution of output into three segments: bunching firms (with output below $\overline{y}$ and y_u), below bunching firms (with output below $\overline{y}$) and above bunching firms (with output above y_u). The bunching firms and the

Table 3.5: Results of Maximum Likelihood Estimation

Scale	Estimated Mean	Estimated Std.	Bunching Firms	Output-Highest	Flat Tax
(γ)	(μ)	Deviation (σ)		Bunching Firm	
	(log Rs.mn)	(log Rs.mn)	(No.)	(Rs.mn)	(%)
0.70	2.77	2.19	6520	14.6	15.1
0.75	2.83	2.20	6859	15.0	14.6
0.80	2.92	2.23	7362	15.5	13.7
0.85	3.06	2.29	8187	16.6	10.5
0.90	3.34	2.38	9774	19.0	4.6
0.95	4.20	2.7	13900	30.3	0.4

Table 3.6: Gains in Counterfactual Economy (MLE) under Flat Tax Regime

Scale	Flat Tax	Total Output	Bunching	Below Bunching	Above Bunching
(γ)	(%)	Gains(%)	Gains(%)	Gains (%)	Gains (%)
0.70	15.1	3.3	96.7	-31.8	2.5
0.75	14.6	6.1	114.3	-37.7	5.1
0.80	13.7	13.0	143.0	-44.5	11.4
0.85	10.5	46.7	243.5	-46.7	43.2
0.90	4.6	230.0	774.8	-34.5	214.4
0.95	0.4	2941.0	9373.0	-7.3	2444.6

67

above bunching firms always gain and the below bunching firms always loose under the flat tax regime. These results are intuitive. The below bunching firms are exempt from paying any taxes in the current tax regime and therefore will loose when they pay a tax under the flat tax regime. The firms above the bunching area will gain for the reason that they pay lower taxes than before in the flat tax regime. The gains to the bunching firms can be substantial, depending on the choice of the scale parameter, in the flat tax regime. Using the empirical estimates of the scale parameter, which are typically in the range of 0.8 and 0.9, we find that the overall gains could range from 13.0% to 230%.

2. The gains to total output for a given scale parameter depends largely on the gains to the large firms, that is, the firms above the bunching region. This is because large firms, as highlighted in Tables 3.1 and 3.2, contribute to more than 75% of the total output in the economy. Having said that, the number of firms in the region below the bunching constitute nearly 30% of all the firms in the economy. Given that the flat tax has differential effects on small versus large firms it is important to study the welfare effects of such a tax. How does the flat tax affect the allocation of other inputs, such as labor and capital? We need a general equilibrium model to answer such a question. In this chapter we restrict our focus to studying the partial effect of removing a tax distortion on output.

3. The relationship between the scale parameter and output gains is non linear in nature. The output gains are quite small for values of the scale parameter below 0.5. The output gains increase exponentially as the scale parameter increases and become explosive when the value of the scale parameter is very close to one. And once again the gains are more sensitive at higher values of the scale parameter. This is again intuitive because when the scale parameter is high, the large firms gain much more than the small firms when the tax rate is reduced. The gains to total output become exponential when the scale parameter approaches one.

In the next section we carry out the estimation of the counterfactual distribution, the number of bunching firms and the output gains or losses using non=linear least squares (NLLS) estimation method.

3.8 Counterfactual Output using Non-Linear Least Square Estimation

In this section we estimate the parameters of the undistorted counterfactual distribution using non linear least squares estimation. Using this alternative estimation technique serves as a robustness check for the maximum likelihood estimation of the previous section. Here the estimation is done on the empirical density of output using count data similar to Saez (2010) and Chetty et al. (2009) instead of the firm-level data as in the previous section. However, these authors use nonparametric methods to estimate counterfactual densities while we make use of the stronger but more powerful assumption of log-normality given its good fit with the overall distribution of output in our sample.[8]

As in the previous section, we adjust output levels for firms above the excise threshold to recover the pre-tax level of production. Given our model in section 3.6, this is obtained by dividing observed output by $(1 - 0.16)^{\frac{\gamma}{(1-\gamma)}}$ for all firms producing more than 10 million rupees. In order to generate the empirical density, we take the log value of output and group firms into bins of width equal to 0.05 log points. This value generates around 350 bins for our sample. The estimation results are not sensitive to the choice of the bin width.

We then estimate the following non-linear equation by excluding firms in the bunching region $y_i \in [y_l, y_u]$

$$D_i = \frac{1}{\sigma} \phi \left(\frac{y_i - \mu}{\sigma} \right) + \sum_{j=l}^{u} \beta_j \mathbf{1}[y_i = y_j] + \epsilon_i$$

where D_i is the density of firms in bin i[9], $\phi(.)$ is the normal density and ϵ_i a mean-zero error term. We select the limiting values y_l and y_u of the bunching range by the same visual inspection as in the previous section. The parameters to be estimated are μ and σ, the mean and standard deviation of the counterfactual density $\phi(.)$ and the $\beta_j's$. These coefficients measure the difference between the actual density of firms and the normal density of bin i for all bins in the bunching range. Recovering the amount of bunching using these estimated parameters is straightforward. This estimated number of bunching firms is given by

$$B = 2N \times bin\ width \times \sum_{j=l}^{thr} \beta_j$$

where we only select bunching firms below the threshold $\overline{y}$. Because firms bunch at the threshold only from above, the underlying distribution immediately to the left of $\overline{y}$ is identified by the lognormal assumption. On the other hand, as illustrated in Figure 3.5 firms located immediately to the right that fill the gap between $\overline{y}$ and y_u come from a mixed distribution of bunching firms

---

[8]Monte Carlo simulations indicate that our estimator performs better both in terms of bias and variance than using polynomials to approximate the density of output around the threshold.

[9]The density D_i is obtained by dividing the number of firms in bin i by the product of the bin width and total number of firms in the sample. This transformation ensures that the sum of densities is equal to one.

Figure 3.8: Counterfactual Density - NLLS Estimation
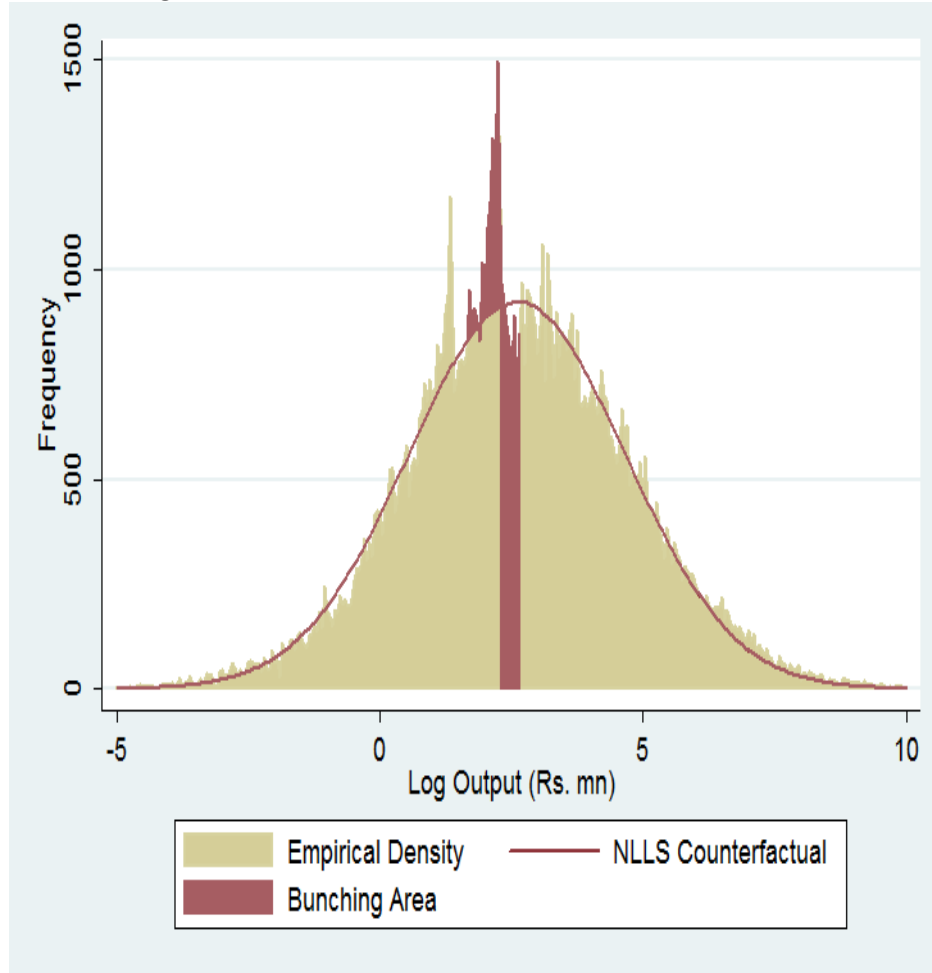
as well as unconstrained firms that cannot perfectly control their desired output levels. Selecting only the lower portion of bunching firms will give us an unbiased estimate of the total number of bunching firms if the output noise is symmetrically distributed around zero.

Figure 3.8 plots the empirical density with a bin width of 0.05 log points along with the estimated counterfactual distribution. The bunching area using the same limiting values as for the previous estimation is shaded in red. The estimation is carried out for a initial value of the scale parameter $\gamma = .5$. Table 3.7 presents the estimated mean μ, standard deviation σ of the undistorted counterfactual distribution, the number of bunching firms, output of the highest bunching firm and the revenue neutral flat tax rate for different values of the scale parameter. Table 3.8 presents the output gain or loss under the flat tax regime using results from the NLLS estimation. We once again show the differential effects of the flat tax on the firms below the bunching region, firms in the bunching region and firms above the bunching region. The results are presented for scale parameter values 0.70, 0.75, 0.80, 0.85, 0.90 and 0.95. The estimation results using NLLS are very similar to the results obtained using maximum likelihood techniques.

Table 3.7: Results of Non Linear Least Squares Estimation

Scale	Estimated Mean	Estimated Std.Deviation	Bunching Firms	Output-Highest	Flat Tax
(γ)	(μ)	(σ)		Bunching Firm	
	(log Rs.mn)	(log Rs.mn)	(No.)	(Rs.mn)	(%)
0.70	2.72	2.10	6065	14.0	15.2
0.75	2.78	2.11	6418	14.3	15.0
0.80	2.88	2.14	6986	14.9	14.4
0.85	3.04	2.19	8093	16.2	11.7
0.90	3.39	2.34	10543	19.8	5.0
0.95	4.59	2.97	16843	45.6	0.6

Table 3.8: Gains in Counterfactual Economy (NLLS): Flat Tax Regime

Scale	Flat Tax	Total Output	Bunching	Below Bunching	Above Bunching
(γ)	(%)	Gains(%)	Gains(%)	Gains (%)	Gains (%)
0.70	15.2	2.6	111.7	-31.9	2.2
0.75	15.0	4.0	134.0	-38.6	3.6
0.80	14.4	8.3	170.3	-46.3	7.8
0.85	11.7	33.4	278.1	-50.6	32.7
0.90	5.00	204.5	920.3	-37.0	202.7
0.95	0.60	2355.1	10797.9	-10.8	2349.3

71

The results of both the maximum likelihood estimation and the non linear least squares estimation rely crucially on the assumption on γ, the scale parameter of the production function. The gains from removing the tax distortion and moving to a flat tax regime can vary a lot depending on the size of the scale parameter. In the next section we try to estimate the scale parameter for the formal manufacturing firms in India.

3.9 Returns to Scale and Welfare Gains

The output gains in our model are highly sensitive to the value of the scale parameter. We find that the relationship between output gains and the scale parameter, γ, is non linear in nature. The gains are less than 1% for $\gamma \in (0, 0.5)$ and increase in a non linear fashion for higher values of γ. The gains become explosive as γ approaches one. This is illustrated in Figure 3.9. Empirical estimates of the returns to scale parameter in the literature are typically between 0.8 and 0.9. For these values of the scale paramter our model estimates overall output gains from removing the tax distortion and moving to a flat tax regime to be in the range of 13% to 230%. Why are gains so explosive as γ approaches one?

The returns to scale parameter is important for determining how many firms will populate an industry. When increasing returns to scale exist, one large firm will produce more cheaply than two small firms. Theory would predict that all the resources in the economy would be allocated to one big firm rather than being spread over many firms. Small firms will thus have a tendency to merge to increase profits. This is why gains become explosive in our model when the scale parameter approaches one. All the firms in our model would want to allocate their resources to one large firm in order to reap the benefits of scale. On the other hand, if an industry has decreasing returns to scale, a merger of two small firms to create a large firm will cut output, raise average costs, and lower profits. In such industries, many small firms should exist rather than a few large firms. In the real world however we see evidence that small firms exist in large numbers both in developed and developing countries. How do we then reconcile the large values of the scale parameter that are found in the literature with the existence of large number of small firms?

In our model we are able to estimate the scale parameter by using the output of the highest bunching firm. The highest bunching firm is one that is at the margin and is indifferent between restricting output to the threshold level and paying taxes. In our model a firm's profits are increasing and monotone in productivity, θ. This ensures that there exists a unique θ^* such that $\prod(0; \theta)|_{y=\overline{y}} = \prod(1 - \tau; \theta)$ and $y(1 - \tau; \theta^*) \geq \overline{y}$, namely that the firm's profit when output is constrained is equal to its unconstrained profit when taxed. θ^* is the productivity of the firm that is indifferent between restricting output equal to the threshold level $\overline{y}$ or producing the unconstrained but taxed level of output given by $y_c^* = y(1 - \tau; \theta^*)$. Equating the equilibrium profit functions $\prod(0; \theta)|_{y=\overline{y}}$ and $\prod(1 - \tau; \theta)$ given by equations (3.4) and (3.5) and combining it with equation (3.11) gives the following non linear equation:

Figure 3.9: Returns to Scale and Gains
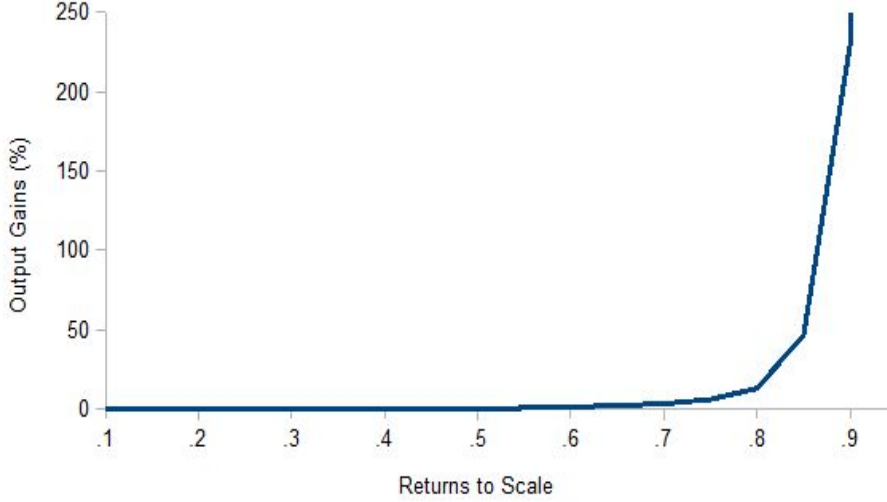
$$\frac{\overline{y}}{y_a^*} - \gamma \left(\frac{\overline{y}}{y_a^*} \right)^{\frac{1}{\gamma}} - (1-\gamma)(1-\tau)^{\frac{1}{1-\gamma}} = 0 \tag{3.17}$$

y_a^* is the undistorted level of output which we can calculate using the estimated parameters of the counterfactual distribution. The estimated value of γ is equal to 0.11. This value of the scale parameter appears to be very low both in terms of the other empirical estimates of the scale parameter in the literature, which are typically between 0.8 and 0.9, and the low profit margins of firms observed in the data. The model also predicts that when the scale parameter is close to one the gains become explosive as all firms merge together to form one large firm. In the real world we do not see such a phenomenon. On the contrary we see that a large number of firms are small and have very low profit margins. Thus, our model where, firms produce homogeneous goods and face decreasing returns fails to account for the existence of a large number of small firms and large values of the scale parameter at the same time.

A better approach to modeling the coexistence of a large number of firms and a large returns to scale parameter perhaps would be the Dixit-Stiglitz type of model where firms produce differentiated products. The consumers' "love for variety" in such a model would ensure that resources are spread over a large number of small firms. Output gains in such a model would be dependent on the trade off between the consumers' love for variety measured by the elasticity of substitution and the returns to scale. We hope to explore this idea more in our future work.

3.10 Conclusion

In this chapter we try to answer the question how costly is the excise tax exemption granted to small firms in India. Firms in India are exempt from paying excise tax if their output level is less than Rs. 10 million. Using firm level data on Indian manufacturing firms we provide strong evidence of bunching of firms around the threshold of Rs. 10 million. We use maximum likelihood estimation and non linear least squares estimation techniques to construct the counterfactual distribution where there no tax distortions. We estimate the number of bunching firms, the counterfactual level of undistorted output and the revenue neutral flat tax rate. We find that moving to a flat tax regime where are all firms are taxed at a uniform rate can lead to welfare gains in the range of 13% to 230% using the values of the scale parameter in the literature. Our estimate of the scale parameter is very low compared to the estimates found in the literature on retruns scale and data on profit margins of firms. We find that our model with homogenous goods and descreasing returns to scale is unable to explain the coexistence of a large number of firms and a large value of the scale parameter. We think that a model where consumers have love for variety and firms produce differentiated goods would be able to explain this phenomenon.

Chapter 4

Conclusion

4.1   Summary of Findings

This study contributes to the growing literature on resource misallocation in understanding cross country income differences. The contributions of this study can be summarized under two broad categories. The first one is with regard to the proper measurement of resource misallocation and the second one is with regard to the identification of the nature and sources of misallocation. We first summarize our findings on the topic of measurement of resource misallocation.

With regard to measurement, the first thing we do is to measure misallocation for the entire manufacturing sector by incorporating the informal sector in our analysis. Earlier work by Hsieh and Klenow (2009) included the study of the formal manufacturing sector in India only. The presence of the informal sector in manufacturing in India is huge, when viewed especially in terms of employment generation. By incorporating the informal sector in our study we are able to measure resource misallocation for the entire manufacturing sector in India. Using data on the informal manufacturing sector in India for the years 1994-95, 2000-01 and 2005-06 and data on the formal manufacturing sector for 2004-05 we first replicate the methodology adopted by Hsieh and Klenow to measure distortions using value added. We find that misallocation had increased in the formal sector for the period 1994-95 to 2004-05 most of which came from the increase in the 90/10 percentiles of log TFPR. Efficient reallocation would result in nearly 111% TFP gains for the manufacturing sector as a whole. We also find that the misallocation in the formal sector to be larger compared to the informal sector for the period under study suggesting that the formal firms face larger distortions than informal firms. This result however is not very robust and is sensitive to the methodology used to measure productivity.

Our second contribution to the measurement of resource misallocation is the inclusion of intermediate inputs. We show that including intermediate inputs to measure resource misallocation is important for two reasons. One, to allow for potential distortions to intermediate inputs. Two, to account for the possible amplification of distortions across firms or industries via intermediate

inputs. We develop a general equilibrium model where firms are heterogeneous in productivity and face non-symmetric distortions. We show how distortions at the firm level can affect aggregate TFP. We also show that the degree of amplification of distortions depends on the share of intermediate inputs in production. We recalculate misallocation for the period mentioned in the above paragraph using gross output. We find that the composite measure of misallocation, given by the dispersion in total factor revenue productivity (TFPR), is much lower when gross output is used. Using gross output we find that the gains from reallocation are around 28% for the manufacturing sector as a whole. Moreover when we use gross output, misallocation is only slightly higher in the formal sector compared to the informal sector. We attribute the difference in the magnitude of misallocation between gross output and value added measures to omitted variable bias which typically occurs when value added measures are used to measure productivity at the industry level. We show that measuring the distortions to factors individually and looking at the variance-covariance matrix of distortions is important to understand the source and magnitude of this bias.

On the topic of identification of the nature and sources of misallocation we first try to answer the question if certain inputs are more distorted than others. The decomposition of the the variance of TFPR reveals that the distortions to intermediate inputs are the largest followed by distortions to capital. How these distortions contribute to overall misallocation is dependent on the factor shares and the covariance of distortions to individual factor inputs. We then use qualitative information on informal firms in our data set to identify a broad set of potential sources of resource misallocation. We find misallocation to be positively correlated with shortage of capital and non availability of intermediate inputs like electricity and raw materials and negatively correlated with market size of the firm, indicating that larger firms face more distortions.

We then move on to study the distortionary effects of a particular size dependent policy in India. A firm in India can get up to 100% exemption from paying excise tax if its annual turnover is less than Rs. 10 million. Using data on Indian manufacturing for 2004-05 we first show graphical evidence of bunching of firms at the threshold level of Rs. 10 million. We develop a simple model to describe this bunching behavior where firms are heterogeneous in productivity and face decreasing returns to scale in production. We then use graphical evidence of bunching to construct the counterfactual distribution of output without any tax distortion using maximum likelihood and non-linear least squares estimation methods. The estimated parameters of the counterfactual distribution are used to calculate the undistorted output, the number of bunching firms and the revenue neutral flat tax rate.

Using the empirical estimates of the scale parameter in the literature we find that there are output gains in the range of 13% to 230% under a revenue neutral flat tax regime. These gains are however sensitive to the scale parameter of the production function. The gains are explosive as the scale parameter approaches one. Using output of the highest bunching firm we estimate the value of the scale parameter to be 0.11. We argue that this number is very low when compared to the empirical estimates of the scale parameter found in the literature and when compared to the low profit margins of firms in our data. We find that our model with homogenous goods and decreasing returns to scale is unable to reconcile the existence of a large number of firms with a large value of the scale parameter implied by low profit margins. We conclude by suggesting alternative modeling approach to reconcile the above fact.

4.2 Suggestions for Future Work

This study fills some of the gaps in the existing literature on resource misallocation and efficiency. There is however a lot more work to be done in this area. We highlight some of the topics for future research below:

- Understanding the linkages between the formal and informal sectors: How does one aggregate output and productivity from the formal and informal sectors? Is the informal sector a drag on the formal sector or a help to cut costs via sub-contracting? How does the introduction of the value added sales tax (VAT) in 2006 in India affect the linkages between formal and informal firms?

- Identifying the nature and sources of misallocation: Why are the distortions to intermediate inputs and capital more than the distortions to labor? Are the sources of distortions different for the formal and informal sectors? How do distortions affect the movement of firms between the formal and informal sectors?

- Political economy concerns: What are the incentives for the government to remove some of the distortions identified in the literature? What should be the speed at which these distortions should be eliminated? What are the welfare implications of removing concessions to small firms? Is one big Walmart better than fifty small firms?

Bibliography

Aghion, Philippe, Robin Burgess, Stephen J. Redding, and Fabrizio Zilibotti, "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India," *American Economic Review*, 2008, *98*, 1397–1412.

Amaral, Pedro and Erwan Quintin, "A Competitive Model of the Informal Sector," *Journal of Monetary Economics*, 2006, *53*, 1541–53.

Arnold, Jens, Beata Javorcik, Molly Lipscomb, and Aaditya Mattoo, "Services Reform and Manufacturing Performance: Evidence from India," *CEPR Working Paper*, 2010.

Atkeson, Andrew G. and Patrick J. Kehoe, "Modeling and Measuring Organizational Capital," *Journal of Political Economy*, 2005, *113*, 10261053.

Azuma, Yoshiaki and Herschel I. Grossman, "A Theory of the Informal Sector," *Economics and Politics*, 2008, *20*, 62–79.

Bailey, Martin, Diana Farrell, and Janna Remes, "Domestic Services: The Hidden Key to Growth," 2005.

Banerjee, Abhijit and Esther Duflo, "Growth Theory through the Lens of Development Economics," in "Handbook of Economic Growth," Elsevier, 2005.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross Country Differences in Productivity: The Role of Allocative Efficiency," *NBER Working Paper*, 2009.

Basu, Susanto and John G. Fernald, "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 1997, *105*, 249–283.

Besley, Timothy and Robin Burgess, "Can Labor Regulation Hinder Economic Performance? Evidence from India," *Quarterly Journal of Economics*, 2004.

Bhide, Amar, "What holds back Bangalore Businesses?," *Asian Economic Papers*, 2008, *7*.

Bloom, Nick, Benn Eifert, Aparajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *World Bank Working Paper Series*, 2011.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006.

Buera, Francisco and Yongseok Shin, "The Dynamics of Reforms: Productivity Growth and Capital Flows," Technical Report, UCLA 2009.

Caselli, Francesco, "Accounting for Income Differences across Countries," in P. Aghion and S. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1A of *Handbook of Economic Growth*, Elsevier, 2005, chapter 9.

Chetty, Raj, Anders Frederiksen, John N. Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses and Labor Supply Elasticities: Evidence from Danish Tax Records," *NBER Working Paper*, 2009.

de Paula, Aureo and Jose Alexandre Scheinkman, "The Informal Sector," *NBER Working Paper*, 2007.

Farrell, Diana, "The Hidden Dangers of the Informal Economy," 2004, *3*, 26–37.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability ?," *American Economic Review*, 2008.

Gollin, Douglas., "Do Taxes on Large Firms Impede Growth? Evidence of Ghana," bulletins, Economic Development Center, University of Minnesota 2005.

Gorodnichenko, Yuriy, "Using Firm Optimization to Evaluate and Estimate Returns to Scale," *NBER Working Paper*, 2007.

_ , Jorge Martinez-Vazquez, and Klara Sabirianova Peter, "Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia," *International Studies Program Working Paper Series*, 2007.

Guner, Nezih, Gustavo Ventura, and Yi Xu, "Macroeconomic implications of size-dependent policies," *Review of Economic Dynamics*, 2008.

Hall, Robert E. and Charles I. Jones, "Why Do Some Countries Produce So Much More Output per Worker than Others?," *Quarterly Journal of Economics*, 1999, *114*, 83–116.

Harris, John R. and Michael P. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 1970, *60*, 126–142.

Hart, Keith, "Employment, Incomes and Equity: A Strategy for Increasing Productive Employment in Kenya," *American Economic Review*, 1972.

Howitt, Peter, "EndogenousGrowth and Cross-Country Income Differences," *American Economic Review*, 2000, *90*, 829–846.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009.

Jones, Charles I., "Comment on La Porta and Scleifer 'The Unofficial Economy and Economic Development'," comment, Brookings Papers on Economic Activity 2008.

__ , "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macro*, 2011.

__ , "Misallocation, Economic Growth and Input-Output Economics," *Proceedings of the 10th World Congress of the Econometric Society*, 2011.

__ and Paul M. Romer, "New Kaldor Fact: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics*, 2010, *2*.

Jorgenson, Dale W. and Kevin J. Stiroh, "Raising the Speed Limit: U.S: Economic Growth in the Information Age," 2000, *1:2000*, 125–234.

Klenow, Peter and Andres Rodriguez-Claire, "The Neoclassical Revival in Growth - Has it gone too far," in "NBER Macroeconomics Annual," Vol. 12, Cambridge, MA: MIT Press, 1997, pp. 73–114.

__ and __ , "Externalities and Growth," in P. Aghion and S. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1 of *Handbook of Economic Growth*, Elsevier, 2005, pp. 817–861.

La-Porta, Rafael and Andrei Schleifer, "The Unofficial Economy and Economic Development," *NBER Working Paper*, 2008.

Lewis, Arthur W., "Economic Development with Unlimited Supplies of Labour," *Manchester School of Economic and Social Studies*, 1954, *22*, 139–191.

Loayza, Norman V., "The economics of the informal sector: a simple model and some empirical evidence from Latin America," *Policy Research Working Paper Series*, 1997.

__ and Jamele Rigolini, "Informality trends and cycles," *Policy Research Working Paper Series*, 2006.

Lucas, Robert, "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9*, 508–523.

Mankiw, Gregory N., David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 1992, *107*, 407–437.

Midrigan, Virgiliu and Daniel Xu, "Finance and Misallocation: Evidence from Plant-Level Data," *NBER Working Paper*, 2010.

Nguyen, Sang V. and Arnold P. Reznek, "Returns to Scale in Small and Large US Manufacturing Establishments," *Center for Economic Studies Working Paper, US Census Bureau*, 1990.

Onji, Kazuki, "The response of !rms to eligibility thresholds: Evidence from the Japanese value-added tax," *Journal of Public Economics*, 2009.

Parente, Stephen L. and Edward P. Prescott, *Barriers to Riches*, Cambridge, MA: MIT Press, 2000.

Pavnick, Nina and Penny Goldberg, "The Effects of the Colombian Trade Liberalization on Urban Poverty," *NBER Working Paper*, 2005.

Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Journal of Economic Dynamics*, 2008.

Saez, Emmanuel, "Do Tax Payers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2*, 180–212.

Schneider, Friedrich, "Shadow Economies and Corruption All over the World: New Estimates for 145 Countries," *Economics: The OpenAccess*, 2007, *1.*

Soto, Hernando De, *The Other Path: The Invisible Revolution in the Third Worlds*, New York: Harper and Row, 1989.

Viner, Jacob, *The Doctrine of Comparative Costs*, Fisher, 1932.

Weinzierl, Matthew C. Mankiw Gregory N. , "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," 2009, pp. 09–139.