Analyzing the Sources of Misallocation in Indian Manufacturing: A Gross-Output Approach

Sujana Kabiraj*

It is well established that misallocation of factor resources lowers productivity. In this paper, I use data from both formal and informal firms to study distortions in input and output markets as sources of misallocation in the Indian manufacturing sector. My work extends the seminal work of Hsieh and Klenow (2009). I consider output, capital, raw material, energy, and service sector distortions in a monopolistically competitive framework to measure the aggregate dispersion in total factor revenue productivity (TFPR). I also decompose the variance in TFPR and show that raw material and output distortions play a major role in defining aggregate misallocation.

Keywords: distortion, Indian manufacturing, misallocation, productivity
JEL codes: E10, O41, O47

I. Introduction

According to the World Bank, the per capita income of the United States (US) was 30 times that of India in 2017. Explaining such differences is one of the fundamental problems in growth economics. Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) demonstrate that the disparity in total factor productivity (TFP) is the primary source of cross-country income differences. In this context, another debate is about the sources of TFP differences among rich and poor nations. Banerjee and Duflo (2005), Restuccia and Rogerson (2008), and Hsieh and Klenow (2009) argue that in poor countries, some TFP differences are generated from a misallocation of resources across firms. In this paper, I follow the aforementioned notion that resource misallocation is a primary source of variation in TFP. I include intermediate inputs such as raw materials, energy, and services into the model of Hsieh and Klenow (2009) to obtain the extent of misallocation that originates from factor market distortions in a developing country such as India.

*Sujana Kabiraj: University of Wisconsin-Stevens Point, United States. E-mail: skabiraj@uwsp.edu. I am grateful to Jenny Minier of the University of Kentucky for providing the data used for the empirical analysis. I thank Areendam Chanda and other members of the faculty at Louisiana State University for their valuable comments as well as the managing editor and anonymous referees for helpful suggestions. The Asian Development Bank recognizes "China" as the People's Republic of China and "Bangalore" as Bengaluru. The usual ADB disclaimer applies.

When measuring physical TFP, one can adopt either of the two known approaches to measuring a firm's output: value added or gross output. The former excludes intermediate inputs, whereas the latter includes them. The difference between the two measures of TFP is more pronounced at the firm or industry level rather than in aggregate output. Gullickson and Harper (1999); van der Wiel (1999); Hulten, Dean, and Harper (2001); and Cobbold (2003) have demonstrated the benefits of the gross-output approach over the value-added method. The productivity manual published by the Organisation for Economic Co-operation and Development (2001) concludes that the gross-output approach is more appropriate for productivity measurement because it reduces productivity measurement bias. Based on these findings, I extend the Hsieh–Klenow model to measure productivity using the gross-output approach by including raw material, energy, and service sector intermediate inputs as factors of production.[1] The inclusion of these factors separately into the production process enables a more detailed representation of factor misallocation. Furthermore, the decomposition of factor market distortions by considering each factor input distortion separately provides a way to distinguish the level of misallocation in each factor market and to identify the corresponding potential gain from reallocation. I find that distortions in the output market and raw material market explain the lion's share of the variation in productivity.

TFP is a residual in the production process and is not observed directly. Moreover, it is difficult to measure firm-level TFP as the unit of production varies across firms. Therefore, I measure the variation in total factor revenue productivity (TFPR), which by definition is the product of output price and the physical TFP of a firm. In the absence of any factor market misallocation, TFPR should be equal for all firms within an industry. The intuition behind this claim is as follows: *if a firm has a high TFP, the marginal cost as well as the output price for that firm will be proportionally lower compared to a low-TFP firm in a particular industry, thus equalizing TFPR*. Based on this intuition from Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), I build my empirical results by using data from both formal and informal manufacturing sector firms in India for the survey year 2005–2006. In such a developing country, the informal sector plays an extensive role in shaping the economy. The informal manufacturing sector in India consists of around 17 million firms that provide 82% of total employment in that sector. Hence, it seems appropriate to include informal sector data in the empirical analysis.

My work has the closest resemblance to the paper by Chatterjee (2011). I extend the paper by including service sector inputs and energy inputs in the model separately. There exists a severe distortion in tariff rates in India's energy sector. For example, during 1999–2000, the industrial sector paid a tariff on electricity almost

---

[1] The KLEMS gross-output approach decomposes the factors of production into capital (K), labor (L), energy (E), materials (M), and services (S). The use of the KLEMS approach facilitates a decomposition of sources of growth in the production of firms as well as the inputs used in such production.

15 times higher than that paid by the agriculture sector and 2.1 times higher than that paid by the domestic sector (Thakur et al. 2005). Although the Electricity Act of 2003 worked toward the reduction and gradual elimination of cross subsidies, such distortions may have some impact on the cost of energy usage for small and large firms as well as in formal and informal sectors. Besides, small firms often have to use other electricity sources (such as generators), which in turn may impact resource allocation differently in smaller firms as compared to their bigger counterparts. Additionally, liberalization in the service sector in the early 1990s has resulted in significant growth in the sector. According to Chanda and Gupta (2011), service sector reforms along with external market linkages led to substantial growth in the most liberalized service sectors such as business services, banking, insurance, education, medical and health, and others. There is evidence in the literature that can link service sector reform to productivity in the manufacturing sector. For example, Arnold, Javorcik, and Mattoo (2011) demonstrate a positive relationship between service sector reform and the performance of manufacturing firms in the Czech Republic. In India, the cost share of service inputs is around 10% and that of energy is around 7% for the manufacturing sector. Exclusion of these factor inputs might lead to misleading measurements of output and productivity. I also include distortions in the energy and service sectors to verify whether some of the variation in firm-level TFPR is attributed to these factors. I find that there is very little variation in TFPR due to energy input distortions and that misallocation in service inputs is more pronounced in the dispersion of TFPR. On the other hand, I find output and raw material distortions are the primary sources of misallocation in the manufacturing sector. Another interesting result is that the distortions, when taken from several factor markets, together reduce the variation in TFPR. This surprising result will be the subject of further research.

The rest of the paper is organized as follows. Section II discusses the relevant literature. In section III, I present a theoretical model to show how TFPR is affected by firm-level distortions. Section IV describes the data, and section V analyzes the empirical results and the decomposition of the variance of TFPR. Section VI sheds some light on the misallocation among different groups of industries within the manufacturing sector. In section VII, I construe some relationship between firm size and misallocation in factor markets. Finally, I conclude in section VIII.

II. Literature Review

My work is related to a large body of literature that has accumulated over the last few decades. Hsieh and Klenow (2009) argue that in a monopolistically competitive framework, misallocation in factor markets can result in large differences in TFP and in output among firms within an industry. For example, a capital market distortion caused by the disparity in access to cheap credit will result in differences in the marginal product of capital among firms. Hsieh and Klenow

argue that in such a situation, the aggregate economy will be better off by allocating more capital to the firm with the higher marginal product of capital. Using firm-level data from India and the People's Republic of China (PRC), they calculate the TFP gain from reallocating capital, equalizing TFPR within the industry, to be 30%–50% in the PRC and 40%–60% in India. I follow the same intuition in this paper. I include raw materials, energy, and service sector inputs as factors of production and find the effect of distortions in all those inputs on firm-level TFPR. The goal is to find the empirical measurement of distortions in individual factor markets on aggregate TFPR.

Restuccia and Rogerson (2008) demonstrate the effect of factor distortion on TFP. They state that different taxes and policies across firms create disparities in prices and lead to a 30%–50% decrease in output and TFP in developing countries. Midrigan and Xu (2014) argue that financial frictions cause variations in TFP across firms through two channels. In particular, financial frictions distort entry decisions and technological adoption of producers. Furthermore, they create disparities in return to capital among producers. Fernald and Neiman (2011) deviate from the standard setup of monopolistic competition. They show that, in a two-sector economy with heterogeneous financial policies and monopoly power, TFP measured in terms of quantities and real factor prices can diverge.

There is a body of literature based on Hsieh and Klenow's framework. Camacho and Conover (2010) use Hsieh and Klenow's methodology to measure productivity differences through misallocation in resources for Colombian industries. Taking the US as the benchmark economy, they find a wide TFPR distribution for Colombia, which implies large resource misallocation across firms. They also calculate that the reallocation of labor and capital among firms will improve aggregate TFP by 47%–55%. Another paper by Kalemli-Ozcan and Sørensen (2014) measures TFP dispersion through capital misallocation for 10 African countries using the World Bank enterprise survey data. They argue that access to finance is one of the main sources of substantial capital misallocation. Dias, Marques, and Richmond (2016) extend the Hsieh–Klenow model to include intermediate inputs and measure TFP disparity using firm-level data from Portugal. They consider data from all sectors of the economy. Consequently, the endogenous intermediate inputs in their model take into account goods produced by all sectors. They find huge misallocation across industries. According to their results, in the absence of misallocation within industries, there would have been a 48%–79% gain in value-added output during 1996–2011. In India, it is rather difficult to find firm-level data for sectors other than manufacturing; hence, I take aggregate input produced by other sectors as exogenously given in the model.

The most closely related work to my research is the paper by Chatterjee (2011), which extends the Hsieh–Klenow framework for both formal and informal manufacturing sectors in India. Chatterjee also includes intermediate input market distortions in the model as a source of variation in TFP. She assumes that the economy has an intermediate input, aggregated from a fraction of total production

by each existing firm. The data used in Chatterjee's paper was obtained from
the Annual Survey of Industries (ASI) for formal firms and the National Sample
Survey Office (NSSO) for informal sector firms, as in the case in my paper.
Since both of these surveys primarily focus on manufacturing sector firms, the
aggregated intermediate input produced from these firms will take into account only
manufacturing sector products. Consequently, Chatterjee ignores inputs from other
sectors such as energy and services in her model. On the other hand, I consider
aggregated energy and service inputs as exogenously given in the model apart from
the combined raw materials produced by the existing firms. In the next section, I
extend Hsieh and Klenow's model to measure the degree of misallocation in the
economy.

III. Model

I consider a static one-period model without uncertainty, used by Hsieh and
Klenow (2009). I assume that the economy consists of J manufacturing industries
indexed as $j = 1, 2, \ldots, J$. Each industry consists of N_j monopolistically competitive
firms indexed as $i = 1, 2, \ldots, N_j$. Each firm produces differentiated products and
thus has substantial market power. The firms have heterogeneous productivity A_{ij}
as exogenously given and an endowment of capital K_{ij}, labor L_{ij}, raw material M_{ij},
energy E_{ij}, and service sector input Z_{ij}. Firms combine the factors to produce a
good using a Cobb–Douglas production function. The firm's production function is
as follows:

$$Y_{ij} = A_{ij} K_{ij}^{\alpha_{K_j}} L_{ij}^{\alpha_{L_j}} M_{ij}^{\alpha_{M_j}} E_{ij}^{\alpha_{E_j}} Z_{ij}^{\alpha_{Z_j}}$$

where $\sum_S \alpha_{S_j} = 1$ and $S \in \{K, L, M, E, Z\}$.

I consider only manufacturing sector firms in the model because I could
find data only for the manufacturing sector in India, which I use for the empirical
analysis. For the sake of simplicity, I assume that all raw materials coming from
the manufacturing sector are aggregated into a single raw material M, whereas all
energy inputs and service sector inputs are aggregated into factor inputs E and Z,
respectively. A fraction of the output produced by manufacturing firms considered
in the model is aggregated as the manufacturing input M and used by the same
firms; hence, the price of M is taken as endogenously determined. On the other hand,
since service and energy inputs that are produced by firms in their respective sectors
are not considered in the model, I take the output prices of such firms as exogenous.
Some manufacturing products may also be used by service and energy sector firms
as intermediate inputs; these products are considered as part of consumption goods
in the model. I further assume that all firms in an industry have the same cost share
of factor inputs α_{S_j}, but there is a variation in factor shares between industries.

In this paper, I measure the misallocation in resources that affects firm-level
TFPR. Distortion in an input or output market does not always uniformly increase

(or decrease) the marginal product of the factors of production (MPF) for all firms. As firms equalize price with the marginal product of factor inputs, a firm that faces taxes will have higher MPF for service inputs than the firms facing subsidies. The intuition behind the entire literature based on Hsieh and Klenow (2009) originates from the hypothesis that aggregate productivity will be higher if factors can be reallocated from lower MPF firms to higher MPF firms.

I assume several types of factor market distortions in the model. Some elements that change the MPF for all inputs by the same proportion are called output distortions ($\tau_{Y_{ij}}$). Tax on the output of a firm affects all inputs proportionally and can be identified as an example of an output distortion. Moreover, if the distortion creates a discrepancy in only the marginal product of capital, I call it capital distortion ($\tau_{K_{ij}}$) in accordance with Hsieh and Klenow. Similar remarks hold for raw material distortion ($\tau_{M_{ij}}$), energy distortion ($\tau_{E_{ij}}$), and service sector input distortion ($\tau_{Z_{ij}}$). For example, differentiation in electricity price between small and large businesses is perceived as an energy distortion as it affects only the marginal product of energy. Note that labor distortion is not considered separately but that every other distortion affects the respective MPF, relative to the marginal productivity of labor.

Each firm produces a single good Y_{ij} that is used both as a final consumption good and as an intermediate raw material. C_{ij} and X_{ij} denote final consumption good and intermediate raw material, respectively, which are produced by the ith firm from the jth industry. Firms face a downward sloping demand schedule that resulted from the assumption of a differentiated product environment in a monopolistically competitive market. Hence, the industry's final good appears to be a constant elasticity of substitution aggregation of all firms' final goods represented as

$$Y_j = \left(\sum_{i=1}^{N_j} Y_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$$

where $\rho > 1$ is the elasticity of substitution. For simplicity, I assume the elasticity of substitution is the same for all industries. This assumption follows from the literature. Each industry's output is sold as consumption good C_j and intermediate raw material X_j as was the case with firm-level output. I further assume that the markets for consumption goods and raw materials, produced by each industry, is perfectly competitive. Hence, the final consumption good is aggregated from industry-level consumption goods by a Cobb–Douglas production function:

$$C = \prod_{j=1}^{J} C_j^{\theta_j}$$

where $\sum_{j=1}^{J} \theta_j = 1$

The intermediate raw material is produced endogenously by aggregating each industry's production of raw materials, again using a Cobb–Douglas production function:

$$M = \prod_{j=1}^{J} X_j^{\lambda_j}$$

where $\displaystyle\sum_{j=1}^{J} \lambda_j = 1$

In the above two equations, θ_j and λ_j are the factor shares of each industry in total consumption and total intermediate raw materials production, respectively. Each firm chooses intermediate raw materials from the aggregated M according to their productivity. The aggregate quantity of other inputs such as energy E and services Z are exogenous in the model. Hence, each firm chooses the optimal amount E_{ij} and Z_{ij} based on its production function. The industry aggregates E_j and Z_j are given by the sum over each firm's use of energy and services in that industry.

I will now solve the model for optimal factor resources and output by maximizing profit for the firm, industry, and economy. I assume that total factor resources are limited in the manufacturing sector by the aggregate use of factor resources of the firms in the sector.

For each $S \in \{K, L, M, E, Z\}$, we can write the aggregate factor resources as

$$S = \sum_{j=1}^{J} \sum_{i=1}^{N_j} S_{ij}$$

and solve for the equilibrium to identify the effects of distortion on productivity.

A.   Equilibrium Analysis

In this section, I present a comprehensive equilibrium structure for firms, industries, and the economy. The equilibrium consists of the quantities of the consumption good and the intermediate raw materials produced at the level of the firm, industry, and aggregate economy. It also takes into account the optimal amount of capital, labor, raw materials, energy, and services used by each firm. The input markets and final goods markets clear at equilibrium. I now solve the optimization problems for each market.

1.   Final Goods Problem

I assume a representative firm produces a final good Y that is used in consumption C and in raw material M for further production. C is produced using

consumption good C_j produced by the industries. I assume C to be a numeraire commodity with unit price P. Likewise, P_j represents the price of the fraction of output or final good produced by each industry Y_j. I do not distinguish between price of final good C_j and raw material X_j, produced by each industry, on the assumption that both are of the same good and are subject to the same cost and market structure. Hence, the optimization problem for the final consumption good is given by

$$\max_{C_j} PC - \sum_{j=1}^{J} P_j C_j \qquad (1)$$

subject to

$$C = \prod_{j=1}^{J} C_j^{\theta_j} \qquad (2)$$

2. Intermediate Raw Materials Problem

The fraction of output used as the intermediate raw material (M) is constructed by the representative firms aggregating the produced raw materials (X_j) from each industry. The price of the aggregated intermediate raw material M is given by p_m. The representative firm optimizes the production of M as follows:

$$\max_{M_j} p_m M - \sum_{j=1}^{J} P_j X_j \qquad (3)$$

subject to

$$M = \prod_{j=1}^{J} X_j^{\lambda_j} \qquad (4)$$

We can solve the final goods problem from equations (1) and (2) and the intermediate raw materials problem from equations (3) and (4) to find the prices set by representative firms. The market clearing price of the final good is

$$P = \prod_{j=1}^{J} \left(\frac{P_j}{\theta_j} \right)^{\theta_j} = 1 \qquad (5)$$

and the intermediate raw material's price is

$$p_m = \prod_{j=1}^{J} \left(\frac{P_j}{\lambda_j} \right)^{\lambda_j} \qquad (6)$$

The second equality in equation (5) follows from the assumption that C is a numeraire good. Both prices are functions of the industry price (P_j) and the share of each industry in producing the same good (θ_j and λ_j, respectively).

3. The Industry's Problem

The final goods produced by each industry Y_j are used as both final consumption good C_j and intermediate raw material X_j. I assume that C_j and X_j are fractions of the same good, hence they face the same optimization problem. Furthermore, C_{ij} and X_{ij} are fractions of a firm's output Y_{ij}; therefore, I assume that they are produced using the same production function and that they also incur the same marginal cost. It is safe to assume that the firms charge the same price P_{ij} for both parts of their output. I represent the industry's problem as

$$\max_{Y_j} P_j Y_j - \sum_{j=1}^{J} P_{ij} Y_{ij} \tag{7}$$

subject to

$$Y_j = \left(\sum_{i=1}^{N_j} Y_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \tag{8}$$

The market clearing industry price is

$$P_j = \left(\sum_{i=1}^{N_j} P_{ij}^{1-\rho} \right)^{\frac{\rho}{\rho-1}} \tag{9}$$

4. The Firm's Problem

To allow for factor misallocation in the input and output markets, I consider several types of distortions. I assume that there exists an output distortion ($\tau_{Y_{ij}}$) that affects the marginal product of each factor of production by the same proportion. I also consider capital distortion ($\tau_{K_{ij}}$), raw material distortion ($\tau_{M_{ij}}$), energy distortion ($\tau_{E_{ij}}$), and service sector input distortion ($\tau_{Z_{ij}}$) that affect the marginal product of capital, raw materials, energy, and service inputs, respectively, relative to the marginal product of labor.[2] Each firm solves the following profit maximization

---

[2]Since all distortions are measured relative to the labor market, I do not explicitly use labor distortion ($\tau_{L_{ij}} = 0$).

problem to choose the optimized level of capital, labor, raw materials, energy, and service inputs:[3]

$$\max_{Y_j} P_{ij} Y_{ij} \left(1 - \tau_{Y_{ij}}\right) - w L_{ij} - r \left(1 + \tau_{K_{ij}}\right) K_{ij} - p_m \left(1 + \tau_{M_{ij}}\right) M_{ij}$$

$$- p_e \left(1 + \tau_{E_{ij}}\right) E_{ij} - p_z \left(1 + \tau_{Z_{ij}}\right) Z_{ij} \tag{10}$$

subject to

$$Y_{ij} = A_{ij} K_{ij}^{\alpha_{K_j}} L_{ij}^{\alpha_{L_j}} M_{ij}^{\alpha_{M_j}} E_{ij}^{\alpha_{E_j}} Z_{ij}^{\alpha_{Z_j}} \tag{11}$$

Solving firm i's problem yields

$$K_{ij}^* = \left(\frac{\rho - 1}{\rho}\right) \frac{\alpha_{K_j} \left(1 - \tau_{Y_{ij}}\right) P_{ij} Y_{ij}}{\left(1 + \tau_{K_{ij}}\right) r} \tag{12a}$$

$$L_{ij}^* = \left(\frac{\rho - 1}{\rho}\right) \frac{\alpha_{L_j} \left(1 - \tau_{Y_{ij}}\right) P_{ij} Y_{ij}}{w} \tag{12b}$$

$$M_{ij}^* = \left(\frac{\rho - 1}{\rho}\right) \frac{\alpha_{M_j} \left(1 - \tau_{Y_{ij}}\right) P_{ij} Y_{ij}}{\left(1 + \tau_{M_{ij}}\right) p_m} \tag{12c}$$

$$E_{ij}^* = \left(\frac{\rho - 1}{\rho}\right) \frac{\alpha_{E_j} \left(1 - \tau_{Y_{ij}}\right) P_{ij} Y_{ij}}{\left(1 + \tau_{E_{ij}}\right) p_e} \tag{12d}$$

$$Z_{ij}^* = \left(\frac{\rho - 1}{\rho}\right) \frac{\alpha_{Z_j} \left(1 - \tau_{Y_{ij}}\right) P_{ij} Y_{ij}}{\left(1 + \tau_{Z_{ij}}\right) p_z} \tag{12e}$$

Optimal quantities of factor inputs contain both output distortion and distortion in their respective factor markets. By combining equations (12a)–(12e) with the firm's objective function in equation (10), we can find the market clearing price for each firm:

$$P_{ij} = \left(\frac{\rho - 1}{\rho}\right) \left(\frac{MC}{\epsilon}\right) \frac{\left(1 + \tau_{K_{ij}}\right)^{\alpha_{K_j}} \left(1 + \tau_{M_{ij}}\right)^{\alpha_{M_j}} \left(1 + \tau_{E_{ij}}\right)^{\alpha_{E_j}} \left(1 + \tau_{Z_{ij}}\right)^{\alpha_{Z_j}}}{\left(1 - \tau_{Y_{ij}}\right) A_{ij}} \tag{13}$$

where

$$\epsilon = \prod_S \alpha_{S_j}^{\alpha_{S_j}}$$

$$MC = r^{\alpha_{K_j}} w^{\alpha_{L_j}} p_m^{\alpha_{M_j}} p_e^{\alpha_{E_j}} p_z^{\alpha_{Z_j}}$$

---

[3]It is important to note that, although the price of aggregated raw materials (M) produced in the manufacturing sector is determined using the profit maximization problem in section III.A.2, firms may consume different combinations (M_{ij}) of such raw materials for their production; hence, they take the price (p_m) as exogenously given.

Note that the firm-level price in equation (13) comprises the marginal cost of production, markup, distortions, and the reciprocal of firm-level productivity. Given the assumption that firms in an industry have the same factor shares and input costs, I can infer that in the absence of distortions, the price of each firm in an industry would have been inversely proportional to the TFP of the firm. This inference is in line with my conjecture that all firms in an industry will have the same revenue productivity in the absence of any misallocation in factor resources.

I define firm-level total revenue productivity as $TFPR_{ij} = P_{ij}A_{ij}$. Solving $TFPR_{ij}$ from equation (13) yields

$$TFPR_{ij} = \left(\frac{\rho-1}{\rho}\right)\left(\frac{MC}{\epsilon}\right)\frac{\left(1+\tau_{K_{ij}}\right)^{\alpha_{K_j}}\left(1+\tau_{M_{ij}}\right)^{\alpha_{M_j}}\left(1+\tau_{E_{ij}}\right)^{\alpha_{E_j}}\left(1+\tau_{Z_{ij}}\right)^{\alpha_{Z_j}}}{\left(1-\tau_{Y_{ij}}\right)}$$

(14)

Revenue productivity given by equation (14) is a measure of firm-level distortion. Variation in $TFPR_{ij}$ gives us the degree of misallocation in input and output markets. I build my empirical findings on this intuition and try to measure the extent of variation in firm-level revenue productivity in the presence of distortions.

I define the marginal revenue products of factor inputs for an industry as the weighted average of the value of firm-level marginal revenue products, where the weights are calculated as a share of a firm's output in the industry:

$$MRPS_j = \frac{P_S}{\sum_{i=1}^{N_j}\frac{\left(1-\tau_{Y_{ij}}\right)P_{ij}Y_{ij}}{\left(1+\tau_{S_{ij}}\right)P_jY_j}}$$

(15)

Recall that S consists of all factor inputs such as K, L, M, E, and Z. P_S denotes the corresponding factor prices r, w, p_m, p_e, and p_z, respectively, and $\tau_{S_{ij}}$ indicates the corresponding factor distortions relative to labor.

I define industry-level total factor revenue productivity ($TFPR_j$) to be proportional to the geometric average of the average marginal revenue products of factor inputs in the industry (given in equation [15]):

$$TFPR_j = \left(\frac{\rho-1}{\rho}\right)\left(\frac{MC}{\epsilon}\right)\left[\frac{1}{\sum_{i=1}^{N_j}\frac{\left(1-\tau_{Y_{ij}}\right)P_{ij}Y_{ij}}{\left(1+\tau_{K_{ij}}\right)P_jY_j}}\right]^{\alpha_{K_j}}\left[\frac{1}{\sum_{i=1}^{N_j}\frac{\left(1-\tau_{Y_{ij}}\right)P_{ij}Y_{ij}}{P_jY_j}}\right]^{\alpha_{L_j}}$$

$$\left[\frac{1}{\sum_{i=1}^{N_j}\frac{\left(1-\tau_{Y_{ij}}\right)P_{ij}Y_{ij}}{\left(1+\tau_{M_{ij}}\right)P_jY_j}}\right]^{\alpha_{M_j}}\left[\frac{1}{\sum_{i=1}^{N_j}\frac{\left(1-\tau_{Y_{ij}}\right)P_{ij}Y_{ij}}{\left(1+\tau_{E_{ij}}\right)P_jY_j}}\right]^{\alpha_{E_j}}\left[\frac{1}{\sum_{i=1}^{N_j}\frac{\left(1-\tau_{Y_{ij}}\right)P_{ij}Y_{ij}}{\left(1+\tau_{Z_{ij}}\right)P_jY_j}}\right]^{\alpha_{Z_j}}.$$

(16)

B.   Allocation of Factors in the Industry

I now solve for the allocation of factor resources for each industry. I aggregate factor resources used by all firms in an industry using their marginal products to get the following:

$$S_j = \sum_{i=1}^{N_j} S_{ij} = S \frac{\alpha_{S_j} \theta_j / MRPS_j}{\sum_{j=1}^{J} \alpha_{S_j} \theta_j / MRPS_j} \tag{17}$$

Recall that $S \in \{K, L, M, E, Z\}$. and $S = \sum_{j=1}^{J} S_j$ are aggregate supplies of factor inputs in the economy. Also recall that θ_j is the share of each industry in producing the final consumption good. Note that factor accumulations in each industry are affected by factor distortions only through the corresponding marginal revenue products. This result is due to the Cobb–Douglas aggregation at the industry level. Combining industry-level factor inputs (17) and revenue productivity (16), we can derive

$$P_j Y_j = TFPR_j K_j^{\alpha_{K_j}} L_j^{\alpha_{L_j}} M_j^{\alpha_{M_j}} E_j^{\alpha_{E_j}} Z_j^{\alpha_{Z_j}} \tag{18}$$

Combining industry price P_j from (9) and firm's price P_{ij} from (13) together with firm-level revenue productivity from (14), we can simplify

$$P_j = \left[\sum_{i=1}^{N_j} \left(\frac{TFPR_{ij}}{A_{ij}} \right)^{(1-\rho)} \right]^{\frac{1}{1-\rho}} \tag{19}$$

Equating (18) and (19), we get

$$Y_j = TFP_j K_j^{\alpha_{K_j}} L_j^{\alpha_{L_j}} M_j^{\alpha_{M_j}} E_j^{\alpha_{E_j}} Z_j^{\alpha_{Z_j}} \tag{20}$$

where

$$TFP_j = \left[\sum_{i=1}^{N_j} \left(\frac{A_{ij} TFPR_j}{TFPR_{ij}} \right)^{(\rho-1)} \right]^{\frac{1}{1-\rho}} \tag{21}$$

Hence, the total factor productivity of each firm is a function of the firm-level TFP, TFPR, and industry-level revenue productivity. Now, we can write the final consumption outcome of the economy as follows:

$$C^* = \prod_{j=1}^{J} \left(TFP_j K_j^{\alpha_{K_j}} L_j^{\alpha_{L_j}} M_j^{\alpha_{M_j}} E_j^{\alpha_{E_j}} Z_j^{\alpha_{Z_j}} \right)^{\theta_S} \tag{22}$$

and the intermediate good of the economy as

$$M^* = \prod_{j=1}^{J} \left(TFP_j K_j^{\alpha_{K_j}} L_j^{\alpha_{L_j}} M_j^{\alpha_{M_j}} E_j^{\alpha_{E_j}} Z_j^{\alpha_{Z_j}} \right)^{\lambda_S} \tag{23}$$

Following Hsieh and Klenow (2009), I now assume that TFP (A_{ij}) and revenue productivity ($TFPR_{ij}$) are jointly log-normally distributed to depict the effect of firm-level distortion on the productivity of an industry. By this assumption, the logarithm of industry-level TFP can be expressed as

$$\log TFP_j = \frac{1}{1-\rho} \log \left(\sum_{i=1}^{N_j} A_{ij}^{(\rho-1)} \right) - \frac{\rho}{2} Var \left(\log TFPR_{ij} \right) \tag{24}$$

Equation (24) shows that factor distortions reduce overall productivity of an industry through the variance of firm-level TFPR. On the basis of this finding, I will now proceed to show how factor distortions contribute to firm-level TFPR variation. Note that I assume that the number of firms are unaffected by factor market distortions. This assumption is elaborated in more detail in Hsieh and Klenow (2009).

IV. Data

This study uses data on the formal manufacturing sector from the Annual Survey of Industries (ASI) collected by the Central Statistical Office of India. ASI is the primary source of industrial statistics in India, which covers all factories as defined in the Factories Act of 1948. ASI data is an annual survey of formal manufacturing firms with more than 50 workers and a random one-third sample survey of firms with more than 10 workers (with electricity) or firms with more than 20 workers (without electricity). I use the 62nd round of ASI data collected in the survey year 2005–2006.

I also take into account data for the unorganized manufacturing sector collected by the National Sample Survey Office (NSSO) of India for the survey year 2005–2006. The NSSO collects firm-level data for the informal manufacturing sector in India every 5 years. The dataset includes small manufacturing firms along with some service sector firms and some unincorporated proprietary firms. These firms are not registered under the Factories Act of 1948; hence, they are not included in the ASI data. The data for the informal sector consists of a large number of firms that use one or two workers. These firms have missing values for most of the variables I take into consideration. Also, they contribute a very small percentage of total value added.

Table 1 summarizes the distribution of informal firms and the corresponding cumulative percentages of contributions to total value added, according to the number of employees. There are over 30,000 one-employee firms that contribute only 1.6% of total value added and almost none of them have data for labor and capital in the corresponding dataset. In my analysis, I do not include such firms. I only consider informal firms that have at least six employees, a cutoff set on the basis of these firms' substantial market share. To keep the two datasets comparable,

Table 1.   **Informal Firms Distribution**

No. of Employees	No. of Firms	Cumulative % of Value Added
1	31,874	1.6
2	23,734	4.4
3	9,468	6.9
4	4,658	9.1
5	2,601	11.6
6	1,948	17.8
7	1,349	21.8
8	1,054	26.0
9	732	31.0
10	648	37.0
11–20	1,712	57.4
21–30	342	62.8
31– 50	190	73.0
51–100	92	95.0
More than 100	61	100.0

Source: National Sample Survey Office (NSSO) survey of unorganized manufacturing enterprises, 2005–2006.

Table 2.   **Distribution of Firms: Annual Survey of Industries versus National Sample Survey Office**

ASI Data		NSSO Data	
No. of Employees	No. of Firms	No. of Employees	No. of Firms
1–10	4,663	6–10	3,512
11–20	6,818	11–20	1,194
21–50	6,771	21–50	397
51–100	3,719	51–100	77
101–500	7,648	101–500	31
More than 500	2,254	More than 500	5

ASI = Annual Survey of Industries, NSSO = National Sample Survey Office.
Sources: Government of India, Ministry of Statistics and Programme Implementation (2005–2006a and 2005–2006b).

I only consider manufacturing industries that are covered in both the ASI and NSSO datasets. The literature in the field (La Porta and Shleifer 2008) argues that informal sector firms are small and highly unproductive compared to formal sector firms. The assumption of monopolistic competition among firms in the model allows for firms with different levels of productivity to coexist in the market.

Table 2 shows the distribution of firms in the analysis. There are around 31,000 formal sector firms taken from the ASI data, whereas the number of informal sector firms from the NSSO data is around 5,000. For this analysis, I had to drop some observations from both sectors due to missing data. Formal firms consist of all sizes, while informal firms are mostly small. To simplify the analysis, I use 2-digit industry-level data developed in the National Industrial Classification (NIC)

system. I consider 23 different industries, including food and beverage, machinery and equipment, wood, paper, publishing, computing machinery, and others (see Table 3).

Hsieh and Klenow (2009) use the value-added method to measure productivity and distortion in capital and output. They did not incorporate raw materials, services, or energy inputs in the production function. I first replicate their results using the value-added method and then extend the model to incorporate intermediate inputs as factors of production. This extension will adopt the gross-output method. I use nominal revenue of the firm as the output variable.

Aside from firm revenue, the variables that I use for this analysis are the firm's industry (2-digit NIC), labor compensation, net book value of fixed capital stock, rent on capital, intermediate input costs, and fuel and energy costs. I assume that service input cost is the same as the residual cost. I use labor compensation including wages, bonuses, and benefits as proxy for labor input. Capital is measured by the average of net book value of capital at the beginning and end of the year. I deviate from Hsieh and Klenow (2009) and Chatterjee (2011), as well as other previous works based on the measurement of the rental cost of capital. Existing literature in this field uses an exogenous percentage of capital as rental cost, whereas I measure the same by variables such as rent on machinery, building, and land, interest paid on loans, and other miscellaneous capital cost, which are taken from the ASI data for the formal sector.

However, for informal firms, the NSSO data do not explicitly provide rent on capital. I measured rental cost from the residual of value added after subtracting total labor cost. The costs of raw materials and energy are calculated explicitly from the cost of inputs of production. Service input costs consist of transport and communication costs, insurance charges, license costs, and other operative expenses.

The elasticity of substitution (ρ) is assumed to be constant in the model. Based on the literature in this field, I assume the value of ρ to be 3. In most of my empirical analysis, I use factor shares from industries in the US as a benchmark to identify the effect of distortion on productivity. The factor shares data are from the KLEMS measures found in the National Income and Product Accounts industry database (2005) provided by the US Bureau of Labor Statistics.

V. Empirical Analysis

My identification strategy is similar to that of Hsieh and Klenow (2009) and Chatterjee (2011). I established identification of distortions based on the rationale that, in the absence of distortions, revenue factor shares of output will be proportional to the parameters α_{Kj}, α_{Lj}, α_{Mj}, α_{Ej}, and α_{Zj} in a market with monopolistic competition. Because I assume distortions in factor markets, the

Table 3. Factor Shares: India versus United States

NIC 2-Digit Code	Manufacturing Industry	Capital Share		Labor Share		Raw Material Share		Energy Share		Services Share	
		India	US	India	US	India	US	India	US	India	US
15	Food and beverage	0.03	0.12	0.05	0.15	0.83	0.57	0.03	0.03	0.06	0.13
16	Tobacco	0.02	0.12	0.17	0.15	0.62	0.57	0.02	0.03	0.18	0.13
17	Textiles	0.04	0.10	0.10	0.23	0.68	0.53	0.09	0.03	0.09	0.11
18	Wearing apparel	0.03	0.11	0.15	0.34	0.55	0.32	0.03	0.02	0.24	0.21
19	Leather	0.02	0.11	0.08	0.34	0.76	0.32	0.03	0.02	0.11	0.21
20	Wood	0.04	0.07	0.08	0.24	0.73	0.53	0.05	0.03	0.09	0.14
21	Paper	0.04	0.14	0.09	0.21	0.67	0.46	0.13	0.08	0.08	0.12
22	Publishing	0.04	0.05	0.15	0.32	0.58	0.39	0.03	0.02	0.20	0.23
23	Petroleum products	0.00	0.28	0.01	0.03	0.95	0.67	0.01	0.01	0.03	0.01
24	Chemical	0.03	0.19	0.07	0.17	0.77	0.40	0.06	0.06	0.08	0.19
25	Rubber and plastics	0.03	0.13	0.06	0.22	0.79	0.49	0.05	0.02	0.07	0.14
26	Nonmetallic minerals	0.05	0.16	0.09	0.29	0.50	0.31	0.26	0.08	0.11	0.16
27	Basic metals	0.04	0.14	0.06	0.20	0.74	0.39	0.11	0.10	0.06	0.17
28	Fabricated metals	0.03	0.14	0.09	0.31	0.74	0.37	0.03	0.02	0.12	0.17
29	Machinery and equipment	0.02	0.11	0.10	0.28	0.73	0.45	0.02	0.01	0.13	0.15
30	Computing machinery	0.02	0.17	0.08	0.37	0.85	0.24	0.01	0.01	0.06	0.20
31	Electrical machinery	0.02	0.11	0.09	0.27	0.78	0.47	0.02	0.01	0.09	0.14
32	Communication machines	0.04	0.11	0.08	0.27	0.76	0.47	0.02	0.01	0.11	0.14
33	Medical instruments	0.04	0.17	0.13	0.33	0.66	0.31	0.02	0.01	0.16	0.18
34	Motor vehicles	0.01	0.10	0.07	0.25	0.82	0.51	0.02	0.01	0.08	0.14
35	Transport equipment	0.02	0.10	0.09	0.25	0.78	0.51	0.03	0.01	0.10	0.14
36	Furniture	0.02	0.11	0.05	0.31	0.85	0.40	0.01	0.01	0.07	0.17
37	Recycling	0.02	0.17	0.06	0.33	0.74	0.31	0.07	0.01	0.10	0.18

NIC = National Industrial Classification, US = United States.
Notes: For US factor shares, multifactor productivity and related KLEMS measures are from the National Income and Product Accounts industry database (using base year 2005). For India factor shares, data are from the Input–Output Transactions Table 2003–04.
Sources: US Bureau of Labor Statistics (2005); Government of India, Ministry of Statistics and Programme Implementation (2008); Government of India, Ministry of Statistics and Programme Implementation. http://mospi.nic.in/sites/default/files/main_menu/national_industrial_classification/nic_2004_struc_2digit.pdf.

revenue shares will give a biased estimation of the parameters. We can validate this from the first-order conditions of the firms:

$$\alpha_{Sj} = \left(\frac{\rho}{\rho - 1}\right) \frac{P_S S_{ij} \left(1 + \tau_{S_{ij}}\right)}{P_{ij} Y_{ij} \left(1 - \tau_{Y_{ij}}\right)} \qquad (25)$$

where $\alpha_{Sj} = \{\alpha_{Kj}, \alpha_{Lj}, \alpha_{Mj}, \alpha_{Ej}, \alpha_{Zj}\}$ and the respective $P_S = \{r, w, p_m, p_e, p_z\}$.[4] Recall that S consists of all factor inputs and $\tau_{S_{ij}}$ denotes corresponding distortions relative to the labor market. In the presence of distortions, I cannot distinguish the misallocation in resources from the bias in the parameters. Following Hsieh and Klenow (2009), I take into account US factor shares. The strategy is based on the assumption that US factor markets are less distorted than in India and the technology used in the industries is the same for both countries. A more detailed discussion on the assumptions are presented in Chatterjee (2011). Factor shares for both countries, described in Table 3, represent the average of the cost share for each factor in each industry.

Figure 1 illustrates the bias in factor shares in Indian industries with respect to benchmark US industries. Any deviation from the 45-degree line shows misallocation in the corresponding factor markets in India. I find a similar pattern in capital, labor, and raw material shares presented in Chatterjee (2011). It is evident from the figure that cost shares of capital, labor, and service inputs are significantly higher in the US than in India, whereas shares of raw materials and energy are higher in India. Next, I analyze within-industry variation in average revenue product of labor. Figure 2 illustrates the distribution of the logarithm of firm-level average revenue product of labor ($APRL$) relative to the industry mean, $\log \left(APRL_{ij}/APRL_j\right)$. I trim 1 percentile from both ends to avoid outliers. The horizontal axis shows $\log \left(APRL_{ij}/APRL_j\right)$, whereas the vertical axis measures the density of firms. There is a substantial variation in average revenue product of labor within an industry with a variance of 3.76.

A. Value-Added versus Gross-Output Approach

The goal in this section is to measure the variation in firm-level TFPR as an indicator of misallocation in factor markets. The variable of interest is the logarithm of firm-level TFPR relative to the industry TFPR, $\log \left(TFPR_{ij}/TFPR_j\right)$. I depict both value-added and gross-output approaches to measure TFPR. First, I replicate the results from Hsieh and Klenow (2009) using the value-added approach. They estimate the distribution of TFPR using formal manufacturing sector data for

---

[4]The assumption of common factor prices is a hypothetical frictionless situation, which allows me to identify the factor allocation distortion apart from any exogenous variation that may affect factor prices. As long as distortions exist, firms pay different factor prices, and the dispersion in distortions provides us with a sense of how much firms actually deviate from the equilibrium.

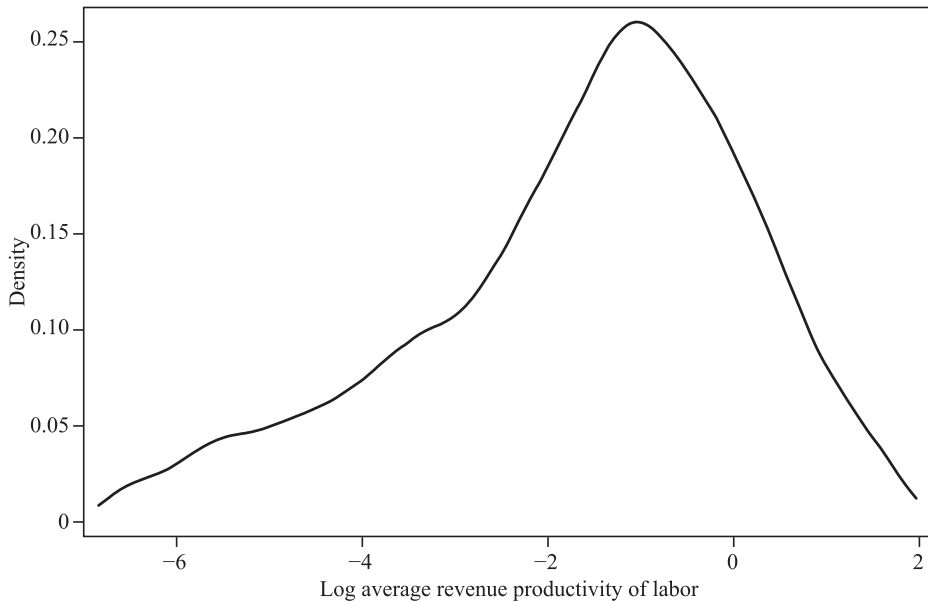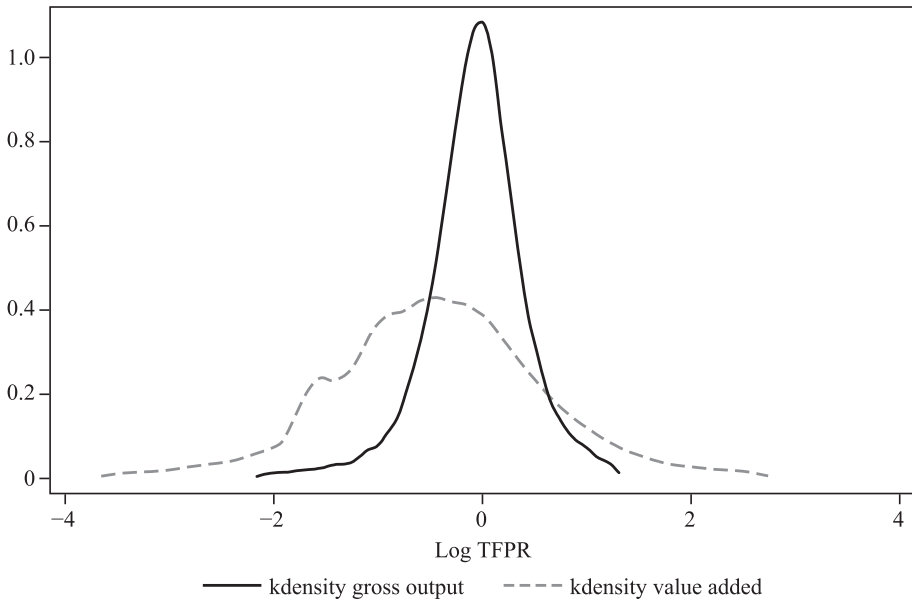Figure 1.   **Factor Shares for the United States and India**
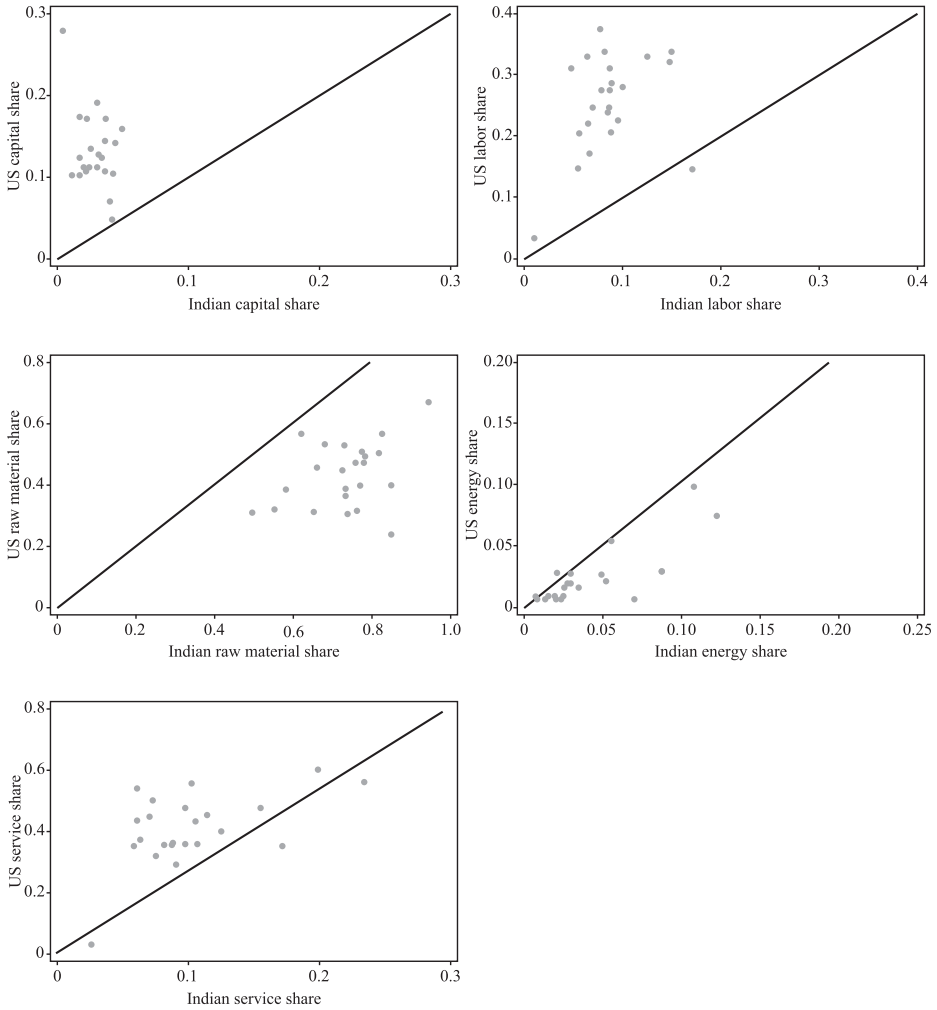
US = United States.
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b); and KLEMS measures from the National Income and Product Accounts of the US Bureau of Labor Statistics (2005).

1987–1988 and 1994–1995. I repeat their method using 2005–2006 data for both formal and informal sectors. I also illustrate the TFPR distribution using the gross-output method using the same data. Cobbold (2003) presented the formal relationship between value-added and gross-output TFP as

$$TFP_{VA} = \frac{G}{VA} \times TFP_{GO}$$

Figure 2. **Distribution of the Logarithm of Average Revenue Product of Labor**

Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

where G and VA represent nominal values of total revenue and total value added, respectively.

Several studies, such as Oulton and O'Mahony (1994) and van der Wiel (1999), show that productivity growth measured using value added is much higher than the measurement considering all inputs. It naturally follows from the above equation that given G and VA, TFP as well as TFPR measured using the value-added approach will be larger than if measured by the gross-output approach.

Before calculating the variance, I trim 1% tails of log ($TFPR_{ij}/TFPR_j$) to get rid of outliers. Figure 3 plots the distributions of the logarithm of TFPR relative to the industry mean. The dashed line shows the value-added TFPR distribution whereas the solid line shows the distribution using the gross-output approach. The variation in value-added TFPR is much higher than the variation in gross-output TFPR. Table 4 presents the TFPR dispersion statistics in firm-level TFPR. Standard deviation (SD) in value-added TFPR is around 0.99 compared to 0.47 using the gross-output approach. The difference in both methods is more pronounced when the variation in TFPR is estimated at higher percentiles.

Table 5 shows the dispersion in the logarithm of TFPR in Hsieh and Klenow (2009) using the value-added approach and the same variable in Chatterjee (2011) using the gross-output approach. The results display a larger value-added SD than

Figure 3. **Distribution of the Logarithm of Total Factor Revenue Productivity**

kdensity = kernel density, TFPR = total factor revenue productivity.
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

Table 4. **Dispersion of the Logarithm of Total Factor Revenue Productivity**

Statistics	Value Added	Gross Output
Standard deviation	0.99	0.47
75th to 25th percentiles	1.23	0.51
90th to 10th percentiles	2.45	1.08

Note: The variable is log (TFPRij/TFPRj).
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b); and KLEMS measures from the National Income and Product Accounts of the US Bureau of Labor Statistics (2005).

Table 5. **Dispersion of the Logarithm of Total Factor Revenue Productivity in the Literature**

Statistics	Hsieh–Klenow (1994–1995)	Chatterjee (2004–2005)
Standard deviation	0.67	0.49
75th to 25th percentiles	0.81	0.56
90th to 10th percentiles	1.60	1.19

Notes: Column 2 shows dispersion of total factor revenue productivity estimated by Hsieh and Klenow (2009) for 1994–1995 data using the value-added approach. Column 3 depicts the same variable estimated by Chatterjee (2011) for 2004–2005 data using the gross-output approach.
Sources: Hsieh and Klenow (2009) and Chatterjee (2011).

in Hsieh and Klenow (2009), who use the same approach with formal sector data from 1994 to 1995. This might be the consequence of an increase in overall level of misallocation in the last decade or inclusion of the informal sector in my analysis.

Furthermore, I find comparable results (shown in Table 4) with those of Chatterjee (2011) in the dispersion of gross-output TFPR. After including energy and service sector distortions, the SD in firm-level TFPR in my study dropped by 0.02 from an overall 0.49 as shown by Chatterjee using 2004–2005 data. The gap between the results is more conspicuous in the 75th to 25th percentiles and 90th to 10th percentiles.

B.   Decomposing the Misallocation in Factor Markets

I now turn to separating the effect of each component attributed to the variance of firm-level TFPR. Moving forward, only the gross-output approach will be considered. I take into account several types of distortions in input and output markets. The calculation for each type as a function of total revenue, cost of inputs, and factor shares is derived from first-order conditions of a firm as

$$1 - \tau_{Y_{ij}} = \left(\frac{\rho}{\rho - 1} \right) \ \frac{wL_{ij}}{\alpha_{L_j} P_{ij} Y_{ij}} \tag{26a}$$

$$1 + \tau_{K_{ij}} = \frac{\alpha_{K_j} wL_{ij}}{\alpha_{L_j} r K_{ij}} \tag{26b}$$

$$1 + \tau_{M_{ij}} = \frac{\alpha_{M_j} wL_{ij}}{\alpha_{L_j} p_m M_{ij}} \tag{26c}$$

$$1 + \tau_{E_{ij}} = \frac{\alpha_{E_j} wL_{ij}}{\alpha_{L_j} p_e E_{ij}} \tag{26d}$$

$$1 + \tau_{Z_{ij}} = \frac{\alpha_{Z_j} wL_{ij}}{\alpha_{L_j} p_z Z_{ij}} \tag{26e}$$

where all input market distortions are measured relative to the labor market. The intuition behind equations (26b)–(26e) is that, in the presence of distortions, input costs relative to labor compensation will be lower than given by the output elasticity. Equation (26a) demonstrates that a deviation of labor share from output elasticity with respect to labor will result in an output distortion.

To give a more elaborate presentation of the above result, I now find the variance of log ($TFPR_{ij}/TFPR_j$). The total misallocation is measured by the following variance:

$$Var\left[\log(TFPR_{ij}/TFPR_j)\right] = Var\left(D_K + D_L + D_M + D_E + D_Z - D_Y\right) \tag{27}$$

Table 6. **Variance Decomposition**

Component	Variance or Covariance
$Var(D_K)$	0.0472
$Var(D_L)$	0.0119
$Var(D_M)$	0.3490
$Var(D_E)$	0.0048
$Var(D_Z)$	0.0345
$Var(D_Y)$	1.1203
$Cov(D_K, D_L)$	0.0030
$Cov(D_K, D_M)$	0.0346
$Cov(D_K, D_E)$	0.0029
$Cov(D_K, D_Z)$	0.0082
$Cov(D_K, D_Y)$	0.0727
$Cov(D_L, D_M)$	−0.0026
$Cov(D_L, D_E)$	0.0001
$Cov(D_L, D_Z)$	0.0001
$Cov(D_L, D_Y)$	0.0067
$Cov(D_M, D_E)$	0.0129
$Cov(D_M, D_Z)$	0.0533
$Cov(D_M, D_Y)$	0.5716
$Cov(D_E, D_Z)$	0.0034
$Cov(D_E, D_Y)$	0.0302
$Cov(D_Z, D_Y)$	0.1110
$Var[log(TFPR_{ij}/TFPR_j)]$	0.2194

TFPR = total factor revenue productivity.
Note: The table shows variances and covariances of the components of log TFPR, where D_S ($S \in \{K, L, M, E, Z\}$) and D_Y are given by equations (28) and (29).
Source: Author's calculation based on the Annual Survey of Industries of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a); and KLEMS measures from the National Income and Product Accounts of the US Bureau of Labor Statistics (2005).

where

$$D_S = \alpha_{S_j} \log \left[\left(1 + \tau_{S_{ij}}\right) \sum_{i=1}^{N_j} \frac{\left(1 - \tau_{Y_{ij}}\right) P_{ij} Y_{ij}}{\left(1 + \tau_{S_{ij}}\right) P_j Y_j} \right] \tag{28}$$

$$D_Y = \log \left(1 - \tau_{Y_{ij}}\right) \tag{29}$$

Recall that S consists of all factor inputs such as $K, L, M, E,$ and Z. α_{S_j} denotes the corresponding factor shares, and $\tau_{S_{ij}}$ indicates the corresponding factor distortions. Also recall that I measure factor distortions relative to the labor market, implying $\tau_{L_{ij}}$ to be 0. In the above equations, D_S can be inferred as components of each factor input in the variance of TFPR. Table 6 describes the variance and covariances of each of the above components.

The variance of the components of equation (16) depict the contribution of factor distortions in explaining the variation in firm-level TFPR. Since I measure distortions in factor markets relative to the labor market in my analysis, the variance of D_L measures the variation in industry TFPR in the presence of only output distortions, multiplied by the cost share of labor. Moreover, the variance of D_Y determines the variation in firm TFPR attributed to only output distortion. Dispersions in D_Y and D_M are very high compared to the overall variance of $\log(TFPR_{ij}/TFPR_j)$, implying that misallocation is highest in output and raw materials.

Overall variance in $\log(TFPR_{ij}/TFPR_j)$ includes the pairwise covariance between the components of equation (16) as well. It is interesting to note that the covariance between output and raw material distortions is the highest (0.5716). This result may follow from the fact that in my framework, raw materials are endogenous, thus the output of one firm is used as raw materials in another.

Next, I examine the distinct effect of each distortion on the logarithm of TFPR relative to the industry mean. In Figure 4, the solid lines illustrate the distribution of the variable of interest, taking one factor distortion at a time. The dashed line represents the actual firm-level TFPR distribution taking all distortions together. The top panels of Figure 4 show TFPR distributions taking either output or capital distortion. Similarly, the middle panels and bottom panel depict scenarios with only raw material, energy, or service input distortions, respectively.

In the absence of any distortion, I expect the TFPR of all firms to equalize, which should reflect in a distribution that shows a vertical line in a graph centered at 0. Any deviation from such a line shows signs of distortion. Taking one factor distortion at a time facilitates the comparison between the contribution of each factor input distortion toward the overall distortion in the market.

Since higher dispersion in the distribution shows higher distortion, it is perceptible from Figure 4 that output and raw material distortions play the main role in the overall distortion within an industry. Capital and service input distortions contribute a modest share in the measurement of misallocation. Energy distortion is almost negligible. These results emphasize the findings in Table 6.[5]

The intriguing observation from Figure 4 is that the misallocation in TFPR is lower when all factor market distortions are considered than when considering only output distortion or only raw material distortion. Such findings imply that factor input distortions offset each other's effects in describing total misallocation. This is an area I would like to work on in the future in order to understand the underlying intuition.

---

[5]Since the distortion in other markets are measured relative to the labor market, I cannot estimate the magnitude of the distortion in the labor market. The Appendix shows factor market distortions in different sectors similar to Figure 4, relative to the energy input market.

Figure 4.   **Distribution of Firm-Level Total Factor Revenue Productivity Taking One Distortion at a Time (relative to the labor market)**
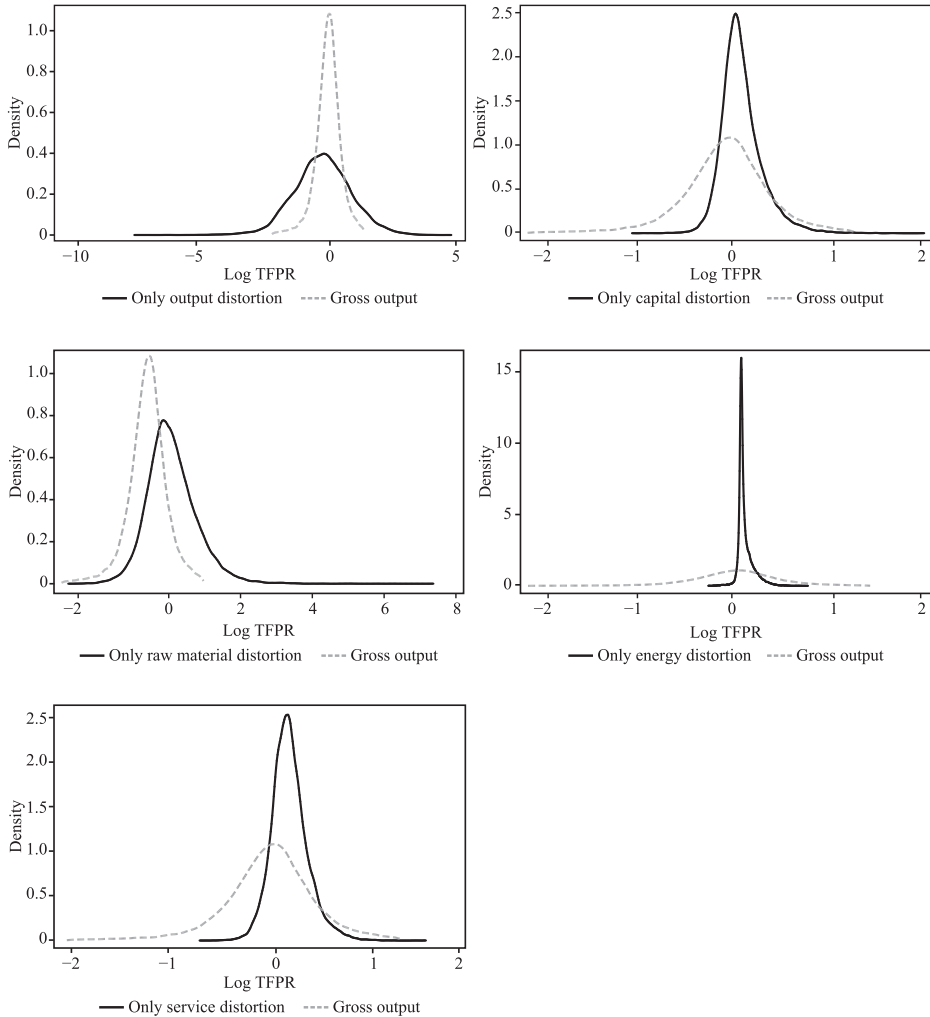
TFPR = total factor revenue productivity.
Note: All TFPR distributions are in logarithm and relative to the industry average.
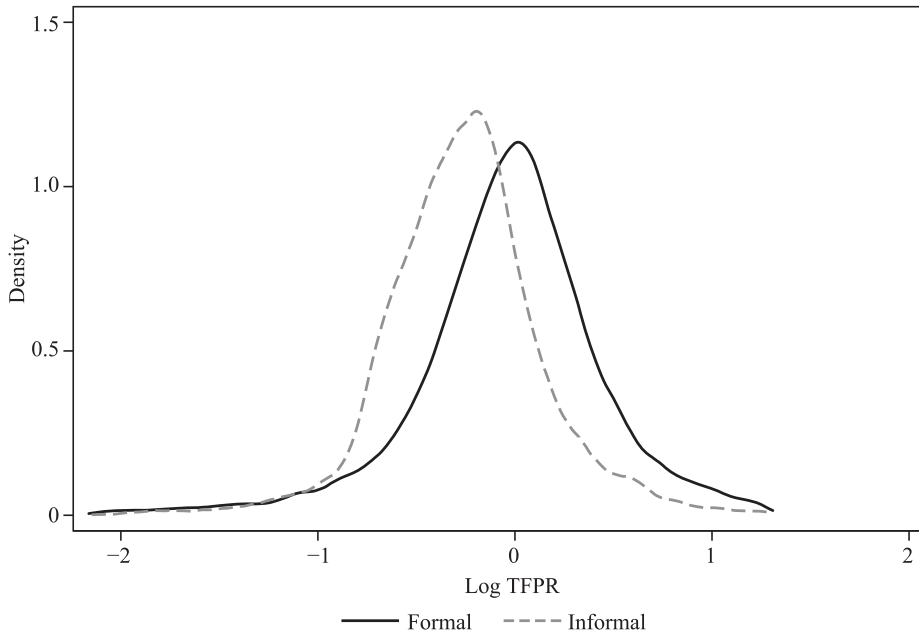Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

C.   Formal versus Informal Sectors

Since I use data from both formal and informal sector firms in my empirical analysis, it is important to examine if there is any inherent difference in the pattern of input and output distortions between these two sectors. Figure 5 illustrates the

Figure 5. **Distribution of Firm-Level Total Factor Revenue Productivity in Formal and Informal Sectors**

TFPR = total factor revenue productivity.
Note: All TFPR distributions are in logarithm and relative to the industry average.
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

distribution of TFPR relative to the industry mean for firms in the formal and informal sectors. The solid line represents formal sector firms, whereas the dotted line represents firms in the informal sector. The dispersions in the distributions are similar in both sectors while the mean of the distribution is higher in the formal sector than in the informal sector.

Next, I examine the distribution of factor distortions separately for both sectors. Figure 6 shows the distribution of output and input distortions for formal and informal firms. The distributions show that in the informal sector, output, raw material, energy, and service input distortions are a little higher than in the formal sector. Capital distortion on the other hand is more dispersed and higher in the formal sector.

VI. Misallocation within the Manufacturing Sector

Misallocation may vary between industries within the manufacturing sector. Therefore, it is useful to look further into the distribution of TFPR in separate

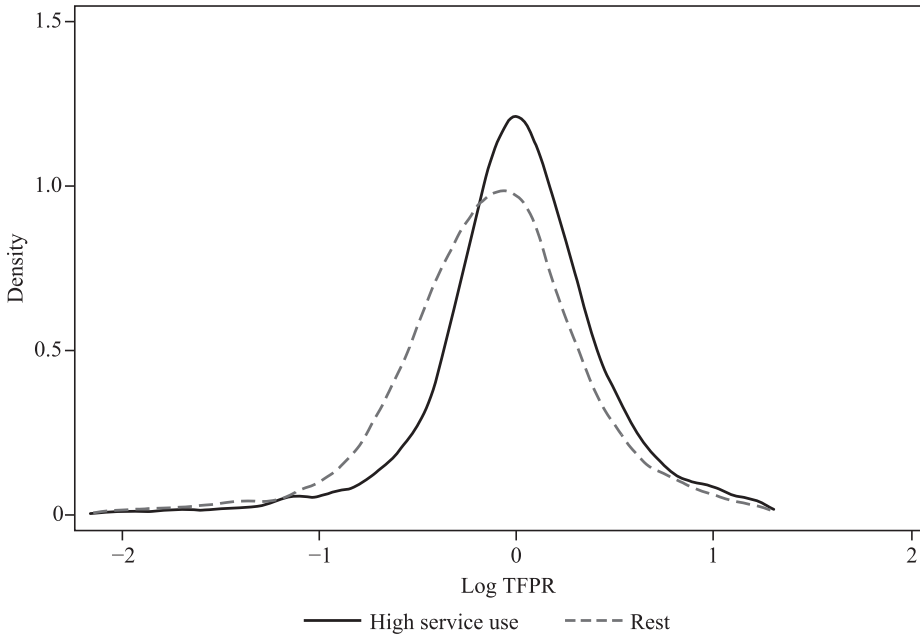Figure 6. **Distribution of Factor Market Distortions in Formal and Informal Sectors**
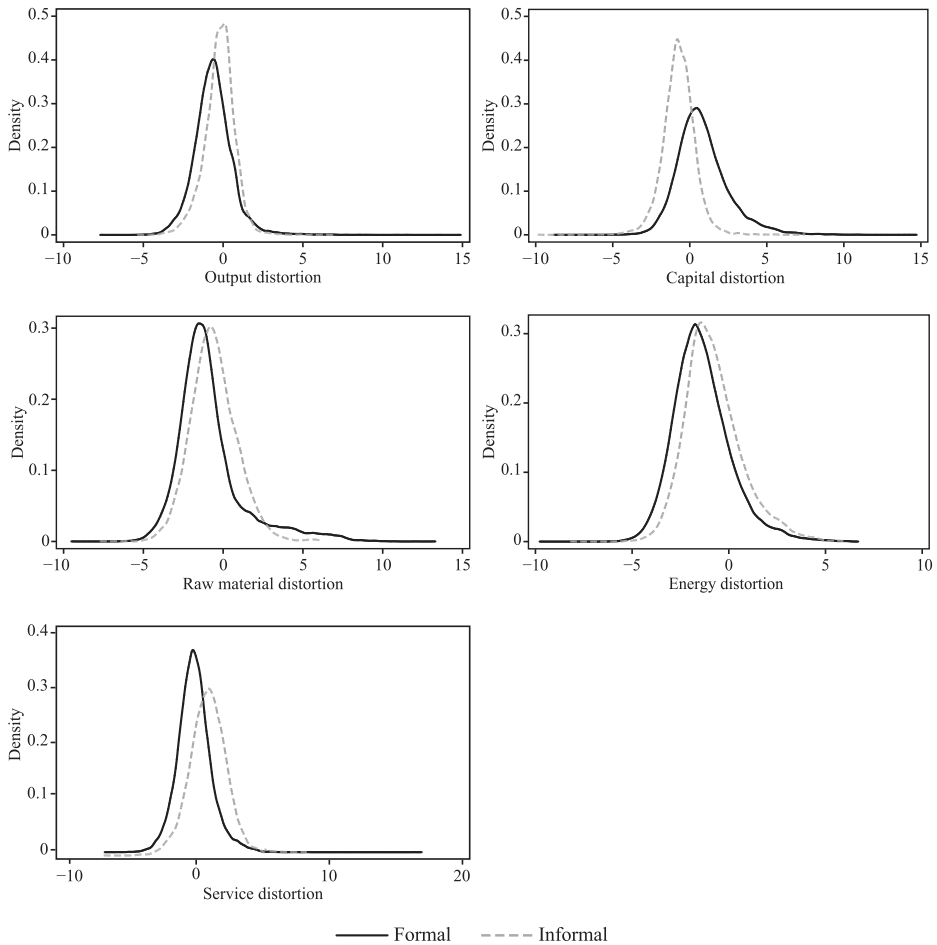
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

groups of industries to infer any inherent pattern of misallocation that may exist within the manufacturing sector.

A.   According to Use of Service Inputs

Services include a vast range of inputs used by manufacturing sector firms. Also, manufacturing industries vary widely in their use of such inputs. Using manufacturing firm data from the Czech Republic, Arnold, Javorcik, and Mattoo (2011) show that the productivity of manufacturing industries that rely extensively on service inputs is affected more by reforms in the service sector. To

Figure 7. **Distribution of Firm-Level Total Factor Revenue Productivity according to Intensity of Use of Service Inputs**

TFPR = total factor revenue productivity.
Note: All TFPR distributions are in logarithm and relative to the industry average.
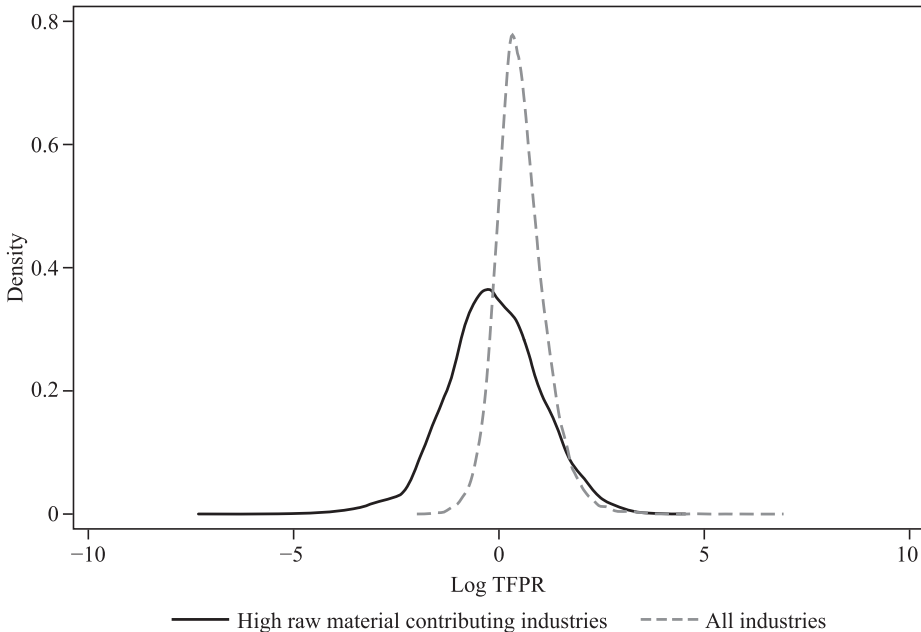Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

investigate if such a connection exists in Indian manufacturing, I use the 2003–2004 input–output table provided by the Ministry of Statistics and Programme Implementation of the Government of India to rank the industries according to their use of services (Government of India, Ministry of Statistics and Programme Implementation 2008). Five industries that used more than 55% of the total service inputs used by the manufacturing sector are food and beverage, basic metals, chemical products, wearing apparel, and electric machinery. Figure 7 shows the TFPR distribution of these five industries compared to others. The dispersion in TFPR is lower in the industries that use service inputs more intensively, reflecting a lower misallocation in these industries.

B.   According to Raw Materials Contribution

Variance decomposition of the misallocation in factor markets in section V.B. revealed that not only does raw material distortion play a vital role in explaining overall misallocation in TFPR, it is also highly correlated with output

Figure 8. **Distribution of Firm-Level Total Factor Revenue Productivity for All Industries (only raw material distortion) and Distribution of Total Factor Revenue Productivity for High Raw Material Contributing Industries (only output distortion)**

TFPR = total factor revenue productivity.
Note: All TFPR distributions are in logarithm and relative to the industry average.
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).

distortion. Since raw materials in the model are produced by the manufacturing firms themselves, raw material distortion may partly reflect output distortion in the industries supplying raw materials. To investigate the source of such a distortion, I categorize industries according to their raw materials contribution, using the input–output table mentioned previously. Five manufacturing sector industries that contributed almost 63% of raw materials to the same sector are food and beverage, textile, petroleum products, basic metals, and chemical and chemical products.

Next, I compare the output distortion in industries that contribute a large share of raw materials with the raw material distortion for all manufacturing sector firms. The solid line in Figure 8 shows the TFPR distribution with only output distortion using firms in the top five raw materials contributing industries, while the dashed line shows the TFPR distribution with only raw material distortion using all firms in the manufacturing sector. The former distribution is more skewed than the latter, which suggests that raw material distortion in the manufacturing sector may be partly reflecting output distortions in the industries contributing most to

Table 7. **Regression of Firm Size on Distortion**

	Log of Distortions				
	Output (1)	**Capital** (2)	**Raw Materials** (3)	**Energy** (4)	**Services** (5)
Log (*labor*)	0.075**	0.274**	0.044**	0.190**	0.076**
	(0.005)	(0.006)	(0.008)	(0.005)	(0.005)
Industry effect	Yes	Yes	Yes	Yes	Yes
Ownership effect	Yes	Yes	Yes	Yes	Yes
Organization effect	Yes	Yes	Yes	Yes	Yes
Region effect	Yes	Yes	Yes	Yes	Yes
No. of observations	41,237	44,726	45,589	47,755	47,829

Notes: The dependent variables in the regressions are the logarithm of output and input (capital, raw materials, energy, and services, respectively) distortions. Standard errors in parentheses. ** shows p-value $< .01$.

Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b); and KLEMS measures from the National Income and Product Accounts of the US Bureau of Labor Statistics.

manufacturing raw materials. This finding suggests that policies that can reduce output distortion in industries that supply the lion's share of manufacturing raw materials should result in lower raw material distortion in the overall manufacturing sector.

VII. Misallocation and Firm Size

There is a body of literature on the sources of factor distortions. Banerjee and Duflo (2005) discovered that capital market distortions might be originating from disparities in credit policy. Chatterjee (2011) mentions unavailability of raw materials as a reason behind intermediate input distortions. Bhidé (2008) shows that in a developing country such as India, electricity connection from private and public enterprises might cause a distortion in energy prices. Hsieh and Klenow (2009) argue that government policy, especially size restrictions, might prevent firms from achieving an optimal scale, thereby creating an output distortion. They also considered firm size as an explanation for TFPR dispersion within an industry. Ha, Kiyota, and Yamanouchi (2016) show a nonlinear relationship between firm employment size and factor market distortions in the context of manufacturing firms in Viet Nam. I now proceed to examine the relationship between firm size and distortion in factor markets.

Table 7 presents regression coefficients from estimating this relationship. I use the logarithm of total labor employed as a measure of firm size. Column (1) uses the logarithm of firm-level output distortion as the dependent variable. Similarly, the dependent variables for columns (2), (3), (4), and (5) are the logarithms of firm-level capital, raw material, energy, and service input distortions, respectively. I control for industry fixed effects, ownership type (private, central government owned,

state government owned, etc.), type of organization (individual proprietorship, partnership, co-operative society, etc.), and location of the firms.

I find a positive relationship between firm size and each type of distortion.[6] Smaller firms in the formal or informal sector might be able to avoid some policy restrictions, unlike their larger counterparts. An assumption of monopolistic competition includes the provision of a markup in the model. Though I assume all firms in an industry have the same markup, larger firms might have greater market power and larger markups, which in turn will create more output distortion as well as raw material distortion. It will be fascinating to see the effect of firm size on distortion once we relax the assumption of a constant elasticity of substitution within an industry.

VIII. Conclusion

I measure the aggregate misallocation in resources using firm-level data from both formal and informal manufacturing sectors in India for the survey year 2005–2006. I include energy distortion and service input distortion to extend existing research such as those by Hsieh and Klenow (2009) and Chatterjee (2011). The dispersion in TFPR within each industry is substantial, implying misallocation caused by distortion of factor resources. While energy distortion does not contribute much to aggregate misallocation, the effect of service sector input distortion is more pronounced. I further decompose the variance of TFPR to find the effect of each factor market distortion separately. I discover that output distortion and raw material distortion contribute the largest share in aggregate misallocation. Reallocation of such factors within industries should result in the largest gain in TFP. Moreover, I find a high level of covariance between output and raw material distortion which, along with a further exploration within the manufacturing sector, suggests that some of the distortion in raw materials may reflect the output distortion in industries producing a larger share of the raw materials. I also uncover a puzzling result that the inclusion of many factor distortions together offset each other's effects and results in a lower aggregate misallocation. Although unexpected, this result may inspire further research in this field.

References

Arnold, Jens, Beata Javorcik, and Aaditya Mattoo. 2011. "The Productivity Effects of Services Liberalization: Evidence from the Czech Republic." *Journal of International Economics* 85 (1): 136–46.

---

[6]Note that the distortions are calculated using equations (26a)–(26e). Since the sign of output distortion is inverse in equation (26a), a higher numerical value of output distortion shows a lower magnitude of the same.
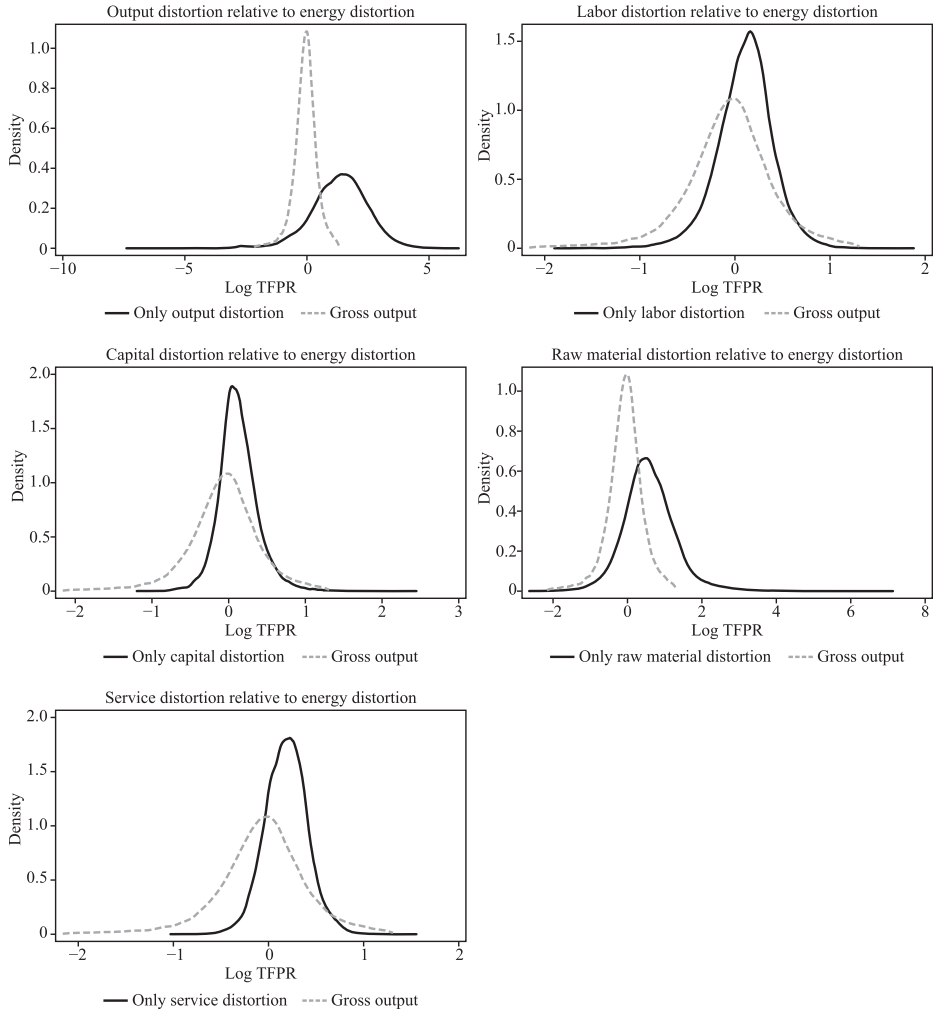
Banerjee, Abhijit V., and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." In *Handbook of Economic Growth Volume 1, Part A*, edited by Philippe Aghion and Steven N. Durlauf, 473–552. Amsterdam: North Holland-Elsevier.

Bhidé, Amar. 2008. "What Holds Back Bangalore Businesses?" *Asian Economic Papers* 7 (1): 120–53.

Camacho, Adriana, and Emily Conover. 2010. "Misallocation and Productivity in Colombia's Manufacturing Industries." Inter-American Development Bank Working Paper Series.

Chanda, Rupa, and Pralok Gupta. 2011. "Service Sector Liberalisation in India: Key Lessons and Challenges." South African Institute of International Affairs (SAIIA) Occasional Paper No. 88.

Chatterjee, Urmila. 2011. "Resource Allocation and Efficiency in Developing Countries." PhD Dissertation, University of California Berkeley. https://escholarship.org/uc/item/6458x094.

Cobbold, Trevor. 2003. "A Comparison of Gross Output and Value-Added Methods of Productivity Estimation." Productivity Commission Research Memorandum. Canberra, Australia.

Dias, Daniel A., Carlos Robalo Marques, and Christine Richmond. 2016. "Misallocation and Productivity in the Lead Up to the Eurozone Crisis." *Journal of Macroeconomics* 49: 46–70.

Fernald, John, and Brent Neiman. 2011. "Growth Accounting with Misallocation: Or, Doing Less with More in Singapore." *American Economic Journal: Macroeconomics* 3 (2): 29–74.

Government of India, Ministry of Statistics and Programme Implementation. http://mospi.nic.in/sites/default/files/main_menu/national_industrial_classification/nic_2004_struc_2digit.pdf.

_____. 2005–2006a. Annual Survey of Industries. http://www.csoisw.gov.in/cms/en/1023-annual-survey-of-industries.aspx (accessed June 1, 2014).

_____. 2005–2006b. National Sample Survey Office. http://mospi.nic.in/NSSOa (accessed June 1, 2014).

_____. 2008. Input–Output Transactions Table 2003–04. http://mospi.nic.in/publication/input-output-transactions-table-2003-04 (accessed August 1, 2014).

Gullickson, William, and Michael J. Harper. 1999. "Possible Measurement Bias in Aggregate Productivity Growth." *Monthly Labor Review* 122 (2): 47–67.

Ha, Doan Thi Thanh, Kozo Kiyota, and Kenta Yamanouchi. 2016. "Misallocation and Productivity: The Case of Vietnamese Manufacturing." *Asian Development Review* 33 (2): 94–118.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *The Quarterly Journal of Economics* 114 (1): 83–116.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124 (4): 1403–48.

Hulten, Charles R., Edwin R. Dean, and Michael J. Harper. 2001. "Total Factor Productivity: A Short Biography." In *New Developments in Productivity Analysis*, 1–54. Chicago: University of Chicago Press.

Kalemli-Ozcan, Sebnem, and Bent E. Sørensen. 2014. "Misallocation, Property Rights, and Access to Finance: Evidence from within and across Africa." In *African Successes, Volume III: Modernization and Development,* edited by Sebastian Edwards, Simon Johnson, and David N. Weil, 183–211. Chicago: University of Chicago Press.

Klenow, Peter J., and Andres Rodriguez-Clare. 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual, Volume 12*, edited by Ben S. Bernanke and Julio Rotemberg, 73–103. Cambridge, MA: MIT Press.

La Porta, Rafael, and Andrei Shleifer. 2008. "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity* 47 (1): 123–35.

Midrigan, Virgiliu, and Daniel Y. Xu, 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review* 104 (2): 422–58.

Organisation for Economic Co-operation and Development. 2001. *Measuring Productivity OECD Manual: Measurement of Aggregate and Industry-Level Productivity Growth.* http://www.oecd.org/sdd/productivity-stats/2352458.pdf.

Oulton, Nicholas, and Mary O'Mahony. 1994. *Productivity and Growth: A Study of British Industry 1954–86.* National Institute of Economic and Social Research Occasional Papers, Volume 46. Cambridge, UK: Cambridge University Press.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics* 11 (4): 707–20.

Thakur, Tripta, S.G. Deshmukh, S.C. Kaushik, and Mukul Kulshrestha. 2005. "Impact Assessment of the Electricity Act 2003 on the Indian Power Sector." *Energy Policy* 33 (9): 1187–98.

US Bureau of Labor Statistics. 2005. National Income and Product Accounts Industry Database. https://www.bls.gov/mfp/mprdload.htm (accessed June 1, 2014).

van der Wiel, Henry P. 1999. "Sectoral Labour Productivity Growth." Research Memorandum No. 159, CPB Netherlands Bureau for Economic Policy Analysis.

Appendix

Distribution of Firm-Level Total Factor Revenue Productivity Taking One Distortion at a Time (relative to the energy sector)

TFPR = total factor revenue productivity.
Note: All TFPR distributions are in logarithm and relative to the industry average.
Source: Author's calculation based on the Annual Survey of Industries and the National Sample Survey Office of the Government of India, Ministry of Statistics and Programme Implementation (2005–2006a, 2005–2006b).