Misallocation, Labor Informality and TFP: Theory & Evidence from the Formal Indian Manufacturing Sector

Samiran Dutta

Supervisor: Prof. François Fontaine
Referee: Prof. Tobias Broer

June 2023

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

Master's Thesis
Analysis and Policy in Economics
Paris School of Economics

Misallocation, Labor Informality and TFP: Theory & Evidence from the Formal Indian Manufacturing Sector*

Samiran Dutta

Abstract

This dissertation aims to re-evaluate the question of resource misallocation in the context of labor informality. I present a theoretical framework of monopolistic competition among heterogeneous firms using two types of labor, capital and intermediate inputs. The firms face idiosyncratic distortions to factor prices, generating misallocation of resources. The model predicts that the ratio of informal to formal labor increases in the presence of labor regulations. This effect is further augmented by a higher elasticity of substitution between the two. Owing to this flexibility, larger firms facing high regulatory costs to formal employment can easily sidestep them by hiring informally. Taking the model to plant-level data on formal Indian manufacturing, I find that aggregate misallocation has declined over 2010-11 to 2018-19. Majority of this decline was a construct of high labor informality, which reduced the relevance of labor misallocation for larger units. Most of the bite from factor misallocation is faced by small and mid-sized units that suffer from high capital and intermediate input distortions, respectively. On the other hand, larger units benefit from lower capital and intermediate input costs, in addition to hiring informally, allowing them to produce more than they should under an efficient allocation. Due to labor informality, output gains from removing size-based labor regulations does not substantially benefit manufacturing output. Instead, removing distortions in intermediate input lead to the highest gains. There are two main takeaways – First, policymakers aiming to reduce formal labor misallocation in developing economies cannot ignore the phenomenon of labor informality; Second, an isolated focus on labor regulations misses the importance of misallocation in other inputs.

JEL codes: D24, J46, O11, O14, O41, O47
Keywords: Misallocation, Productivity, Informality, Labor Regulations.

***Acknowledgments**: I would like to express my deepest gratitude to Prof. François Fontaine for his extremely valuable support and close mentoring throughout all stages of this dissertation. I also benefited greatly from the early comments of Dr. Sampreet S. Goraya on several methodological aspects. A special thanks goes to the members of the India KLEMS team who allowed me the independence to conduct research on growth and productivity in India, leading to my ideation behind this work. I dedicate this dissertation to the memory of Dr. Deb Kusum Das, the greatest influence in my research and life.

I Introduction

Misallocation of resources within an economy stands as a well documented phenomenon, with negative consequences for aggregate TFP and output. Following the seminal work of Hsieh and Klenow (2009) (H&K, henceforth), resources may not be allocated efficiently *across* firms, such that reallocating resources towards those who could use them to produce more will benefit aggregate outcomes. Given its parsimonious nature, the H&K framework has been widely applied to evaluate resource misallocation in developing economies. However, such studies often ignore a crucial feature of the developing world, namely – the issue of informality. In a more general context, the literature has ignored the fact that if the ease of substitution across inputs is high, those who face constraints in employing one input may shift to another to meet their production needs. Following these observations, this dissertation raises two main questions – *(i) Can we improve our understanding of misallocation by explicitly accounting for informality?* and *(ii) How does misallocation behave when there is an ease of substitution across resources?*

Informality is often associated with rather unromantic views. The work of Farrell (2004) and Levy (2008) suggests that the informal sector is 'parasitic' such that certain firms remain small and unregistered to avoid regulations, allowing them to compete with their formal counterparts. In the language of misallocation, the basic intuition is that resources can be freed from the informal sector and shifted to the more productive formal firms, raising aggregate output. This line of thought often misses the second dimension of informality. As highlighted by Ulyssea (2018), informality must be understood in two margins. First, firms may decide to stay unregistered (the *extensive* margin); Second, the formal sector firms can choose to hire informal workers as a recipe to sidestep labor regulations (the *intensive* margin).[1] This dissertation puts the intensive margin of informality at the center stage. Particularly, in the context of resource misallocation, I argue that policy implications generated by ignoring this margin might be misleading.

Suppose that a highly productive formal firm wishes to hire a certain number of workers in its quest to achieve an optimal scale of production. However, if this firm is hit by labor regulations, it will consequently hire less and thus produce less than it otherwise would. This will clearly imply losses in aggregate output. In this scenario, the flexibility afforded by the intensive margin of informality becomes crucial. If there is enough degree of substitutability across formal and informal workers, the constrained firms can opt to increase the hiring of the latter and achieve their desired scale of production. This would in turn imply gains in aggregate output, even in the presence of labor regulations. Using the H&K framework to study misallocation in developing economies, several studies have accounted for the extensive mar-

[1] Formal workers are defined as those who get in-kind benefits owing to labor regulations and cannot be fired instantaneously without adjustment costs, while the informal workers usually do not benefit from such regulations and can be fired instantaneously.

gin (Chatterjee, 2011; Misch and Saborowski, 2018; Kabiraj, 2020; Mohommad et al., 2021). However, the implications of the intensive margin are virtually absent in these studies.[2] There are two main reasons for this – First, not many developing countries report detailed data on the distinction of informal and formal employees hired by the registered firms. Second, even if such data exists, it is not immediate how to augment the original H&K framework with the intensive margin of informality. This dissertation is aimed at filling this gap in the existing literature. I propose an extension of the H&K framework by explicitly accounting for the degree of substitutability across both types of labor and across combined labor, capital and intermediate inputs.[3] This is done by positing a two-level CES specification for micro-level output, rather than working with the usual Cobb-Douglas specification originally used by H&K and the subsequent literature.[4] Accounting for the elasticity of substitution has important consequences for the extent of misallocation generated by this framework. Intuitively, if the more productive firms can easily substitute to informal labor to meet their labor requirements, the importance of the constraints imposed on hiring formal workers by labor regulations reduces. Subsequently, the contribution of labor misallocation towards the aggregate losses in output also reduces. The ease of substitution across other inputs also matters. For example, if a firm faces strong constraints in renting capital, but can easily shift to intermediate inputs or labor to meet its production needs, the relevance of capital misallocation decreases.

To empirically evaluate the model, I utilise plant-level data from the registered segment of manufacturing in India. The unique feature of this data is that it allows us to distinguish between both types of labor, in terms of the number of employees and wages. However, there is no free lunch – utilising data on only the registered plants implies that the I cannot analyse the extensive margin of informality. Chatterjee (2011) and Kabiraj (2020) use data from the unregistered manufacturing plants in India in their application of the H&K framework. However, I argue that this methodology is inherently flawed. The H&K model is based on representative agents at each level of production. This implies that both of these studies essentially assume the same production technology across both registered and unregistered plants, which is highly debatable. As noted before, I explicitly introduce formal and informal employees as separate inputs in my model, where the former benefits from strong labor regulations. Since unregistered plants do not have to abide by any regulations, it clearly does not make sense for the model to assume the same production technology across such units. The focus thus remains

[2]Mohommad et al. (2021) argues that reducing labor market rigidities in Indian states with high informality positively affects aggregate output. Informality is measured as the share of unregistered manufacturing plants in the states, completely missing the intensive margin. Moreover, the authors' use an index of Employment Protection Legislation (EPL) in their estimations. Such an index is based on inherently subjective assessments of differences in the text of the labor laws across states and is thus problematic (Amirapu and Gechter, 2020).

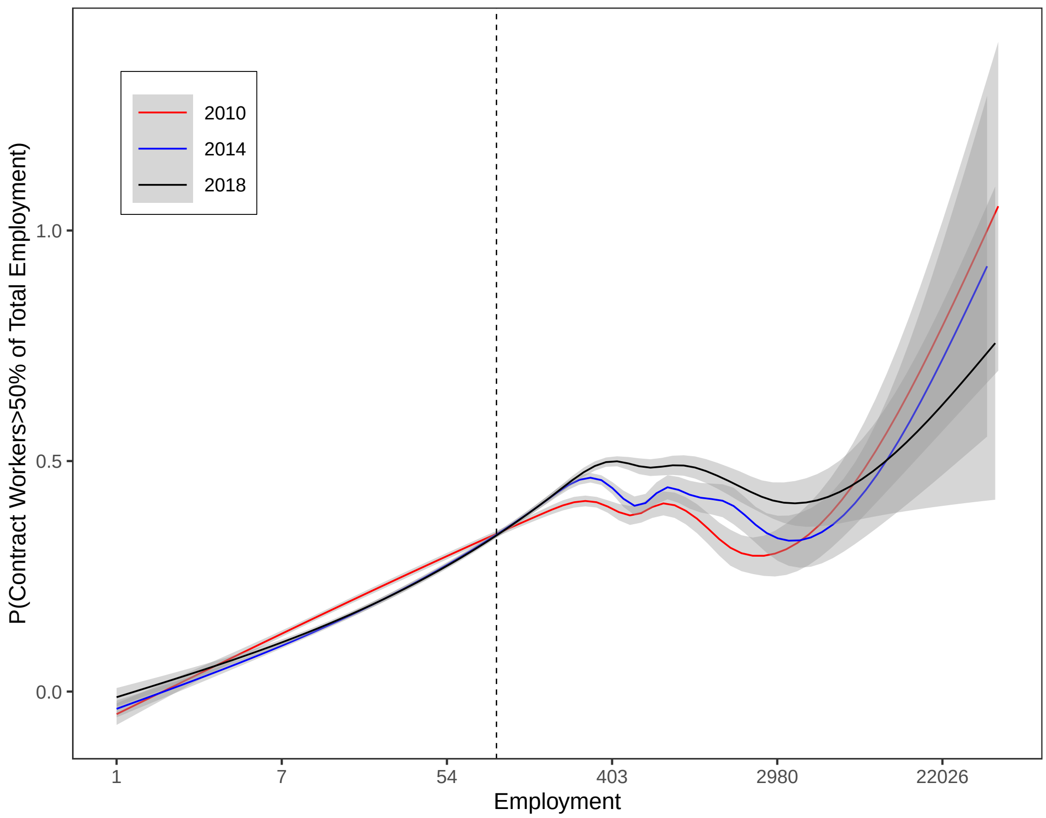
[3]The inclusion of intermediate inputs allows me to push the original value-added framework of H&K to a gross-output approach, widely held as the more appropriate method of TFP measurement.

[4]The Cobb-Douglas specification produces tractable mechanisms but operates on the restrictive assumption of unitary elasticity of substitution across inputs.

on sketching out the implications of the intensive margin of informality.

How relevant is the intensive margin in India's manufacturing story? Figure 1 illustrates this by evaluating the probability that informal labor accounts for more than 50% of total plant employment, as a function of total employment. It is evident that informal hiring has remained more or less stable in smaller plants. On the other hand, there appears to be an increase in informal hiring for plants employing more than 100 total workers.[5] The mechanism of sidestepping regulations by higher informal labor thus appears to be a relevant phenomenon for Indian manufacturing, meriting further investigation.[6]

Figure 1: Contract Labor Usage and Plant Size



Note: The plot shows point estimates and 95% confidence intervals from non-parametric regression of the probability a plant hires more than 50% of its non-managerial workers through contractors on (log) non-managerial employment). The vertical dashed line represents 100 employed persons, beyond which high adjustments costs of formal labor kick in (ref. Section III.2).

Having computed the 'elasticity robust' measure of misallocation, the model is able to generate counterfactual gains in aggregate TFP and output from a hypothetical reallocation of all re-

[5]Table 9 in Appendix A illustrates the same phenomena by reporting the average contract worker usage by plant-size bins, for each year.

[6]Bertrand et al. (2021) carry out an event study to further evince this phenomena, using an important Supreme Court judgement in 2001, saying that – there is no requirement of automatic absorption of contract workers in the permanent workforce. The authors' show that adoption of contract labor rose and constraints on large firms fell.

sources. In the spirit of a more disaggregated analysis, the aggregate gains are broken down by narrowly defined industry groups, as a guide to directing policy towards the industries suffering from high misallocation.[7] Additionally, by comparing the observed data and model-generated counterfactual, I provide evidence on which plants will increase and shrink their size under an efficient allocation of resources. This analysis essentially asks whether on an average, smaller or larger plants face relatively stronger constraints and should thus be receiving more attention from policymakers. In the context of Indian manufacturing, Banerjee and Duflo (2005) argue that policy tends to pamper the inefficient plants while hindering the most efficient ones. For 1994, H&K find that this is synonymous to saying – larger plants in India's formal manufacturing are more constrained while the smaller plants benefit from several implicit subsidies. I re-evaluate this contention in my framework with more recent data.

Lastly, and perhaps most importantly, the model allows the elimination of each factor price distortion at a time, such that we can gauge the relative importance of misallocation in each resource towards aggregate misallocation. Due to the separate introduction of both types of labor inputs, we can specifically back out a formal labor input distortion from the data. Reduced form evidence shows that this separation is important, as the estimated formal labor distortion is greatly in line with size-based labor regulations – a widely debated policy instrument in India. Such regulations are known to plague the larger plants when adjusting their formal labor input. As a result, policymakers often argue that these regulations need to be eased, to allow the existence of more large-sized plants in the economy. I build a counterfactual plant size distribution by removing the formal labor distortion in the model, in order to evaluate whether this vision would indeed come true. Further, to grasp the relative importance of this policy, I compare the reallocation gains from removing formal labor distortions with the gains implied by removing distortions in capital and intermediate inputs, one at a time.

I find that aggregate resource misallocation in formal manufacturing has decreased over the 2010-11 to 2018-19 period. Although declining in importance, misallocation appears to be substantially limiting manufacturing output. In particular, an efficient allocation of resources across plants, within each industry, could have increased manufacturing gross output by 63% in 2018-19. Several important industries seem to be constrained by an inefficient allocation, including – manufacturing of refined petroleum products, basic iron and steel, automobiles, textiles and pharmaceuticals. Given the high output share of these industries in the aggregate, this is particularly concerning. Dissecting the observed misallocation within industries, I find that on an average – small and medium sized plants face the strongest constraints in the input markets, relative to the larger plants. This implies that under a hypothetical efficient allocation, the relatively smaller plants will be able to increase their size, while the larger plants will shrink. Investigating the possible sources of misallocation across plants of different sizes,

[7] The H&K framework generates aggregate reallocation gains based on the weighted industry-level gains. It is thus surprising that papers utilizing this framework fail to report disaggregated results.

several important findings emerge. First, misallocation in intermediate inputs appears to be constraining majority of the plants, except those at the top of the size distribution. Moreover, the importance of such distortions appear to have increased for mid-sized plants. Second, capital misallocation only hinders the output of the smallest plants. Third, the larger plants appear to be constrained from size-based labor regulations, while the smaller plants benefit from lower costs of hiring formal labor. However, the larger plants relatively benefit more lower costs of hiring informal labor. Removing the size-based labor regulations, I find that the new plant size distribution will indeed exhibit more large sized plants, while the density of smaller plants reduces. This is in line with the expectations of policymakers aiming to ease such regulations, with the aim of promoting 'ease of business' for larger plants (Economic Survey of India, 2019-20). However, I show that the output gains from this exercise, although substantial in absolute terms, decline in relative importance over 2010-11 to 2018-19. In particular, removing size-based labor regulations would have increased manufacturing output by a towering 41% in 2010-11. More recently, in 2018-19, this policy benefits output by only 15%, while removing misallocation in intermediate inputs leads to a 31% increase in manufacturing output.

Why does the relative relevance of removing size-based labor regulations decrease? Reconciliation from the model shows that the equilibrium ratio of informal to formal labor unambiguously rises in response to higher regulatory costs of the latter. This effect is further augmented by a greater ease of substitution across both labor inputs. In essence, the larger plants constrained by such regulations, easily sidestep them by hiring informal labor to meet their total labor requirements. Removal of such regulations would reallocate more formal labor towards these plants in the hope to improve aggregate output. However, the resulting increase in output runs into diminishing returns, since the total labor input of the constrained plants is already high. Ignoring the intensive margin of informality would thus lead to incorrect expectations of policy outcomes from easing size-based labor regulations. The sidestepping mechanism subsequently reduces the contribution of formal labor misallocation towards aggregate decline in output. On the other hand, misallocation of capital and intermediate inputs maintain their relevance over the 2010-11 to 2018-19 period. In summary, a decade's worth of policy towards reducing misallocation has lead Indian manufacturing nowhere. Rather, the observed decline in aggregate misallocation is a construct of the intensive margin of informality.

This dissertation is structured as follows. Section II outlines the theoretical framework – extending the H&K model with weaker assumptions and an alternative path to deriving the gains from reallocation. Section III reports the details of the dataset, important labor regulations and calibrations. Section IV reports the empirical results of the model. Section V delves into the sources of misallocation and the relative importance of removing each type of distortion. Section VI presents various robustness checks. Finally, Section VII offers some concluding remarks. Several statistics and figures are relegated to Appendix A. The details of the theoretical proofs and derivations are outlined in Appendix B.

II Model

The theoretical framework is built on the seminal work of Hsieh and Klenow (2009), where plants are heterogeneous in productivity, facing exogenous idiosyncratic 'distortions' in factor prices, which in turn generates resource misallocation and limits the output of the economy. Although devoid of international trade, the model draws inspiration from Melitz (2003) in terms of heterogeneous plants facing monopolistic competition within industries. The main deviation from H&K comes from my assumptions on plant-level output. In particular, instead of assuming a unitary elasticity of substitution across inputs and hence working with a Cobb-Douglas specification, I introduce a more general two-level CES production function. A simple change in this specification allows me to grasp rich implications.

In particular, it allows me to explicitly capture the elasticity of substitution between formal and informal labor input, relevant for generating mechanisms that explain the 'sidestepping' of labor regulations. Further, I also capture the elasticity of substitution across combined labor, capital and intermediate inputs. This is particularly relevant when explaining whether misallocation is indeed an issue when all inputs are highly substitutable. Moreover, introducing both types of labor as separate inputs allows the model to capture a formal labor input distortion, which can be subsequently backed out from the data. This is important since the size-based labor regulations only apply to the formal labor input, and not labor input as a whole. The counterfactual exercise of removing these regulations should thus be based on this specific distortion alone.

The introduction of intermediate inputs in the production function allows me to use the gross-output approach, widely held as the more appropriate method when dealing with TFP measurement (Jorgenson et al., 1987; Oulton and O'Mahony, 1994; Jorgenson and Stiroh, 2001).[8] The analytical solutions of the model are inspired from the work of Dias et al. (2016, 2018), who take an alternative route compared to H&K, when arriving at the expression for the hypothetical reallocation gains. This method is much more tractable when working with elaborate production functions, such as the one I employ.

II.1 Environment

II.1.1 The Agents

The model's economy consists of three main agents – heterogeneous producers at the plant-level, a representative firm at the industry-level that combines plant-level output, and a representative final good producer that combines industry-level output. There are a total of S

[8] Gollop and Roberts (1979) remarks that value-added measure of productivity operates on an unrealistic assumption – technical change only affects the usage of capital and labour so that intermediate inputs cannot be the source of improvements in productivity. In my context, a reduction in misallocation of intermediate inputs will benefit aggregate TFP and output, and thus cannot be ignored.

industries, indexed by $s = 1, 2, .., S$ and a total of N_s plants within each industry s, indexed by $i = 1, 2, .., N_s$. The plants are assumed to be heterogeneous in their productivity, denoted by A_{si}.[9] Further, each plant i faces monopolistic competition within a given industry s, and produces an output denoted be Y_{si}. The plant-level output is further sold to a representative firm at the industry level, at price P_{si}. Industry output, denoted by Y_s, is then sold to a representative final goods producer at price P_s. The final goods producer behaves competitively and combines industry output into one final output, denoted by Y, reflecting the GDP of the model's economy. The supply of factors is assumed to be exogenous i.e., the supply of labor, capital and intermediate inputs is inelastic. The respective agents make optimal production decisions in a static setting without any uncertainty, such that each plant is always assumed to be in its long-run equilibrium in each period. Notably, the dynamics of entry and exit are not investigated i.e., producers do not respond to reallocation and corresponding changes in TFP.

II.1.2 Production Functions

Plant-level output is assumed to be produced using a constant returns to scale two-level CES technology[10] of formal labor ($H_{f,si}$), informal labor ($H_{c,si}$), capital input (K_{si}) and an aggregate (combining energy, material and services) of intermediate inputs (Q_{si}). The first-level CES captures the substitutability between formal and informal labor, with the elasticity of substitution parameter $\theta < 1$, and respective factor shares denoted by $\beta_{c,s}$ and $\beta_{f,s}$.

$$X_{si} = (\beta_c H_{c,si}^{\theta} + \beta_f H_{f,si}^{\theta})^{\frac{1}{\theta}} \tag{1}$$

Further, the second-level captures the substitutability across the combined labor input (X_{si}), capital and intermediate inputs, with the elasticity of substitution parameter $\psi < 1$ and the respective factor shares denoted by $\alpha_{L,s}$, $\alpha_{K,s}$ and $\alpha_{Q,s}$.

$$Y_{si} = A_{si} \left[\alpha_{L,s} X_{si}^{\psi} + \alpha_{K,s} K_{si}^{\psi} + \alpha_{Q,s} Q_{si}^{\psi} \right]^{\frac{1}{\psi}} \tag{2}$$

For further derivations and convenience of notation, let the elasticity of substitution in the first and second-level be denoted by $\nu = 1/(1-\theta)$ and $\eta = 1/(1-\psi)$, respectively. Combining both levels, the plant-level production function writes:

$$Y_{si} = A_{si} \left[\alpha_{L,s} (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s} K_{si}^{\psi} + \alpha_{Q,s} Q_{si}^{\psi} \right]^{\frac{1}{\psi}} \tag{3}$$

[9]The Hicks-neutral productivity parameter can be assumed to be drawn exogenously from a distribution $F(A_{si})$, without any consequence to the model's derivations and implications.

[10]This specification is borrowed directly from the work of Sato (1967), which provides an exposition on the more general CES nesting structures.

Notably, the subscript s on the factor shares denotes the assumption that revenue factor shares are assumed to be constant for all plants in a given industry s. The necessity of this rather strong assumption will be described later in Section II.3, and the way I make it less restrictive in Section III.2. Industry output is assumed to be a Dixit and Stiglitz (1977) style CES aggregate of the N_s differentiated plant-level output, as follows:

$$Y_s = \left(\sum_{i=1}^{N_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \qquad ;\sigma > 1 \tag{4}$$

Here, σ denotes the elasticity of substitution across plant-level output, and is assumed to be constant across all plants, implying no heterogeneity in markups.[11] The final good producer combines industry output using a Cobb-Douglas technology, with respective industry shares μ_s, as follows:

$$Y = \prod_{s=1}^{S} (Y_s)^{\mu_s} \qquad ;\sum_{s=1}^{S} \mu_s = 1 \tag{5}$$

II.2 Optimization Problems

The profit maximization problem of the representative final good producer, taking industry output price (P_s) as given, gives an expression for industry shares as follows:[12]

$$\max_{Y_s} \Pi_F = \prod_{s=1}^{S} Y_s^{\mu_s} - \sum_{s=1}^{S} P_s Y_s \implies \mu_s = \frac{P_s Y_s}{Y} \tag{6}$$

The demand curve for plant-level output, within each industry, is arrived at by solving the following industry-level profit maximization problem:

$$\max_{Y_{si}} \Pi_s = P_s \left(\sum_{i}^{N_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} - \sum_{i=1}^{N_s} P_{si} Y_{si} \qquad ;\forall s = 1, 2, .., S \tag{7}$$

The first-order condition gives the following expression for the plant-level demand curve:

$$P_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} \tag{8}$$

Where $P_s Y_s^{\frac{1}{\sigma}}$ is an industry level constant which can set equal to one without any consequence to the relative productivities and hence no implications on the intra-industry reallocation

[11] σ is assumed to be greater than one to ensure well-behaved convex preferences over the 'bundle' of plant-level output. Heterogeneous markups may endogenously generate misallocation (Peters, 2020). This is something I do not capture.

[12] The final good price P is assumed to be the numeraire.

exercise.[13] A formal proof of this assertion is relegated to Appendix B.6. Since the plant-level production function is built on four inputs, it is possible to separately identify distortions that affect the factor prices of all inputs simultaneously from distortions that affect the marginal product of one of the factors relative to the others. Using the absolute informal labor distortion as the base, I introduce four types of idiosyncratic distortions – an output distortions ($\tau_{Y_{si}}$), formal labor distortion ($\tau_{f_{si}}$), capital distortion ($\tau_{K_{si}}$) and an intermediate input distortion ($\tau_{Q_{si}}$). Taking into account the demand curve in equation (8), the resulting profit maximization problem of a given plant i in an industry s is given by:[14]

$$\Pi_{si} = \max_{H_{c,si},H_{f,si},K_{si},Q_{si}} (1-\tau_{Y_{si}})P_{si}Y_{si} - w_{c,s}H_{c,si} - w_{f,s}(1+\tau_{f_{si}})H_{f,si} - r_s(1+\tau_{K_{si}})K_{si} - q_s(1+\tau_{Q_{si}})Q_{si}$$
$$\text{s.t. } P_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} \tag{9}$$

Equation (9) deserves a brief discussion. Notably, the input prices are assumed to be constant for all plants within a given industry s.[15] The indirect approach to understanding resource misallocation i.e., keeping the form of $(1 + \tau_{si})$ unspecified for each input, captures several policies and implicit costs that affects the plants in an idiosyncratic manner, together in a single term. The output wedge captures any distortion that changes the marginal products of informal labor, formal labor, capital and intermediate inputs by the same proportion and be thought of as an implicit tax on the value of production. Subsequently, the formal labor, capital and intermediate input wedges capture the distortions that raise the marginal product of each input, relative to informal labor. For instance, in the context of labor regulations that affect the cost of hiring formal labor, a higher $\tau_{f_{si}}$ will capture the higher adjustment costs of formal labor relative to their informal counterparts. Plants with higher $\tau_{K_{si}}$ can be thought of as credit constrained due to non-competitive banking, while those with a lower $\tau_{K_{si}}$ might be benefiting from subsidised credit due to preferential access. Finally, plants with a higher $\tau_{Q_{si}}$ might be facing high supply chain distortions or a combination of distortions in the prices of energy, material and services.

Using a cost minimization approach to solve the problem in equation (9), I am able to pin down the plant-level pricing equation as a function of input prices, respective input shares, distortions, the productivity parameter and the markup. Let TC_{si} denote the total cost function.

[13] This assumption is similar to the one made in Hsieh and Klenow (2009) and Dias et al. (2016, 2018).

[14] An analogous characterisation is to think of distortions to each input in absolute levels, denoted by $(1+\tau^*_{si})$, and writing the profit function as:

$$\Pi_{si} = \max_{H_{c,si},H_{f,si},K_{si},Q_{si}} P_{si}Y_{si} - w_{c,s}(1+\tau^*_{c_{si}})H_{c,si} - w_{f,s}(1+\tau^*_{f_{si}})H_{f,si} - r_s(1+\tau^*_{K_{si}})K_{si} - q_s(1+\tau^*_{Q_{si}})Q_{si}$$

Assuming that $(1-\tau_{Y_{si}}) = 1/(1+\tau_{c^*_{si}})$ and $(1+\tau_{z_{si}}) = (1+\tau^*_{z_{si}})/(1+\tau^*_{c_{si}})$;$\forall z \in \{f, K, Q\}$, leaves the first-order conditions unchanged and is thus equivalent to working with equation (9).

[15] The relevance of this assumption and the attempt to make as less restrictive as possible, is discussed in Section II.3 and III.2, respectively.

Then the profit maximizing level of production will solve the following optimization problem:

$$\max_{Y_{si}} \ \Pi_{si} = (1 - \tau_{Y_{si}}) P_{si} Y_{si} - TC_{si} \tag{10}$$

$$\text{s.t. } P_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} \implies P_{si} Y_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{\frac{\sigma-1}{\sigma}} \tag{11}$$

The first-order condition gives the standard result of monopolistic competition, plant-level price is a markup over the marginal cost of production:

$$P_{si} = \left(\frac{\sigma}{\sigma - 1}\right) \frac{\partial TC_{si}/\partial Y_{si}}{(1 - \tau_{Y_{si}})} \tag{12}$$

This pricing equation can be made more explicit by solving for the marginal cost of producing a unit of Y_{si}. Since the plant-level production function is assumed to be a two-level CES, there is an implicit additive separability between each level. This separability implies that the equilibrium allocation of factors within each level is determined entirely by the relative prices of the given level only (Sato, 1967). This is important since it implies that the equilibrium ratio of informal to formal workers will only depend on their relative prices. As argued in the previous section, the choice of hiring informal labor appears to be motivated by the regulatory costs of formal labor. This idea can thus be captured by the modelling choice. The separability also implies that the cost minimization problem can be broken down into two stages. First, the plant decides on its optimal mix of formal and informal labor that minimizes the cost of production based only on the combined labor input (X_{si}). Second, the plant then decides on its optimal mix of combined labor input, capital and intermediate inputs, given the minimum cost of production from the first level. The first level cost minimization problem writes:

$$\min_{H_{c,si}, H_{f,si}} \quad w_{c,s} H_{c,si} + (1 + \tau_{f_{si}}) w_{f,s} H_{f,si} \tag{13}$$

$$\text{s.t. } (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{1}{\theta}} \geq X_{si}$$

Solving for the conditional factor demands of formal and informal labor, the resulting first-level cost function, denoted by $TC_{si}^{(1)}$, is given by:[16]

$$TC_{si}^{(1)}(w_{c,s}, w_{f,s}, \beta_{c,s}, \beta_{f,s}, \tau_{f_{si}}, X_{si}, \nu) = X_{si} \left[\beta_{c,s}^{\nu} w_{c,s}^{1-\nu} + \beta_{f,s}^{\nu} w_{f,s}^{1-\nu} (1 + \tau_{f_{si}})^{1-\nu} \right]^{\frac{1}{1-\nu}} \tag{14}$$

Using the expression for $TC_{si}^{(1)}$, the second-level cost minimization problem writes:

[16] Note that the term in the squared parenthesis can be thought of as the imputed price index of producing X_{si}, which also takes a CES form (Sato, 1967).

$$\min_{X_{si},K_{si},Q_{si}} \quad TC^{(1)}_{si}(w_{c,s}, w_{f,s}, \beta_{c,s}, \beta_{f,s}, \tau_{f_{si}}, X_{si}, \nu) + r_s K_{si} + q_s Q_{si} \tag{15}$$
$$\text{s.t. } A_{si}\left[\alpha_{L,s}X^{\psi}_{si} + \alpha_{K,s}K^{\psi}_{si} + \alpha_{Q,s}Q^{\psi}_{si}\right]^{\frac{1}{\psi}} \geq Y_{si}$$

Solving the above, we arrive at the second-level cost function, denoted by $TC^{(2)}_{si}$. Further, using the expression for $TC^{(2)}_{si}$ in equation (12), we can re-express a given plant's pricing equation more explicitly as follows:[17]

$$P_{si} = \frac{\left(\frac{\sigma}{\sigma-1}\right)}{A_{si}(1-\tau_{Y_{si}})}\Bigg\{(\alpha_{L,s})^{\nu}\left[\beta^{\nu}_{c,s}w^{1-\nu}_{c,s} + \beta^{\nu}_{f,s}w^{1-\nu}_{f,s}(1+\tau_{f_{si}})^{1-\nu}\right]^{\frac{1-\eta}{1-\nu}} + (\alpha_{K,s})^{\eta}r^{1-\eta}_{s}(1+\tau_{K_{si}})^{1-\eta} + (\alpha_{Q,s})^{\eta}q^{1-\eta}_{s}(1+\tau_{Q_{si}})^{1-\eta}\Bigg\}^{\frac{1}{1-\eta}} \tag{16}$$

Instead of working with a two-level CES, if plant-level output was specified to be Cobb-Douglas, we can arrive at a much more tractable pricing equation, written as:

$$P_{si} = \left(\frac{\sigma}{\sigma-1}\right)\frac{(1+\tau_{f_{si}})^{\beta_{f,s}}(1+\tau_{K_{si}})^{\alpha_{K,s}}(1+\tau_{Q_{si}})^{\alpha_{Q,s}}}{A_{si}(1-\tau_{Y_{si}})}\left(\frac{w_{c,s}}{\beta_{c,s}}\right)^{\beta_{c,s}}\left(\frac{w_{f,s}}{\beta_{f,s}}\right)^{\beta_{f,s}}\left(\frac{r_s}{\alpha_{K,s}}\right)^{\alpha_{K,s}}\left(\frac{q_s}{\alpha_{Q,s}}\right)^{\alpha_{Q,s}}$$

It is clear that both equations follow the same structure i.e., price is a function of the markup, input prices, input shares, and idiosyncratic factor distortions. However, equation (16) is explicitly augmented with the elasticity of substitution parameters and provides a more general approach to expressing plant-level prices. The novelty of working with equation (16) rather its Cobb-Douglas counterpart can be appreciated by understanding how the plant's pricing equation affects the measure of misallocation. This is described in the next section.

II.3 Revenue Productivity and Physical Productivity

In order to provide a theoretical context behind the association of distortions and plant-level productivity, it is essential to understand the distinction between total factor revenue productivity (henceforth, TFPR_{si}) and total factor physical productivity (henceforth, TFPQ_{si}). As outlined by Foster et al. (2008), researchers seldom have access to plant-level prices (P_{si}), which would allow us to back out real units of output (Y_{si}) from the data. In this scenario, the measure of productivity has to be based on the revenue of the plant, which is denoted by TFPR_{si}. Exploiting the Hicks-neutral nature of the productivity parameter in equation (3),

[17]The algebra behind the cost minimization procedure is outlined in Appendix B.3.

TFPR_{si} in the model can be written as:

$$\text{TFPR}_{si} = P_{si}A_{si} = \frac{P_{si}Y_{si}}{\left[\alpha_{L,s}(\beta_{c,s}H^{\theta}_{c,si} + \beta_{f,s}H^{\theta}_{f,si})^{\frac{\psi}{\theta}} + \alpha_{K,s}K^{\psi}_{si} + \alpha_{Q,s}Q^{\psi}_{si}\right]^{\frac{1}{\psi}}} \tag{17}$$

In the spirit of Solow (1956), the 'true' technological efficiency of the plant, denoted by TFPQ_{si} can be expressed as:

$$\text{TFPQ}_{si} = A_{si} = \frac{Y_{si}}{\left[\alpha_{L,s}(\beta_{c,s}H^{\theta}_{c,si} + \beta_{f,s}H^{\theta}_{f,si})^{\frac{\psi}{\theta}} + \alpha_{K,s}K^{\psi}_{si} + \alpha_{Q,s}Q^{\psi}_{si}\right]^{\frac{1}{\psi}}} \tag{18}$$

Notably, a high plant TFPR_{si} can at best be interpreted as a noisy signal of high TFPQ_{si}, due to the plant-level price acting as a confounding factor. To see this more clearly, we can rewrite equation (16) to get a more edified expression for TFPR_{si}, as follows:

$$\text{TFPR}_{si} = \frac{\left(\frac{\sigma}{\sigma-1}\right)}{(1-\tau_{Y_{si}})}\Bigg\{(\alpha_{L,s})^{\nu}\left[\beta^{\nu}_{c,s}w^{1-\nu}_{c,s} + \beta^{\nu}_{f,s}w^{1-\nu}_{f,s}(1+\tau_{f_{si}})^{1-\nu}\right]^{\frac{1-\eta}{1-\nu}} + (\alpha_{K,s})^{\eta}r^{1-\eta}_{s}(1+\tau_{K_{si}})^{1-\eta} + (\alpha_{Q,s})^{\eta}q^{1-\eta}_{s}(1+\tau_{Q_{si}})^{1-\eta}\Bigg\}^{\frac{1}{1-\eta}} \tag{19}$$

Clearly, a higher TFPR_{si} is proportional to a higher levels of distortions faced by a plant, within their respective industries. Subsequently, it is important to note that under the assumptions of the model, the only source of variation in TFPR_{si} are the idiosyncratic distortions. Thus, if prices reflect idiosyncratic variations in distortions which are in turn transmitted to the measure of TFPR_{si}, the variance of the revenue based productivity measure can be used to capture the extent of misallocation. Subsequently, if there were no variation in the idiosyncratic distortions[18] i.e., τ_{si} does not vary across plants in a given industry, it is clear from equation (19) that TFPR_{si} also loses its variation. As highlighted by H&K, "*..In the absence of distortions, more capital and labor should be allocated to plants with higher TFPQ to the point where their higher output results in a lower price and the exact same TFPR as at smaller plants..*". On the other hand, high TFPR_{si} serves as an indicator that the given plant is facing distortions that raises the marginal costs of inputs, hence, rendering the plant smaller than optimal.[19] The

[18] Equivalently, one could also assume that each plant in a given industry faces the same industry average wedge (Dias et al., 2016).

[19] Note that since the marginal revenue product of inputs is equalized to the marginal cost, this implies that the given plant can produce more from the inputs.

relation between TFPR_{si} and marginal revenue products can be seen more clearly after solving for the first-order conditions from the profit-maximization problem outlined in equation (9). Let MRP_{si} denote the marginal revenue product of the respective inputs, we can then rewrite the expression for TFPR_{si} as:

$$\text{TFPR}_{si} = \left(\frac{\sigma}{\sigma-1}\right)\Bigg\{(\alpha_{L,s})^{\nu}\left[\beta_{c,s}^{\nu}\text{MRPL}_{c,si}^{1-\nu} + \beta_{f,s}^{\nu}\text{MRPL}_{f,si}^{1-\nu}\right]^{\frac{1-\eta}{1-\nu}} + (\alpha_{K,s})^{\eta}\text{MRPK}_{si}^{1-\eta} + (\alpha_{Q,s})^{\eta}\text{MRPL}_{c,si}^{1-\eta}\Bigg\}^{\frac{1}{1-\eta}} \tag{20}$$

Thus, the dispersion in TFPR_{si} is generated from the underlying dispersion in the marginal revenue products of the inputs, which is in turn rooted in the differences in distortions faced by the plants in a given industry. Further, it is important to note how the elasticity parameters affect our estimates of TFPR_{si}. Since equation (20) is highly non-linear, it is difficult to directly establish this relationship. Since the term in the parenthesis of equation (20) is also a two-level CES, we may attempt to use a Taylor approximation. The empirical literature based on CES functions often works with a translog approximation, originally proposed by Kmenta (1967). Although this approximation has been extended to the case of n inputs by Hoff (2004), it only applies to CES functions without a nesting structure, and thus cannot be applied here. However, the intuition behind the influence of the elasticity parameters can be grasped by a simple example. Assume that a plant faces high distortions in capital input due to the lack of credit access. If the substitutability across combined labor, capital and intermediate inputs (η) is high, the plant can instead rely on the other inputs, essentially negating the effect of the capital distortion. Similarly, a plant hit by labor regulations can start relying on informal labor to meet its labor requirements, when the elasticity of substitution across both labor inputs (ν) is high. Thus, high elasticity across inputs will push down the estimates of TFPR_{si}, such that the plants facing high distortions can possibly have a closer TFPR_{si} to those who benefit from lower distortions. This will reduce the dispersion in TFPR_{si} and hence reduce misallocation in the economy.

II.4 The Efficient Counterfactual

In the absence of distortions, TFPR_{si} will be equalized for all plants i within the given industry $s = 1, 2, .., S$ i.e., resources will be allocated efficiently within each industry. Naturally, the next step is to generate this efficient counterfactual in order to study the potential gains from this intra-industry reallocation exercise. Setting the industry constant equal to one, we can rewrite equation (8) to get the following expressions for real and nominal output.

$$Y_{si} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma} \tag{21}$$

$$P_{si}Y_{si} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma-1} \tag{22}$$

In the counterfactual scenario, let TFPR_s^* denote the efficient level of TFPR in a given industry s, when the plant-level distortions are eliminated such that TFPR_{si} equalizes across all plants within the given industry. Subsequently, the efficient level of real and nominal output at the plant-level writes:

$$Y_{si}^* = \left(\frac{A_{si}}{\text{TFPR}_s^*}\right)^{\sigma} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma}\left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma} = Y_{si}\left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma} \tag{23}$$

$$(P_{si}Y_{si})^* = \left(\frac{A_{si}}{\text{TFPR}_s^*}\right)^{\sigma-1} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma-1}\left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma-1} = P_{si}Y_{si}\left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma-1} \tag{24}$$

When the idiosyncratic distortions are eliminated, we essentially have $\text{TFPR}_{si} = \text{TFPR}_s^*$, using this in the expressions shown above would simply lead to $Y_{si}^* = Y_{si}$ and $(P_{si}Y_{si})^* = P_{si}Y_{si}$, contradicting the assertion that equations (23) and (24) represent efficient outcomes. The reader must note a small subtlety here. The TFPR_{si} term used in the second and third equality of the above equations are meant to represent the actual 'observed' TFPR_{si} of the plants, which is potentially distorted and is not equal to the theoretical construct of the efficient TFPR_s^*. The re-expressed equality in both equations is meant to convey the idea that if the observed TFPR_{si} is higher than TFPR_s^* i.e., if the plant faces barriers to optimal scale, then the efficient output for the given plant would be higher in the counterfactual exercise. Analogous to equation (17), developing the definition of TFPR_s^* gives us:[20]

$$\text{TFPR}_s^* = \frac{(P_sY_s)^*}{\left[\alpha_{L,s}(\beta_{c,s}H_{c,s}^{\theta} + \beta_{f,s}H_{f,s}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s}K_s^{\psi} + \alpha_{Q,s}Q_s^{\psi}\right]^{\frac{1}{\psi}}} \tag{25}$$

Let $\lambda_i \geq 1$ denote the sampling weight assigned to each plant in the dataset. Then, industry level inputs can be written as $H_{c,s} = \sum_{i=1}^{N_s} \lambda_i H_{c,si}$, $H_{f,s} = \sum_{i=1}^{N_s} \lambda_i H_{f,si}$, and similarly for K_s and Q_s. Using the expression for efficient nominal output from equation (24), and rewriting

[20] Note that definition of TFPR_s^* is equivalent to assuming that each plant faces the same industry average wedge, such that there is no within industry variation in TFPR_{si} (Dias et al., 2016). This also ensures that industry demand for each input remains constant in the counterfactual. The derivation behind this expression is shown in Appendix B.5.2.

$(P_sY_s)^* = \sum_{i=1}^{N_s} \lambda_i P_{si} Y_{si}$, we can rewrite equation (25) as :

$$\text{TFPR}_s^* = \left\{ \frac{\displaystyle\sum_{i=1}^{N_s} \lambda_i A_{si}^{\sigma-1}}{\left[\alpha_{L,s}(\beta_c H_{c,s}^{\theta} + \beta_f H_{f,s}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s} K_s^{\psi} + \alpha_{Q,s} Q_s^{\psi}\right]^{\frac{1}{\psi}}} \right\}^{\frac{1}{\sigma}} \tag{26}$$

It is important to re-emphasise that the above expression gives us a model estimate of an efficient level of TFPR in a given industry s, which would occur in the absence of plant-level idiosyncratic distortions. In contrast, H&K define an expression denoted as $\overline{\text{TFPR}}_s$, which is meant to capture the average 'observed' TFPR in a given industry, with possible misallocation of resources. These two terms are completely different, and should not be confused with each other. Once we have the model estimates of the counterfactual outcomes, we can arrive at the potential gross output gains at the industry and aggregate level. Since the counterfactual exercise essentially takes the amount of inputs observed in the model's economy as given and generates a model estimate for potential gains under an efficient allocation, the gross output gains will coincide with aggregate TFP gains (Dias et al., 2016). At the industry level, accounting for sampling weights, efficient industry output (Y_s^*) relative to the observed (under misallocation) output (Y_s) can now be written as:

$$\frac{Y_s^*}{Y_s} = \frac{\left[\displaystyle\sum_{i=1}^{N_s} \lambda_i (Y_{si}^*)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}}{\left[\displaystyle\sum_{i=1}^{N_s} \lambda_i (Y_{si})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}} = \left[\frac{\displaystyle\sum_{i=1}^{N_s} \lambda_i (A_{si})^{\sigma-1}}{\displaystyle\sum_{i=1}^{N_s} \lambda_i \left(A_{si} \frac{\text{TFPR}_s^*}{\text{TFPR}_{si}}\right)^{\sigma-1}} \right]^{\frac{\sigma}{\sigma-1}} \tag{27}$$

At present, equation (27) is difficult to interpret. However, some algebraic manipulation leads to a much more intuitive expression:

$$\frac{Y_s^*}{Y_s} = \left[\frac{1}{\displaystyle\sum_{i=1}^{N_s} \Omega_{si} \left(\frac{1}{\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}}\right)^{\sigma-1}} \right]^{\frac{\sigma}{\sigma-1}} \quad ; \Omega_{si} = \frac{\lambda_i A_{si}^{\sigma-1}}{\sum_{i=1}^{N_s} \lambda_i A_{si}^{\sigma-1}} \tag{28}$$

From equation (28), it should be clear that industry reallocation gains are rooted in the weighted (Ω_{si}) sum of inverse scaled TFPR for all plants within the given industry (Dias et al., 2016, 2018). Note that scaled TFPR ($\text{TFPR}_{si}/\text{TFPR}_s^*$) signifies the deviation of observed plant TFPR_{si} with respect to the industry efficient TFPR^*_s. In particular, scaled TFPR will be higher for plants that face implicit 'taxes' in inputs, pushing them away from the efficient industry TFPR. In contrast, plants with a low scaled TFPR can be thought of as those who benefit from implicit 'subsidies' and face relatively less barriers to growth. To flesh out the intuition behind equation (28), it is informative to rewrite the weights in the following manner:

$$\Omega_{si} = \frac{\lambda_i A_{si}^{\sigma-1}}{\sum_{i=1}^{N_s} \lambda_i A_{si}^{\sigma-1}} = \lambda_i \left[\frac{A_{si}}{\left(\sum_{i=1}^{N_s} \lambda_i A_{si}\right)^{\frac{1}{\sigma-1}}} \right]^{\sigma-1} = \lambda_i \left[\frac{A_{si}}{\text{TFPQ}^*_s} \right]^{\sigma-1} \tag{29}$$

Where TFPQ^*_s denotes the efficient industry TFPQ in the absence of distortions, as shown in Hsieh and Klenow (2009), with the application of sample weights (λ_i). Consequently, plants with a higher TFPQ_{si} ($= A_{si}$) i.e., plants that are more productive, will have a higher Ω_{si}. Now, going back to equation (28), it is clear that higher industry reallocation gains are synonymous to a lower weighted sum of inverse scaled TFPR. This weighted sum will in turn be lower if plants with a higher scaled TFPR are also given a higher weight, while those with a lower scaled TFPR are weighted less. In other words, industry reallocation gains will be higher if on an average, more productive plants (higher Ω_{si}) face higher distortions (higher scaled TFPR). Essentially, this is the spirit behind the idea of resource reallocation, which would hypothetically shift resources to more productive plants who can in turn produce more from the given set of resources, generating higher gross output gains at the industry level.[21] To push the intuition further, it is useful to note that in the efficient counterfactual, a higher Ω_{si} will also correspond to a higher gross-output share for the constrained plants, within their given industry. This can be seen from using equation (24) in the expression for Ω_{si}, as follows:

$$\Omega_{si} = \frac{\lambda_i A_{si}^{\sigma-1}}{\sum_{i=1}^{N_s} \lambda_i A_{si}^{\sigma-1}} = \frac{\lambda_i \left(\frac{A_{si}}{\text{TFPR}^*_s} \right)^{\sigma-1}}{\sum\limits_{i=1}^{N_s} \lambda_i \left(\frac{A_{si}}{\text{TFPR}^*_s} \right)^{\sigma-1}} = \frac{\lambda_i (P_{si} Y_{si})^*}{\sum\limits_{i=1}^{N_s} \lambda_i (P_{si} Y_{si})^*} \tag{30}$$

Finally, applying the Cobb-Douglas aggregator of the representative final good firm from equation (5), the potential aggregate gross output reallocation gains can be expressed as:

$$\frac{Y^*}{Y} = \prod_{s=1}^{S} \left\{ \frac{Y^*_s}{Y_s} \right\}^{\mu_s} = \prod_{s=1}^{S} \left\{ \left[\frac{\sum\limits_{i=1}^{N_s} \lambda_i (A_{si})^{\sigma-1}}{\sum\limits_{i=1}^{N_s} \lambda_i \left(A_{si} \frac{\text{TFPR}^*_s}{\text{TFPR}_{si}} \right)^{\sigma-1}} \right]^{\frac{\sigma}{\sigma-1}} \right\}^{\mu_s} \tag{31}$$

[21] Importantly, this does not necessarily mean that resources will be hypothetically shifted from the smaller plants to the bigger ones. Elimination of distortions is meant to generate an efficient distribution of resources. If smaller plants face higher distortions, this exercise will reallocate resources to such plants as well.

Note that equations (27) and (31) will generate a single value for each industry $s = 1, .., S$ and the economy for a given year, respectively. Since these will be model based estimates, they are essentially 'just-identified' and one cannot check whether the standard errors corresponding to these numbers are high. Moreover, these quantities evidently depend on the computations of three terms: TFPQ_{si} ($= A_{si}$), TFPR_{si} and TFPR_s^*. These three measures are in turn sensitive to the way factors of production at the plant-level are computed from the data. This is because all three of these measures are essentially derived from a Solow residual-like (Solow, 1956) exercise, which is well known to be prone to measurement errors.[22] Deferring this discussion to the Section III.2, equations (27) and (31) form the foundation of the counterfactual exercise, while the plant-level output changes are generated from equations (23) and (24).

II.5 Inferring Distortions

The counterfactual exercise outlined in the previous section is completely independent of the way we define the distortions or estimate them. However, once the extent of misallocation and the potential reallocation gains are documented, learning about the individual distortions and their relation to plant characteristics is useful for the purpose of guiding policy. To this end, it is possible to back out the distortion induced wedges from the data, without imposing any further structure on them. This is done by manipulating the first-order conditions of the plant's profit maximization problem, outlined in Appendix B.7. Subsequently, the wedges can be expressed as follows:[23]

$$(1 + \tau_{f_{si}}) = \left(\frac{\beta_{f,s}}{\beta_{c,s}}\right) \left(\frac{w_{c,s} H_{c,si}^{\frac{1}{\nu}}}{w_{f,s} H_{f,si}^{\frac{1}{\eta}}}\right) \tag{32}$$

$$(1 + \tau_{K_{si}}) = \left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}}\right)\left(\frac{w_{c,s} H_{c,si}^{\frac{1}{\nu}}}{r_s K_{si}^{\frac{1}{\eta}}}\right)(X_{si})^{\frac{\nu-1}{\nu} - \frac{\eta-1}{\eta}} \tag{33}$$

$$(1 + \tau_{Q_{si}}) = \left(\frac{\alpha_{Q,s}}{\alpha_{L,s}\beta_{c,s}}\right)\left(\frac{w_{c,s}}{q_s}\right)\left(\frac{H_{c,si}^{\frac{1}{\nu}}}{Q_{si}^{\frac{1}{\eta}}}\right)(X_{si})^{\frac{\nu-1}{\nu} - \frac{\eta-1}{\eta}} \tag{34}$$

$$(1 - \tau_{Y_{si}}) = \left(\frac{\sigma}{\sigma - 1}\right) \frac{w_{c,s} H_{c,si}^{\frac{1}{\nu}}}{(\alpha_{L,s}\beta_{c,s}) P_{si} Y_{si}} \frac{\left[\alpha_L(\beta_c L_{csi}^{\theta} + \beta_f L_{fsi}^{\theta})^{\frac{\psi}{\theta}} + \alpha_K K_{si}^{\psi}\right]}{(X_{si})^{\frac{\nu-1}{\nu} - \frac{\eta-1}{\eta}}} \tag{35}$$

[22]The seminal work of Jorgenson and Griliches (1967) introduced constant quality indexes for labour and capital input, which resulted in capital and labor input accounting for majority of the US economic growth in 1945-65, rather than TFP based on the Solow residual.

[23]Note that these wedges are defined relative to the absolute informal labor input wedge, as described previously in footnote 14.

As noted before, these wedges can possibly represent several frictions. What these expression actually represent is thus difficult to grasp without putting further structure. However, when we assume a unitary elasticity of substitution across all inputs ($\nu = \eta = 1$), the way these expressions capture distortions can be understood. In particular, we would infer a high formal labor distortion ($\tau_{f,si}$) from the data, when the ratio of informal labor to formal labor compensation is higher than what we can expect from their respective factor shares. Since formal employees usually earn more than their informal counterparts,[24] this ratio will be high when the informal hiring of the plant is high. This would in turn indicate that that the given plant faces strong regulatory costs of formal labor, and thus shifts to more informal hiring. Thus, capturing the mechanism of sidestepping labor regulations. A similar line of reasoning holds for equations (33) and (34) i.e., we infer a capital and intermediate input distortion from the data when the compensation of informal employees is higher than the cost of capital and intermediate inputs, relative to what we can expect from the factor shares. This is not exactly intuitive, rather, it is simply a result of using the absolute informal labor distortion as the base, when computing the wedges. Since the focus of this dissertation is mainly on the implications of the formal labor distortion, measuring wedges relative to informal labor is a necessary step, albeit with the loss of intuitive reasoning for the other distortions.[25]

A fundamental issue with 'identifying' distortions in this manner is that the observed revenue factor shares in the data could be distorted themselves. In other words, when the economy faces factor price distortions, the data generating process of the observed revenue factor shares might be a function of the distortions. If this is indeed the case, it will not be possible to disentangle the wedges calculated in the data, using equations (32) to (35), from the respective revenue factor shares. Section III.2 outlines a way in which we can deal with this issue.

II.6 Gauging the Relative Importance of Distortions

As noted before, the dispersion in TFPR_{si} can be used to gauge the extent of misallocation in the observed economy. Even after taking a logarithmic transformation of equation (16), it is evidently difficult to capture how the variation of the formal labor, capital and intermediate input distortions contribute to aggregate misallocation. This exercise is much simpler when we assume a unitary elasticity of substitution, which allows a tractable expression for the variance of TFPR_{si}. However, in the interest of maintaining generality in the elasticity of substitution, I follow an alternative path to decomposing the relevance of each distortion towards aggregate misallocation. In particular, I investigate the hypothetical aggregate reallocation gains by

[24]Formal employees are subject to various in-kind benefits and also have job security, which informal employees usually do not. This naturally implies a higher compensation, even if wages might be comparable.

[25]Notably, the choice of the base input is often arbitrary in the misallocation literature, since we only care about the distortion backed out from the data. This choice is much more explicit in my framework due to the focus on labor regulations.

eliminating one wedge at a time. The intuition behind this exercise is that if a given distortion in an input is particularly problematic for aggregate misallocation, the output gains from efficiently allocating the said input will subsequently be higher as well.

As shown by Dias et al. (2016), this can be done by 'eliminating' variation in a particular wedge, while keeping the other wedges operational. For example, suppose that we are interested in investigating at the reallocation gains implied by only removing idiosyncratic variation in the formal labor input wedge within the given industries, while keeping the other inputs fixed. This is equivalent to imposing an industry-specific formal input wedge $(1+\widetilde{\tau}_{f_s})$, while ensuring that the industry demand of formal labor is equal to the one observed in the data (equal to $H_{f,s}$).[26] In essence, we are shutting down the variation in $(1+\tau_{f,si})$ such that the formal labor input distortion no longer varies across plants in a given industry, and evaluating the new counterfactual output. Since size-based labor regulations vary across plants of different sizes, this exercise is equivalent to removing such regulations. Let the new allocation of formal labor input be denoted by $\widetilde{H}_{f,si}$, then we can simply rewrite equation (32) as:

$$\widetilde{H}_{f,si} = \left[\left(\frac{\beta_f}{\beta_c}\right)\left(\frac{w_{c,s}H_{c,si}^{1-\theta}}{w_{f,s}(1+\widetilde{\tau}_{f,s})}\right)\right]^{\frac{1}{1-\theta}} \tag{36}$$

Next, in the new allocation, in order to make sure that the aggregate demand for formal labor input remains the same at the industry-level, we impose the following:

$$H_{f,s} = \sum_{i=1}^{N_s}\lambda_i\widetilde{H}_{f,si} = \sum_{i=1}^{N_s}\lambda_i\left[\left(\frac{\beta_f}{\beta_c}\right)\left(\frac{w_{c,s}H_{c,si}^{1-\theta}}{w_{f,s}(1+\widetilde{\tau}_{f,s})}\right)\right]^{\frac{1}{1-\theta}} \tag{37}$$

With some manipulation, we arrive at the following expression for the industry-specfic wedge:

$$(1+\widetilde{\tau}_{f,s}) = \left(\frac{\beta_{f,s}}{\beta_{c,s}}\right)\left(\frac{w_{c,s}}{w_{f,s}}\right)\left(\frac{\sum_{i=1}^{N_s}\lambda_i H_{c,si}^{1-\theta}}{H_{f,s}}\right)^{1-\theta} \quad ;\forall i \in s = 1,2,..,S \tag{38}$$

Plugging this industry-specfic wedge back into equation (36) gives us the final expression for the new allocation of formal labor input.

$$\widetilde{H}_{f,si} = \frac{H_{f,s}}{\frac{\sum_{i=1}^{N_s} H_{c,si}^{1-\theta}}{H_{c,si}}} \tag{39}$$

Which further implies the new plant-level output, as follows:

[26] Matching the industry demand to what is observed in the data is a simple way to make sure that the counterfactual exercise reallocates the chosen factor across plants within the given industry, such that the total demand for the said factor remains the same at the industry-level.

$$\widetilde{Y_{si}}^f = A_{si}\left[\alpha_{L,s}(\beta_{c,s}H^{\theta}_{c,si} + \beta_{f,s}\widetilde{H}^{\theta}_{f,si})^{\frac{\psi}{\theta}} + \alpha_{K,s}K^{\psi}_{si} + \alpha_{Q,s}Q^{\psi}_{si}\right]^{\frac{1}{\psi}} \tag{40}$$

Finally, to arrive at the reallocation gains implied by this new allocation, we can simply replace Y^*_{si} with $\widetilde{Y_{si}}^f$ in equation (27), and subsequently rewrite equation (31) as:

$$\frac{\widetilde{Y}^f}{Y} = \prod_{s=1}^{S}\left\{\frac{\widetilde{Y}_s^f}{Y_s}\right\}^{\mu_s} = \prod_{s=1}^{S}\left\{\left[\frac{\sum_{i=1}^{N_s}\lambda_i(\widetilde{Y_{si}}^f)^{\frac{\sigma-1}{\sigma}}}{\sum_{i=1}^{N_s}\lambda_i(Y_{si})^{\frac{\sigma-1}{\sigma}}}\right]^{\frac{\sigma}{\sigma-1}}\right\}^{\mu_s} \tag{41}$$

Similarly, we can arrive at reallocation gains implied by removing the capital and intermediate input wedges, one at a time. The algebra for these two is a bit more involved, and is left for Appendix B.8. Unfortunately, we cannot arrive at the reallocation gains implied by removing the output wedge, due to the lack of closed form solutions in the procedure described above. However, note that since the output wedge is built on the informal labor distortion, which is arguably the least distorted input out of all, most of the reallocation gains should be captured by efficient allocation of formal labor, capital and intermediate input, one at a time.

III Data and Variables

III.1 ASI, Worker Distinction and Labor Laws

Data for the formal (registered) manufacturing units is drawn from the Annual Survey of Industries (ASI), compiled annually[27] by the Central Statistical Organization (CSO) of India and made available for analysis by the Ministry of Statistics and Program Implementation (MOSPI) of the Government of India. The basic unit of enumeration is at plant/factory/enterprise level, identified with a Dispatch Serial Number. The coverage extends nationwide except for the states of Mizoram and Sikkim, and the union territory of Lakshadweep. The dataset covers all manufacturing plants/factories registered under the Sections 2(m)(i) and 2(m)(ii) of the Factories Act, 1948.[28] Additionally, the survey also covers bidi and cigar manufacturing plants registered under the Bidi and Cigar Workers Act, 1966. ASI is a census of all registered manufacturing units with 100 or more and a random sample of those employing less than 100 workers. Sampling weights are applied to all relevant computations to infer statistics about

[27] The cross-section covers one accounting year. With the enactment of Income Tax Act, by and large, the accounting year of all factories is from 1st April of the preceding year to 31st March of the current year.

[28] All manufacturing plants with 10 or more employees (using power) and with 20 or more employees (not using power) are required to be registered under the said Act.

the population of registered manufacturing plants. I utilise three cross-sections of the ASI, including – 2010-11, 2014-15 and 2018-19. Accounting for missing data for relevant variables, the dataset provides detailed information on production and inputs for around 41,000 plants for each given year. The uniqueness of this dataset is that it allows us to distinguish between formal and informal workers in terms of the number of employees and their wages. Since this distinction is crucial, I emphasise on how the data allows me to separate these workers, and some important regulations that influence the use of each. ASI provides data on workers employed 'directly' by the plant (formal labor) and those employed through 'contractors' (informal labor). This implies that those who are employed directly include both full-time and temporary workers.[29]

Under Chapter VB of the Industrial Disputes Act of 1947 (IDA, henceforth), manufacturing plants that employ more than 100 workmen must acquire permission from a government authority before laying off even one workman, inducing high adjustment costs.[30] This can be understood as the opportunity cost in terms of time and effort exerted by employers to go through the long process of adjusting formal labor employment. Moreover, since these regulations are enforced by government officials through arbitrary inspections, there is substantial room for extracting bribes. Amirapu and Gechter (2020) documents a positive association between these regulatory costs and the exposure to corruption. Debroy and Bhandari (2008) highlight that bribes paid by employers to such inspectors typically rises with number of employees. The definition of a 'workman' under this Act includes only those employed directly by the plant. Therefore, those employed through contracting agencies (contract workers, henceforth) do not come under the purview of this Act. Thus, the main friction generated by the IDA comes from the high implicit cost of firing formal workers with an additional cost of dealing with government officials. These costs are not associated with contract workers, since they are not considered direct employees of the plants in which they work. Such workers are thus exempted from severance pay or retrenchment authorization, allowing employers to exploit this flexibility and sidestep the various opportunity and direct costs associated with the IDA.

However, it is important to note that the use of contract workers is not completely unregulated. The Contract Labour (Regulation and Abolition) Act of 1970 (CLA, henceforth) provides conditions for the services of contract workers in plants employing at least 20 such workers. The total contract worker usage must be declared by the employers, in addition to the nature of their work. Several in-kind benefits are also stipulated by this Act, including – minimum wage, health/safety provisions and pensions. As a legal antidote to the contract worker loophole in the IDA, the ease of substitution across formal and informal workers is

[29]Bertrand et al. (2021) denotes the formal and informal distinction as full-time and contract workers, which is misleading.

[30]The government authorities are usually labour courts with the objective to judge industrial disputes between employers and employees. If the court favors the employer, the worker is laid off after being paid a severance pay for 15 days for each year of service. If there is no agreement, the case can be moved to the higher courts.

limited by Section 10 of the CLA. In particular, contract workers cannot be in charge of tasks performed by permanent workers of the establishment, or establishments in the same industry. Essentially, the government can prohibit the use of contract labor at a firm that uses it for its perennial operations, by giving an abolition notification. However, there was a legal 'gray area' as to what would actually happen to contract workers at establishments not following this rule. An important 2001 ruling by the Supreme Court of India lifted this uncertainty. Known as the SAIL judgement, the Supreme Court ruled that there is no requirement of automatic absorption of contract workers in the permanent force of the establishment, following an abolition notification under Section 10 of the CLA. Bertrand et al. (2021) exploit this as an event study and report the subsequent rise in contract worker usage in larger plants. In summary, flexibility comes from the little to no adjustment costs of contract/informal labor.

III.2 Variables and Calibrations

Since the model operates on some strong assumptions regarding constant input shares and input prices for all plants within a given industry, a higher level of disaggregation in the industry definition should be used to maintain relevancy of the model's implications. To this end, this study uses the 4-digit classification for the manufacturing industries, from 1010 to 3320 within Section-C of the National Industry Classication (2008). This 4-digit classification allows me to retain meaningful plant-level variation within each industry whilst also maintaining the idea of disaggregation. As noted in before in Section II.4, the model's implications can be sensitive to the way we measure output and the various inputs. For the baseline computations and for the purposes of inferring distortions, a gross output approach is used. Gross output for a given plant is calculated as a sum of the nominal value of total sales, receipts from both manufacturing and non-manufacturing services, and various rents received. For the purpose of comparison and to generate more meaningful statistics for aggregate reallocation gains, a value-added approach is also used. Value-added for a given plant is measured as gross output short of intermediate input consumption for the given accounting year.

For the baseline computations, the labor inputs are measured by their respective wage bills i.e., $H_{z,si} = w_{z,si}L_{z,si}$ for $z \in \{f, c\}$, where $w_{z,si}$ represents plant-specific average wages per worker and $L_{z,si}$ denotes employment. Correspondingly, the sector specific wage rate is set to one. Computing labor inputs as their wage bill is rudimentary way of controlling for differences in human capital (Hsieh and Klenow, 2009; Dias et al., 2016) i.e., I assume that differences in wages capture differences in hours worked and possible skill differences. As a robustness check, labor input is also measured by the average number of persons employed for both formal and informal labor. A similar approach is used to compute the intermediate inputs used in production, i.e., Q_{si} is set equal to the corresponding value of intermediate inputs used in production. These include energy, material and services consumed by the plant in the given accounting year, with the addition of imported consumption and other expenses incurred on raw materials. The measurement of capital at the firm/plant level remains a persistent issue, especially when the measurement of TFP is concerned.[31] Hulten (1991) recognises that using the book value of capital remains a flawed technique but is often the only recourse without data on past investments. Since I work with cross-sectional data, the perpetual inventory method cannot be used. Following the general resource misallocation literature (Gustavo and Cristobal, 2012; Chen and Irarrazabal, 2013; Dias et al., 2016, 2018), and Hsieh and Klenow (2009) in particular, capital is measured as the average of the net book value of fixed capital at the beginning and end of the accounting year. The rental rate is assumed to be 10 percent of the average net book value of capital.

The elasticity of substitution across the differentiated goods produced by plants is assumed to be equal to 3 i.e., $\sigma = 3$, in the baseline computations. Since the value of σ readily affects the model estimates of reallocation gains, I provide a robustness check by increasing this elasticity. Following Padmakumar (2022), the elasticity of substitution across labor inputs is set to $\nu = 2.7$, and the elasticity of of substitution across combined labor, capital and intermediate inputs is set to $\eta = 4$.[32] Note that these elasticities are usually assumed to be one in the misallocation literature, due to the assumption of a Cobb-Douglas technology. Deviating from this assumption and using a general CES specification, I am able to re-estimate the model using various values of ν and η as robustness checks.

As highlighted in Section II.5, if the economy faces distortions, it is difficult to identify them separately from the industry shares of the corresponding factors. A rather crude way of dealing with this issue is to set the industry factor shares equal to the ones observed in a relatively less distorted economy (Hsieh and Klenow, 2009; Gustavo and Cristobal, 2012; Dias et al., 2016, 2018). Following the literature, the industry shares from the U.S. are used the computations. Data for the respective industry factor shares comes from the NBER Productivity Database. Since the 4-digit industry classification in the U.S. (SIC-1987) does not exactly match the one in India (NIC-2008), an approximate concordence table is built to match the industries. Importantly, to maintain plant-level variation, this study does not drop the handful of Indian manufacturing industries for which no close match was found in the U.S. 4-digit classification. Instead, the average industry shares from the matched Indian industries that are closely related to the activities of the unmatched ones, are used as a proxy. I end up with a total of 125 4-digit industries with revenue factor shares calibrated to the U.S., for each of the three years.

[31] Hicks (1981) (p.204) remarks that, *"The measurement of capital is one of the nastiest jobs that economists have set to statisticians."*

[32] Note that the model allows me to use the conditional factor demand equations (Appendix B.3.2) to estimate these elasticities. However, this requires the use of plant-level fixed effects due to the idiosyncratic distortions, and also industry fixed effects due to the assumed industry constants. Due to the lack of access to the panel version of ASI, I use the recent work of Padmakumar (2022), who indeed works with the panel version and a similar two-level specification.

IV Empirical Results

IV.1 TFP Distributions

As highlighted in Section II.3, the model allows us to get just-identified estimates of two types of TFP distributions, namely, total factor physical productivity ($\text{TFPQ}_{si} = A_{si}$) and total factor revenue productivity ($\text{TFPR}_{si} = P_{si}A_{si}$). Using the former, we can evaluate the density of less and/or more productive plants, relative to the industry efficient TFPQ^*_s. This is done by creating a measure of (log) scaled TFPQ, denoted by $ln(A_{si}W_s^{\frac{1}{\sigma-1}}/\text{TFPQ}^*_s)$, where W_s denotes the sum of sample weights for all plants within the given industry. The few lines of algebra behind this measure is shown in Appendix B.9. If this measure if positive, then plants are more productive than TFPQ^*_s, which prevails under efficient allocation.[33] Correspondingly, if scaled TFPQ is negative, plants are less productive than the industry efficient TFPQ^*_s. Overall, the distribution of scaled TFPQ informs us about the heterogeneity in plant-level productivity.

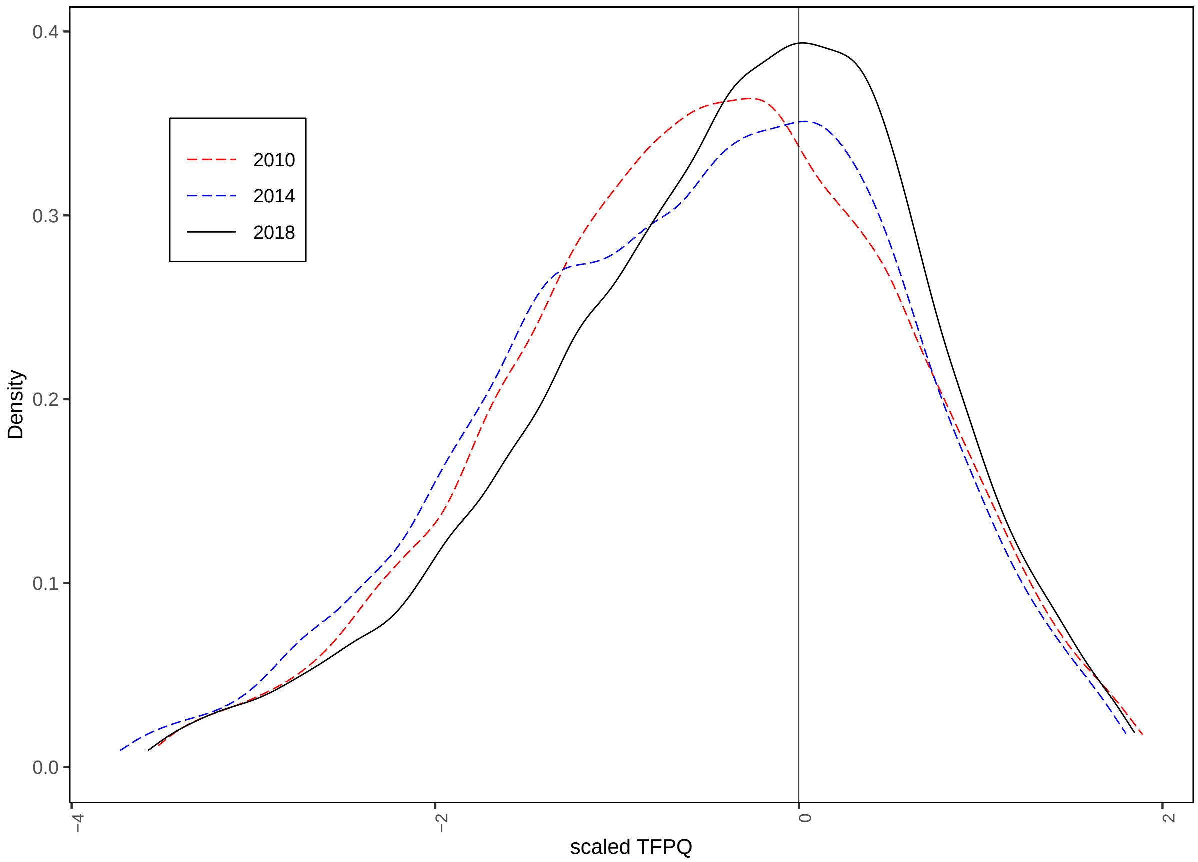
As evident from Figure 2, the left tail of the scaled TFPQ distribution becomes thicker in 2014-15, relative to 2010-11. Further, the left tail becomes substantially thinner in 2018-19, compared to both 2014-15 and 2010-11. This suggests that the policy and competitive environment started favoring the existence of inefficient plants in 2014-15, relative to 2010-11. Subsequently, the environment appears to be favoring the existence of more productive plants in 2018-19. The summary statistics for the scaled TFPQ distribution are reported in Table 1. Consistent with the increase in standard deviation for 2014-15, the difference between the 75th-25th and the 90th-10th percentiles increased as well. This implies that the 'physical' productivity gap between the plants at the top and bottom percentiles increased within their respective industries. As visually apparent from Figure 2, most of the increase in this gap can be attributed to the addition/increase of inefficient plants rather than an addition/increase of more productive units.[34]

This is also evinced by the statistics in the lower panel of Table 1. In particular, owing to the the large addition/increase in inefficient plants, the difference between the 25th-10th percentile increased in 2014-15. While a small addition/increase in productive plants lead to the the 90th-75th percentiles moving closer in 2014-15. In contrast, for 2018-19, the large decline/decrease in inefficiency around the 25th percentile lead to the larger difference between the 25th-10th percentiles. Correspondingly, an addition/increase in efficiency around the 75th percentile lead to a smaller difference between the 90th-75th percentiles. Overall, we can conclude that the registered segment of manufacturing is comprised of more physically productive plants in 2018-19, relative to 2014-15 and 2010-11.

[33] Note that if $x < 1$, then $ln(x) < 0$ and if $x > 1$ then $ln(x) > 0$. This is the basic idea behind this measure.

[34] Note that the cross-sectional ASI data does not allows me to identify plant entry and exit. This means that the thicker left tail can be due to two reasons – reduction in physical productivity of surviving plants or an addition of inefficient plants.

Figure 2: Density of Scaled TFPQ, Gross Output Approach



Note: The top and bottom 1% tails of the $ln(A_{si}W_s^{\frac{1}{\sigma-1}}/\text{TFPQ}_s^*)$ are removed to deal with outliers.

Table 1: Scaled TFPQ Distribution, Summary Statistics

	2010-11	2014-15	2018-19
S.D.	1.05	1.09	1.03
75-25	1.48	1.58	1.43
90-10	2.76	2.84	2.70
25-10	0.70	0.72	0.77
90-75	0.58	0.53	0.50
Obsv.	40,016	40,645	42,238

Note: The reported statistics are for the distribution of $ln(A_{si}W_s^{\frac{1}{\sigma-1}}/\text{TFPQ}_s^*)$ after removing the top and bottom 1% tails. The first row reports the standard deviation. The second and third rows report the difference between the 75th-25th and 90th-10th percentile of the distribution. The lower panel reports the difference between the 25th-10th and 90th-75th percentiles.

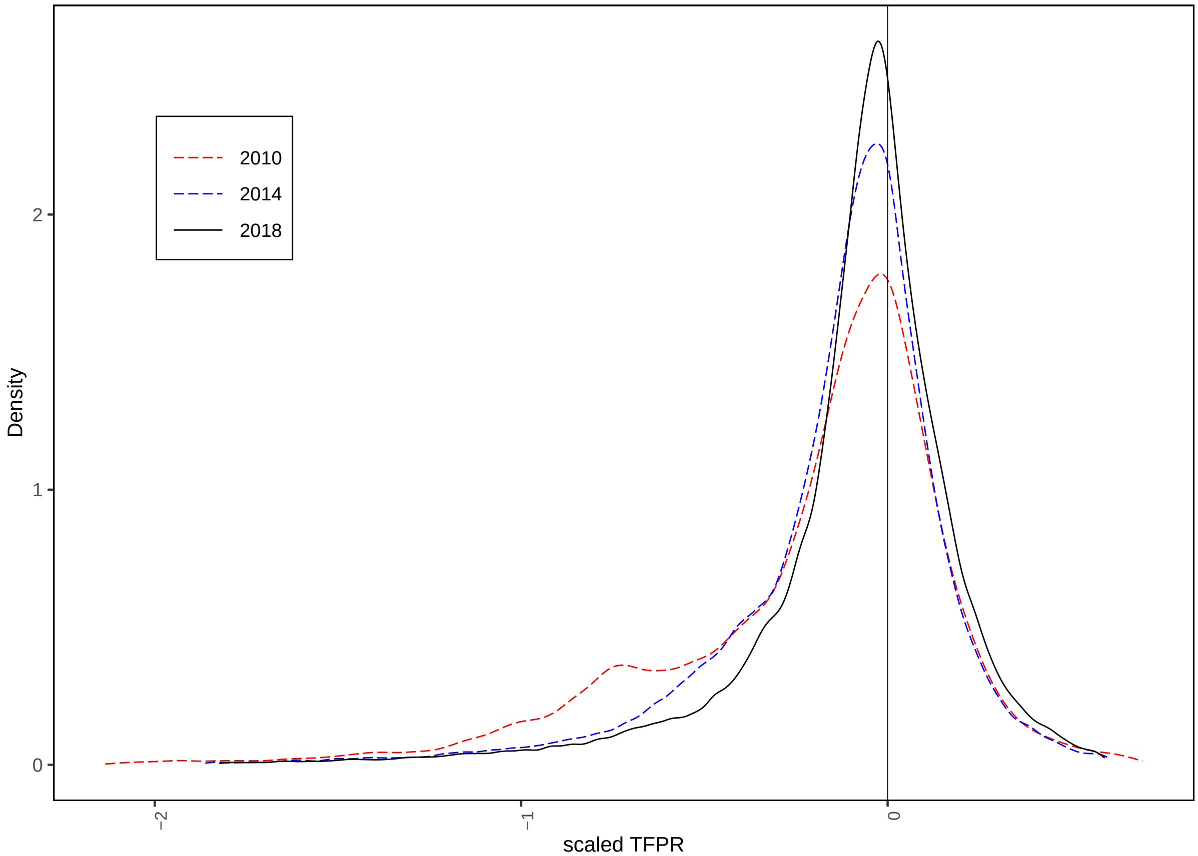
Figure 3 illustrates the the distribution of (log) scaled TFPR i.e., $ln(\text{TFPR}_{si}/\text{TFPR}^*_s)$ for each year. This measure allows us to look at the deviation of observed plant TFPR_{si} from the industry efficient TFPR^*_s. If this deviation is positive, the given plant can be understood as suffering from high distortions within their respective industries, and vice-versa. As described before in Section II.3, the dispersion in TFPR_{si} can be used to gauge the extent of resource misallocation in the economy. In particular, if the distortions are eliminated, there would be no dispersion in TFPR_{si} within each industry, and marginal revenue product of factors would be equalized across plants within a given industry. On the other hand, a higher dispersion in TFPR_{si} would indicate more variation in the idiosyncratic distortions faced by the plants. This would in turn imply higher potential reallocation gains from shifting resources to plants with relatively higher marginal revenue product of factors, within each industry.

The intuition behind looking at this dispersion can be understood with an example. Suppose that there are only two plants $i \in s$, and they differ slightly in their respective measures of TFPR_{si}. Assume that this difference is enough for one to have a negative scaled TFPR (benefit from implicit 'subsidies') and the other to have a positive scaled TFPR (face implicit 'taxes' in employing inputs). Since these plants will be very close to the industry efficient TFPR, reallocating resources from the plant benefiting from implicit subsidies to the one facing barriers will only lead to small output gains. In contrast, assume that TFPR_{si} of each are far apart i.e., one plant benefits from large implicit subsidies while the other faces strong barriers. Then, gains from reallocation will be higher or equivalently misallocation will be high.

Table 2 reports the summary statistics behind this distribution for each year. From Figure 3 and Table 2, it is apparent that the dispersion in log scaled TFPR has declined overtime i.e., on an aggregate, resource misallocation appears to have decreased in the 2010-11 to 2018-19 period. Importantly, we observe a substantial decrease in dispersion for 2014-15, relative to 2010-11. This was followed by only a minor decrease in 2018-19. The middle and extremes of the distribution moved considerably closer in 2014-15, leading to the considerable decline in standard deviation. However, for 2018-19, although the extremes of the distribution moved closer, there appears to be an increase in the dispersion at the middle. This limits the overall decline in dispersion for 2018-19. We can decipher these observations further by focusing on the relative differences in the lower panel of Table 2.

As visually apparent from Figure 3, it should be noted that the decline in dispersion observed for 2014-15 was mostly rooted in the decline of implicit 'subsidies' for the plants with extremely low TFPR_{si} (ref. 25th-10th percentile difference), rather than the decrease in implicit 'taxes' for the plants with high TFPR_{si} (ref. 90th-75th percentile difference). Interestingly, for 2018-19, relative to 2014-15, the difference between the 90th-75th percentile shows that there is a slight increase in barriers faced by the plants already suffering from high distortions. However, since the left tail cut off is the same for 2014-15 and 2018-19, the higher difference between the 25th-10th percentiles in 2018-19 shows that there is a substantial decrease in

Figure 3: Density of Scaled TFPR, Gross Output Approach



Note: The top and bottom 1% tails of the $ln(\text{TFPR}_{si}/\text{TFPR}^*_s)$ are removed to deal with outliers.

Table 2: Scaled TFPR Distribution, Summary Statistics

	2010-11	2014-15	2018-19
S.D.	0.40	0.30	0.28
75-25	0.41	0.27	0.58
90-10	0.92	0.66	0.23
25-10	0.38	0.26	0.53
90-75	0.12	0.11	0.12
Obsv.	40,645	40,016	42,238

Note: The reported statistics are for the distribution of $ln(\text{TFPR}_{si}/\text{TFPR}^*_s)$ after removing the top and bottom 1% tails. The first row reports the standard deviation. The second and third rows report the difference between the 75th-25th and 90th-10th percentile of the distribution. The lower panel reports the difference between the 25th-10th and 90th-75th percentiles.

implicit subsidies for the plants at the 25th percentile of the distribution. Overall, it can be concluded that the decrease in misallocation observed from 2010-11 to 2018-19 was majorly driven by the decline in implicit benefits enjoyed by the relatively unconstrained plants. On the other hand, there was a limited to no decline in implicit barriers faced by plants who were already suffering from high factor price distortions in 2010-11. Thus, although aggregate misallocation has decreased, policies do not appear to have weakened the implicit barriers faced by plants in the factor markets.

IV.2 Hypothetical Reallocation Gains

IV.2.1 Aggregate Gains

Having documented the decline in misallocation over the 2010-11 to 2018-19 period, it is useful to investigate to what extent an efficient allocation could have improved aggregate output. This is done by computing equations (27) and (31) for each given year. The computations gives us gross output gains for 125 industries and aggregate gross output gains for a given year, respectively. Due to the issue of double counting, the concept of value-added is a better reflection of final output when we aggregate across industries. To evaluate the value-added gains, equation (31) is also computed using a value-added production function at the plant level.[35] Correspondingly, for the value-added exercise, industry revenue factor shares are re-calibrated using the value added measure in the NBER Productivity Database, rather than the value of shipments. Table 3 reports the estimates of aggregate reallocation gains using both approaches, for each year.

Table 3: Aggregate Reallocation Gains, Gross Output and Value-Added

Measure	2010-11	2014-15	2018-19
Gross Output	85.29	73.44	63.78
Value Added	212.15	193.73	166.19

Note: The reported statistics are for $(\frac{Y^*}{Y} - 1) * 100$, where $\frac{Y^*}{Y}$ is computed from equation (34) using both gross output and value added measures of plant-level output. Plants with negative value-added are dropped in the computation.

The immediate observation that stands out in Table 3 is that for each year, reallocation gains are substantially higher when plant-level output is computed using value-added, compared to gross output. The observed difference can be understood by a small digression on the relationship between TFP measurement using the two approaches. Following Cobbold (2003),

[35] This simply amounts to using a value-added measure (gross output minus intermediate consumption) for each plant, and ignoring intermediate goods as inputs in the production function. All of the derivations qualitatively remain the same.

we know that the relationship between value-added and gross output measures of TFPR can be formally written as:

$$\text{TFPR}_{\text{VA},si} = \left(\frac{\text{GO}_{si}}{\text{VA}_{si}}\right)\text{TFPR}_{\text{GO},si} \tag{42}$$

This relation simply says that the value-added measure of productivity will exceed the gross output measure by a factor equal to the ratio of gross output to value-added, for the given plant. Notably, if the plant uses more intermediate inputs, the gross output to value-added ratio will be higher as well. Thus, in the presence of intermediate inputs, the value-added measure of productivity will always be inflated in comparison to gross output. Consequently, the value-added method will lead to higher plant TFPR_{si}, with the inflation factor varying according to the intermediate input use of each plant. If there is indeed substantial variation in intermediate input usage across plants, the dispersion of scaled TFPR based on the value-added approach will consequently be higher. Figure 11 in Appendix A compares the density of scaled TFPR using both approaches. The relative difference in dispersion is evident. Hence, reallocation gains based on value-added will be higher than the computation based on gross-output.[36]

Evidently, using both methods, we observe that reallocation gains decline over the 2010-11 to 2018-19 period. This is consistent with the declining dispersion in scaled TFPR i.e., the decline in misallocation, reported previously in Figure 3 and Table 2. Another important component of this decrease in misallocation should be noted. As described before in equations (27) to (30), we know that reallocation gains will be higher if on an average, more productive plants (higher TFPQ_{si}) face stronger distortions (higher TFPR_{si}). Notably, even if the dispersion in scaled TFPR is high, but most of the bite from distortions is not felt by the physically productive plants (high TFPQ_{si}), reallocation gains can still be low. To assess whether on an average, the physically productive plants are facing stronger implicit barriers, we can look at the correlation between scaled TFPR and scaled TFPQ as a broad indicator (Dias et al., 2016, 2018). The corresponding estimates for each year are reported in Table 4.

Table 4: Correlation between scaled TFPR and scaled TFPQ

	2010-11	2014-15	2018-19
ρ(scTFPR, scTFPQ)	0.44	0.36	0.37
Obsv.	40,645	40,016	42,238

Note: The reported statistics are for the correlation between $ln(\text{TFPR}_{si}/\text{TFPR}^*_s)$ and $ln(A_{si}W_s^{\frac{1}{\sigma-1}}/\text{TFPQ}^*_s)$, after removing the top and bottom 1% tails of each distribution. All correlation estimates are significant at the 1 percent level. The gross output approach is used in the computations of the relevant measures.

[36]This result is consistent with the findings of Chatterjee (2011) and Kabiraj (2020), who use the same dataset.

The correlation estimate decreases in 2014-15 relative to 2010-11, corresponding to the relatively lower gross output gains in 2014-15, as reported in Table 3. Interestingly, although reallocation gains decline in 2018-19 compared to 2014-15, the correlation estimate between scaled TFPR and scaled TFPQ saw a minor increase. Note that the right tail of the scaled TFPQ distribution became thicker in 2018-19 (ref. Figure 2). This means that the increase in the correlation estimate observed for 2018-19 could very well be due to the addition of more productive plants, rather than an increase in the barriers faced by the existing productive plants. This cannot be explored with the cross-sectional ASI dataset due to the lack of common unit identification. However, the decline in reallocation gains for 2018-19 suggest that the smaller dispersion in scaled TFPR more than compensated for this phenomenon. In other words, even though the highly productive plants started facing more barriers to growth, the smaller dispersion in scaled TFPR (and hence the distortions) was enough to pull down the implied gains from reallocation. Overall, we can conclude that although aggregate resource misallocation has declined overtime, the gains from an efficient allocation remain high in an absolute sense, and hence deserve policy attention. In particular, for 2018-19, under an efficient allocation of resources within industries, aggregate formal manufacturing output could have been higher by 166 percent in terms of value-added and 63 percent in terms of gross output.[37]

IV.2.2 Disaggregated Industry Gains

The previous section documented high potential reallocation gains for formal manufacturing. In the interest of guiding policy, we can disaggregate these gains to the industry-level. This is done by computing equation (27), returning reallocation gains for each of the 125 4-digit industries, for each year. This exercise gains merit if there is substantial variation in reallocation gains across different industries. Figure 12 in Appendix A illustrates this variation by ranking the observed gains for the industries, for each year.

Notably, even though a given industry may exhibit high reallocation gains, this will only translate to higher aggregate output if the given industry's gross output share is also high. To this end, Table 5 reports the top ten industries in terms of the highest gross output share and their corresponding reallocation gains, for each year. The third column traces the correlation between scaled TFPR and scaled TFPQ as an indicator of whether more physically productive plants faced more distortions within each industry, for each year. The fourth column reports the within industry dispersion in scaled TFPR, as a measure of the observed misallocation in each industry, for each year. The observed (increase) decrease in reallocation gains are a construct of either a (higher) lower correlation between scaled TFPR and scaled TFPQ and/or

[37]Instead of estimating the counterfactual reallocation gains at the aggregate level, we can do the same exercise at the state level. If the idiosyncratic distortions faced by the plants are clustered at the state-level, perhaps due to the variation in the policy environment for each, this exercise gains merit. The results from this exercise are shown in Table 10 in Appendix A, followed by a brief discussion.

Table 5: Industry Reallocation Gains, Ranked by Gross Output Shares

S.No	4-Digit Industry (NIC, 2008)	Gross Output Reallocation Gains	Gross Output Share	corr(scTFPR,scTFPQ)	sd(scTFPR)
		2010-11			
1.	Manufacture of refined petroleum products	136.74	19.00	0.25	0.27
2.	Manufacture of basic iron and steel	130.87	8.26	0.35	0.34
3.	Manufacture of motor vehicles	40.54	4.90	0.32	0.26
4.	Manufacture of parts and accessories for motor vehicles	27.41	3.85	0.30	0.24
5.	Preparation and spinning of textile fibres	74.03	3.33	0.31	0.25
6.	Manufacture of pharmaceuticals, medicinal chemical and botanical products	94.22	3.19	0.46	0.47
7.	Manufacture of vegetable and animal oils and fats	46.67	2.93	0.44	0.26
8.	Manufacture of cement, lime and plaster	23.89	2.46	0.42	0.36
9.	Manufacture of grain mill products	91.84	2.28	0.55	0.37
10.	Manufacture of plastics products	68.32	2.06	0.42	0.26
		2014-15			
1.	Manufacture of refined petroleum products	124.15	18.66	0.24	0.23
2.	Manufacture of basic iron and steel	99.45	8.01	0.25	0.25
3.	Manufacture of motor vehicles	66.24	4.82	0.36	0.32
4.	Manufacture of parts and accessories for motor vehicles	21.47	3.66	0.30	0.22
5.	Manufacture of pharmaceuticals, medicinal chemical and botanical products	125.30	3.37	0.50	0.46
6.	Preparation and spinning of textile fibres	32.22	2.91	0.13	0.15
7.	Manufacture of plastics products	48.43	2.84	0.42	0.20
8.	Manufacture of cement, lime and plaster	35.42	2.56	0.47	0.34
9.	Manufacture of grain mill products	71.42	2.50	0.39	0.24
10.	Manufacture of jewellery and related articles	68.05	2.40	0.48	0.21
		2018-19			
1.	Manufacture of refined petroleum products	159.62	10.21	0.27	0.18
2.	Manufacture of basic iron and steel	118.97	8.68	0.26	0.21
3.	Manufacture of motor vehicles	7.87	5.54	0.14	0.17
4.	Manufacture of pharmaceuticals, medicinal chemical and botanical products	110.93	4.37	0.43	0.44
5.	Manufacture of parts and accessories for motor vehicles	26.27	4.16	0.32	0.21
6.	Manufacture of plastics products	43.31	3.16	0.40	0.21
7.	Manufacture of vegetable and animal oils and fats	24.30	2.87	0.23	0.15
8.	Manufacture of cement, lime and plaster	34.40	2.78	0.37	0.36
9.	Manufacture of grain mill products	66.09	2.53	0.34	0.21
10.	Preparation and spinning of textile fibres	14.41	2.47	0.10	0.15

Note: The reported statistics for a given industry s are from $(\frac{Y_s^*}{Y_s} - 1) * 100$, where $\frac{Y_s^*}{Y_s}$ is computed using equation (27). Gross output share corresponds to $\mu_s = \frac{P_s Y_s}{PY}$ in equation (6), where $P_s Y_s = \sum_{i=1}^{N_s} \lambda_i P_{si} Y_{si}$. All correlation estimates are siginificant at the 1% level.

a (higher) lower dispersion in scaled TFPR.

As we would expect, the observed changes in the aggregate reallocation gains (ref. first row in Table 3) are built on heterogeneous changes in reallocation gains at the industry level. In 2014-15, relative to 2010-11, the industries that saw a decline in reallocation gains include – Preparation and spinning of textile fibres (drop by 41.81 ppts.), Manufacture of iron and steel (31.42 ppts.), Manufacture of grain mill pdts. (20.42 ppts.), Manufacture of plastic pdts. (19.89 ppts.), Manufacture of refined petroleum pdts. (12.59 ppts.) and Manufacture of parts and accessories for motor vehicles (5.94 ppts.). For 2018-19, relative to 2014-15, there is a noticeably large decline in reallocation gains for the Manufacture of motor vehicles (drop by 58.37 ppts.), this is followed by – Preparation and spinning of textile fibres (17.81 ppts.), Manufacture of pharmaceuticals (14.37 ppts.), Manufacture of grain mill pdts. (5.33 ppts.), Manufacture of plastic pdts. (5.12 ppts.) and Manufacture of cement, lime and plaster (1.02 ppts.). Thus, several important industries appear to have improved their allocative efficiency overtime.

On an average over the three years, the refined petroleum industry reflects the highest reallocation gains of around 140 percent, combined with the highest average gross output share of 16 percent. This is followed by the iron and steel and the pharmaceutical industry with average reallocation gains of 116 percent and 110 percent, commanding average gross output shares of 8 percent and 4 percent, respectively. Given the importance of these industries towards aggregate formal manufacturing output, this is particularly problematic. Importantly, although the refined petroleum industry and the iron and steel industry observed a decline in misallocation in 2014-15, both industries appear to be suffering from increased misallocation in 2018-19.[38] The pharmaceutical industry saw an increase in misallocation during 2014-15, with only a small drop in 2018-19. On the more optimistic side, three industries observed a consistent decline in misallocation, including – textiles, grain mill products and the plastic products industry.

Table 11 in Appendix A reports the top ten industries with the highest reallocation gains for each year. Majority of these industries hold a gross output share of around 0.10-1.46 percent. Thus, even though these industries suffer from high misallocation, the high reallocation gains will only lead to limited increase in aggregate output. There appears to be no systematic pattern i.e., majority of the industries that suffered from high misallocation in 2010-11, do not re-appear in the top ten in 2014-15, and similarly for 2018-19.[39] Overall, although aggregate misallocation has declined, the output of several important industries evidently remains constrained. Thus, from the lens of policy making, the decline in aggregate misallocation should not taken at its face value.

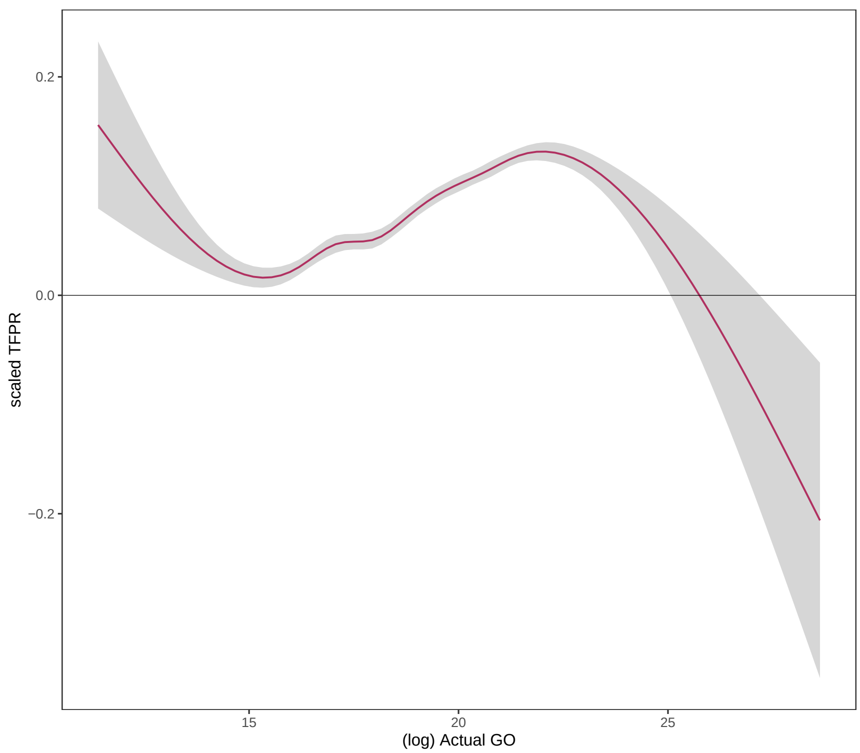
[38] The refined petroleum industry saw a 8.45 ppts. decline in its output share in 2018-19, but still remains as the largest contributor towards aggregate output.

[39] However, there are a couple of exceptions. The manufacturing of bottled drinks and the manufacturing of plastic and synthetic rubber suffered from high misallocation over the entire 2010-11 to 2018-19 period.

IV.3 Efficient Plant Size Distribution

Following the industry-level analysis in the previous section, it is important to document how the output of the underlying plants behaves under an efficient allocation. In particular, in the counterfactual exercise, we can characterise which plants would increase or reduce their size. As previously outlined in Section II.4, the counterfactual gross output for a given plant will be higher if its observed TFPR_{si} is high, relative to the efficient TFPR^*_s. In other words, under a within-industry efficient allocation of resources, the most constrained plants will increase their gross output, while those who benefit from implicit subsidies will shrink in size. In order to characterise whether on an average, smaller or larger plants face more distortions within a given industry, we can explore the association between scaled TFPR and gross output. Figure 4 illustrates this relationship for 2018-19.[40]

Figure 4: Relationship between Scaled TFPR and Plant Size, 2018-19



Note: The plot shows point estimates and 95% confidence intervals from non-parametric regression of (log) scaled TFPR on the log of observed gross output; positive scaled TFPR $\iff$ plants face implicit 'taxes' and negative scaled TFPR $\iff$ plants benefit from implicit 'subsidies'.

There are four main takeaways. First, on an average, the smallest plants seem to face the highest distortions in input markets. Second, as we reach the small to medium sized plants, the observed relationship somewhat weakens but still remains positive and significant.[41] Third, the

[40]The observed relationship remains qualitatively the same for 2010-11 and 2014-15.

[41]It can be argued that the smallest plants face strong capital constraints due to lack of collateral or good

mid size plants face strong distortions, with an increasing intensity as size increases. Finally, the mid to large and the substantially larger plants either do not face distortions or start benefiting from implicit subsidies as size keeps increasing.[42] Given these four observations, on an average, we can expect the intra-industry efficient allocation of resources to increase the output of the small and mid-sized plants, while the mid to large and the largest plants should relatively shrink.

Notably, given the idiosyncratic nature of the distortions (ref. equation (9)), these results will only hold on an average. In other words, some small or mid-sized plants could very well be benefiting from implicit subsidies due to a reduced scale of operation, while some larger plants might be suffering from distortions. This further implies that a certain percentage of small and mid-sized plants will indeed shrink under an efficient allocation of resources, while a percentage of large-sized plants will grow even further. We can explore this heterogeneity and analyse the relationship in Figure 4 more formally. This is done by evaluating the efficient plant size distribution against the observed data, using the following steps. First, the plants are put into five equally sized quintile groups based on their observed gross output. Second, the efficient counterfactual gross output of each plant is computed, such that TFPR equalizes within their respective industries. Third, I compute the ratio of efficient to observed gross output for each plant and sort them into four bins. Finally, I arrive at the percentage of plants belonging to each quintile group, within each specific bin based on the efficient to actual gross output ratio. These bins are specified as follows – 0-50% (plants that should reduce their size by half or more), 50-100% (plants that should reduce in size by less than half), 100-200% (plants that should increase size by less than double), and 200% + (plants that should at least double in size).

Table 6 reports the findings. Indeed, as previously hinted by Figure 4, the most populated bin for the plants belonging to the bottom, second and third quintile group is 200%+. Subsequently, the majority of the plants in the fourth and top quintile fall in the 50-100% bin. This implies that majority of small, small to mid-sized and the mid-sized plants will be able at least double their gross output if resources were to be allocated efficiently, within their respective industries. On the other hand, majority of the plants at the top of the distribution would have to reduce their gross output by less than half of their observed production. As expected, there is considerable heterogeneity in this result.[43] Many plants at the bottom and middle of the distribution will indeed shrink in size, while some at the top will also grow. In particular, for 2018-19, around 30 percent, 35 percent and 46 percent of the plants belonging

credit history, this issue might subsequently reduce as we move to the larger plants.

[42]This relationship can be further broken down by looking at the relationship between plant size and the individual distortions, one at a time. This discussion is delayed to the next section.

[43]Note that for 2010-11, in contrast of 2014-15 and 2018-19, we observe that even the plants at 4th quintile group of the distribution would on an average exhibit growth under an efficient allocation (37 percent of the plants in this group would observe an output increase of 200 percent or more). This is consistent with the findings in Table 4 i.e., larger plants faced more constraints in 2010-11.

Table 6: Observed Plant Size v/s Plant Size Absent Distortions, 2014-15 to 2018-19

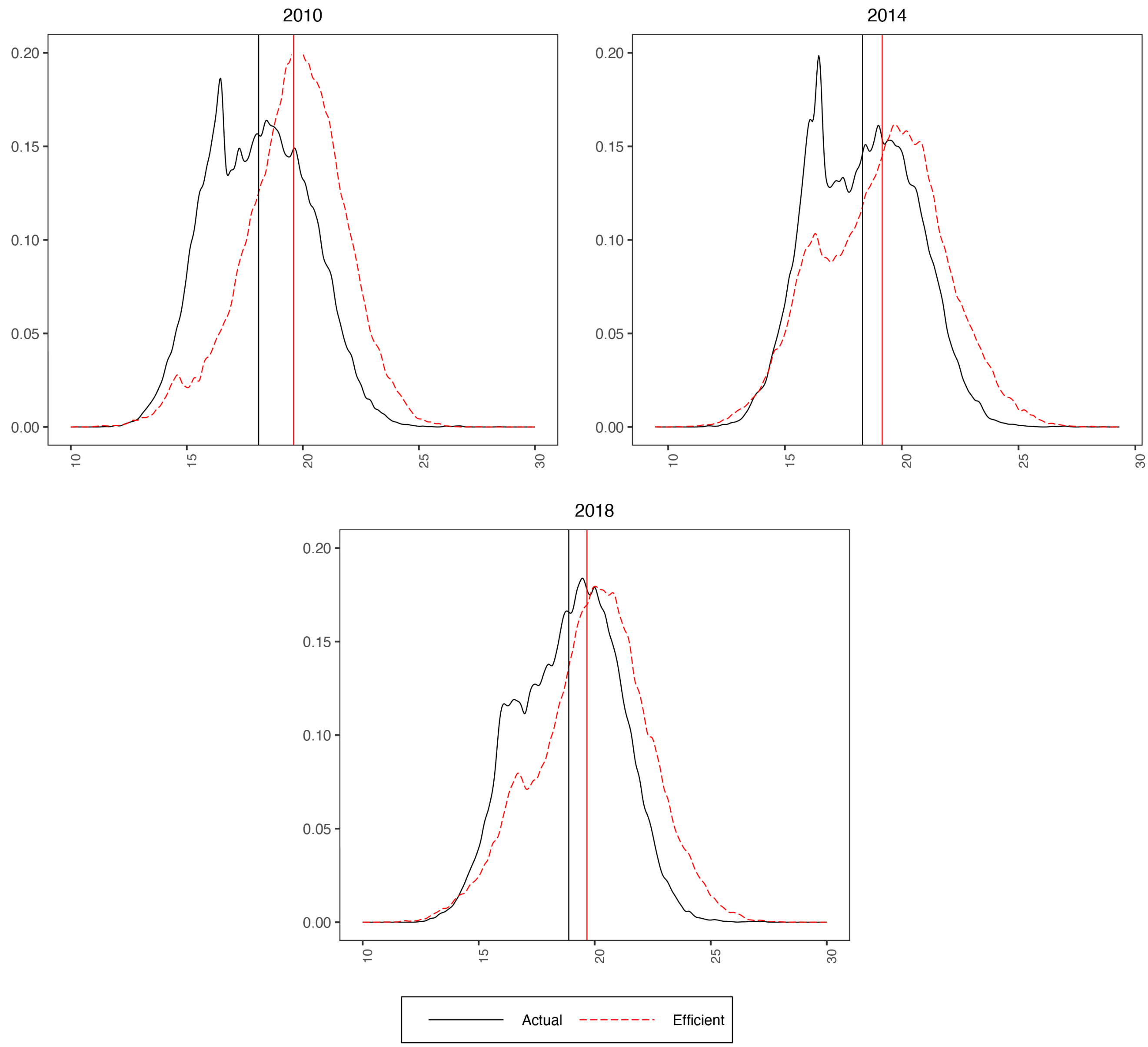
	Bins of Efficient to Actual Output Ratio			
2010-11	0-50	50-100	100-200	200+
Top Quintile	17.06	39.97	29.74	13.19
4th Quintile	16.70	28.58	17.64	37.05
3rd Quintile	11.50	13.16	10.09	65.22
2nd Quintile	7.19	6.34	6.80	79.65
Bottom Quintile	9.10	4.07	5.12	81.69
2014-15	0-50	50-100	100-200	200+
Top Quintile	13.93	44.29	34.74	7.01
4th Quintile	12.18	41.63	28.32	17.83
3rd Quintile	11.85	33.22	19.98	34.93
2nd Quintile	11.62	31.30	15.89	41.18
Bottom Quintile	14.96	26.58	15.90	42.54
2018-19	0-50	50-100	100-200	200+
Top Quintile	9.61	45.88	37.93	6.56
4th Quintile	9.65	44.21	31.38	14.74
3rd Quintile	9.02	37.00	24.05	29.91
2nd Quintile	8.37	27.12	16.58	47.91
Bottom Quintile	9.84	20.59	19.56	50.00

Note: The reported statistics are for the percentage of plants belonging to each bin based on the efficient to observed size ratio, for each quantile group based on the observed size. The efficient gross output is computed using equation (24), after removing the top and bottom 1% tails of the scaled TFPR and scaled TFPQ.

to the bottom, second and third quantile groups, respectively, will shrink in size. Meanwhile, around 44 percent and 46 percent of plants in the fourth and top quintile groups, respectively, would increase their size. Overall, for each year, many plants appear to shrink under an efficient allocation of resources. However, in order for the counterfactual exercise to generate gross output reallocation gains, the average gross output under the efficient distribution should be higher than the one observed in the data. Reassuringly, Figure 5 shows that this is indeed the case i.e., the mean efficient gross output is higher than the corresponding mean observed gross output, for each year.

The efficient distribution is noticeably shifted to the right as compared to actual distribution. This intuition behind this shift deserves a brief discussion. In particular, this shift provides further evidence for the statistics reported in Table 5. We know that the small, small to mid-sized and mid-sized plants exhibit a substantial increase in their output. This pushes

Figure 5: Actual Gross Output v/s Gross Output Absent Distortions



Note: The plots show the kernel density of the logarithm of gross output. The vertical lines show the mean values for each distribution. The 'efficient' gross output i.e., $(P_{si}Y_{si})^{*}$ is computed using equation (24) with the application of sample weights.

the left-tail of the distribution to the right, such that the density of these plants in the efficient distribution is lower compared to the actual. This right push is further augmented by the fact that there are indeed much more small and medium sized plants in the formal manufacturing sector (ref. Table 9 in Appendix A). Further, as previously reported in Table 5, even though the majority of large plants observe a decrease in their output, most of the reduction amounts to less than half of their observed output. On the other hand, most of the relatively smaller plants would at least double their output. This large right push from the relatively smaller plants combined with a relatively lesser left push from the plants at the top of the distribution materializes into higher density near the right tail of the efficient distribution. Overall, this pattern appears to be consistent across all years. Intuitively, the average gap between the actual and efficient gross output should be reflective of reallocation gains documented in Table 3. In particular, this gap should increase in 2014-15, relative to 2010-11, followed by a decline in 2018-19. Comparing the distance between the vertical lines in Figure 5, the average gap between the distributions clearly follows this pattern overtime.

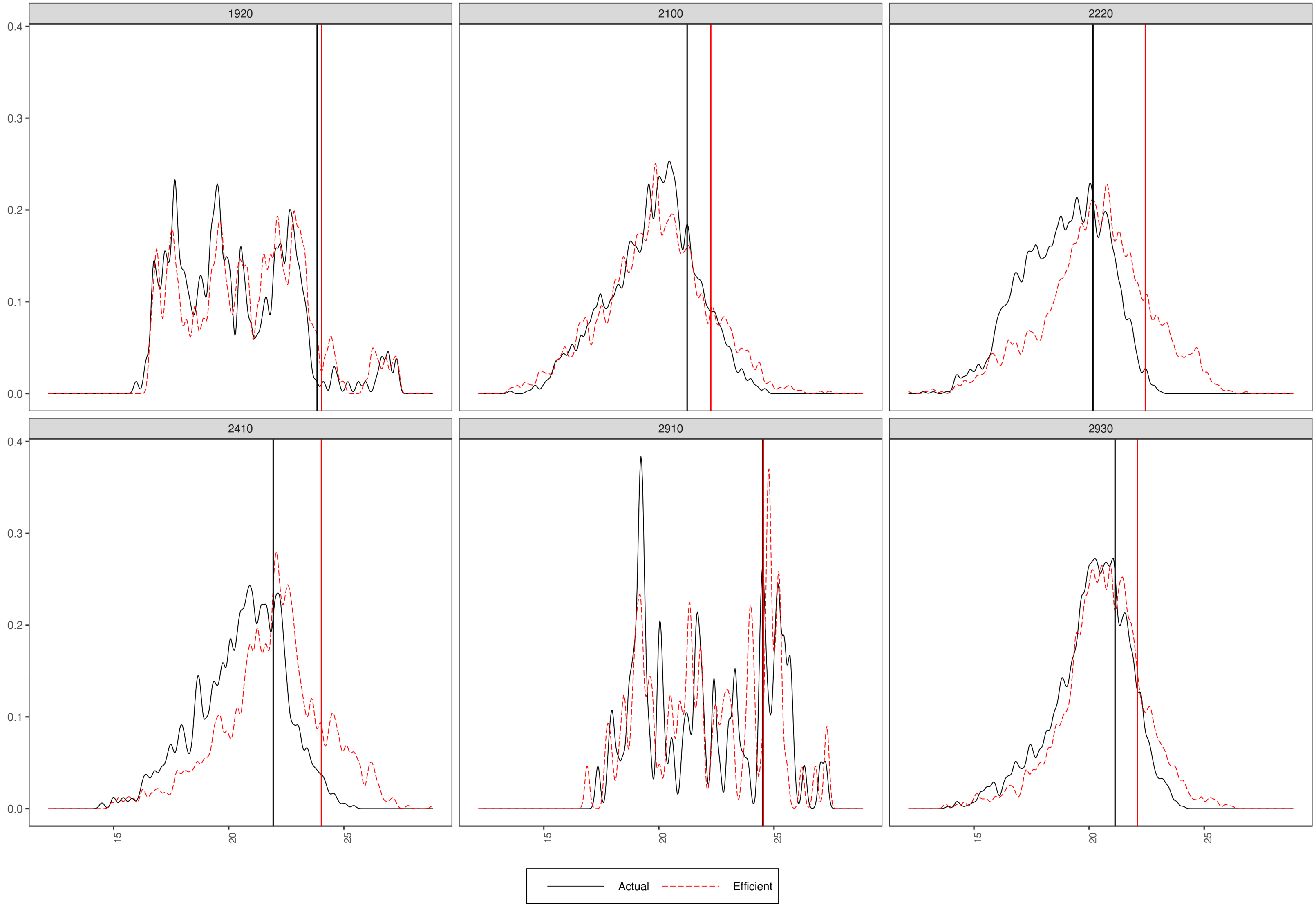
Figure 5 only illustrates the behaviour of the two distributions at the aggregate. Owing to the difference in gains from reallocation at the industry level (ref. Table 5), there is reason to expect heterogeneity in these distributions for the underlying industries. Figure 6 illustrates these two distributions for the top six industries in terms of gross output share in 2018-19. The right shift of the efficient distribution is apparent, with the mean gap reflecting the reallocation statistics reported in the bottom panel of Table 5.[44]

Overall, we observe that under an efficient allocation of resources, majority of the small, small to mid-sized and mid-sized plants would increase their size. On the other hand, majority of the large-sized plants would shrink. In other words, we can conclude that majority of the plants near the bottom of the distribution face high distortions that does not allow them to achieve their optimal scale. Further, majority of the plants at the top of the distribution appear to be benefiting from implicit 'subsidies' in factor markets that allows them to increase their scale, above and beyond their optimal size under an efficient allocation.[45] This is in contrast to the results of Hsieh and Klenow (2009), who do a similar exercise for India's formal manufacturing in 1994-95. The authors' report that majority of the mid-sized plants would shrink in size, while the larger plants are likely to expand. This is attributed to the inefficiency of mid-sized establishments belonging to the public sector. Since majority of the structural reforms of 1991 in terms of industrial de-licensing were implemented in the late 1990's, Hsieh and Klenow (2009) evidently could not capture the effect of these reforms. In the more recent context, given the increasing push towards privatisation, this argument clearly does not apply. Instead, more recently, I argue that majority of the large sized plants who could potentially be benefiting from implicit 'subsidies' in factor prices of intermediate inputs and capital, while also benefiting from high informal labor usage that allows them to circumvent formal labor

[44]Note that the mean gap between the distributions should not be compared across industries. This is because the concavity of the logarithmic transformation will reflect a smaller mean gap for industries which have a higher average observed gross output. For example, for 2018-19, Table 5 reports that the Petroleum industry (1920) could have potentially increased output by 159%, while the potential gains for the Iron and Steel industry (2410) are around 118%, but in Figure 6, the mean gap for the former will appear to be relatively smaller due to its higher average observed gross output.

[45]Note that this line of reasoning comes from equation (24), which highlights that plants facing implicit 'taxes' in factor prices will grow in the counterfactual. Conversely, those benefiting from implicit 'subsidies' will shrink in size.

Figure 6: Actual Gross Output v/s Within Industry Equalized TFPR Gross Output, Top Six Industries by GO Share, 2018-19



Note: The plots show the kernel density of the logarithm of gross output. The vertical lines show the mean values for each distribution. The 'efficient' gross output i.e., $(P_{si}Y_{si})^*$ is computed using equation (24) with the application of sample weights. The panels report the following 4-digit NIC (2008) codes – 2220 = Manufacture of plastic products, 1920 = Manufacture of refined petroleum products, 2100 = Manufacture of pharmaceuticals, medicinal chemical and botanical products, 2410 = Manufacture of basic iron and steel, 2910 = Manufacture of motor vehicles, 2930 = Manufacture of parts and accessories for motor vehicles.

distortions. On the other hand, although relatively smaller plants do not have to suffer from distortions in formal labor, they could potentially be constrained by distortions in intermediate inputs and capital. The next section is devoted to making this argument more concrete.

V Sources of Misallocation

V.1 Correlation Evidence

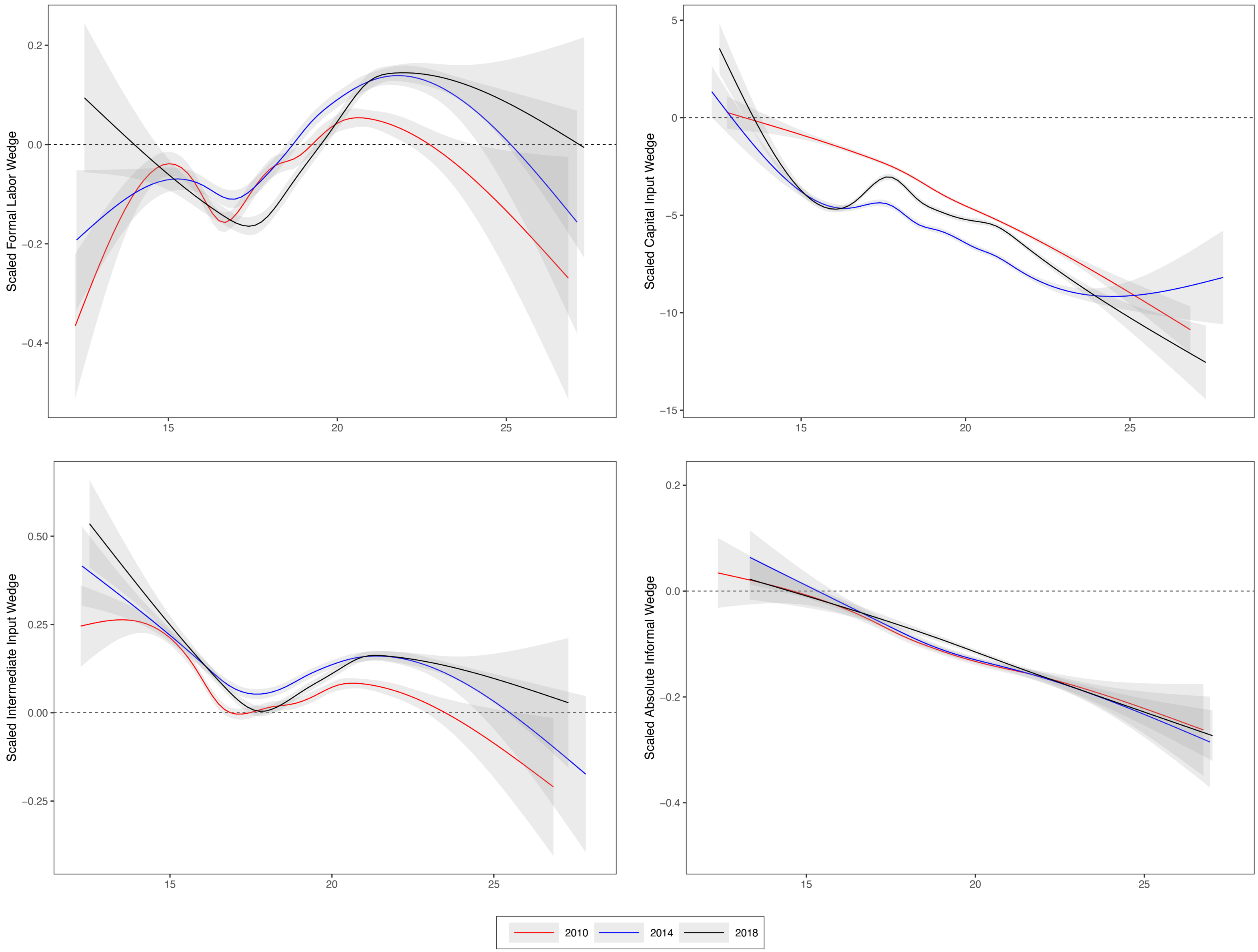
The indirect approach to understanding resource misallocation is essentially a *black box* when it comes to understanding the sources of misallocation. The general formulation of the wedges without imposing any particular structure leaves the sources of misallocation open to many interpretations. However, given the context of size-based labor regulations and the differences in various perceived obstacles as size changes (Sen, 2020), a natural way to understand these distortions is to investigate their relationship against plant size. Moreover, given the appraisal of observed plant size against the efficient counterfactual, in the previous section, exploring the size-based relationship also allows us to understand which distortions prevent the small and mid-sized plants from growing. To this end, the correlation between scaled TFPR and plant size previously shown in Figure 4 can be disentangled by investigating the correlations between the various (scaled) wedges and plant size (Dias et al., 2016, 2018).

Figure 7 shows the relationship between the scaled wedges[46] and plant size measured as gross output, for each year. Consistent with the context of size-based labor regulations, it is observed that the smaller plants either do not face any formal labor distortion (an insignificant relationship) or benefit from implicit subsidies in hiring formal workers. As plant size increases, we observe that the medium and medium to large-sized plants start to face implicit taxes when hiring formal workers. This relationship either becomes insignificant or turns into implicit subsidies as we move to the largest plants. Importantly, for the medium and medium to large sized plants, the obstacles in hiring formal workers seems to increased in 2014-15, followed by an insignificant increase in 2018-19. Overall, based on pure correlation analyses, we observe that medium and larger plants have started to face more constraints in formal hiring, while the smaller plants either do not face constraints or benefit from implicit subsidies in hiring formal workers.[47] Note that the framework does not characterize the formal labor wedge as size dependent adjustment costs implied by labor regulations, however, given with the way the correlation appears to behave against plant size, the link with size-based labor regulations seems inevitable.

[46]The wedges are estimated using equations (32)-(35) for each plant, and then divided by their respective industry mean. This scaling is done to ensure coherence with the scaled TFPR measure. A positive scaled wedge reflects that a plant faces more constraints compared to the industry average. On the other hand, a negative scaled wedge means that the plant benefits from implicit subsidies, relative to the industry mean.

[47]Notably, the implicit subsidies for the smaller plants appears to somewhat increase in 2018-19 as well.

Figure 7: Relationship between Scaled Factor Price Wedges and Plant Size (log of Actual GO)



Note: The plot illustrates the point estimates of a non-parametric fit using smoothed conditional means with a 95% confidence interval, between logarithm of gross output (sample weighted) and the scaled wedges for each year. Positive scaled wedge = implicit ‘taxes’ in that input, negative scaled wedge = implicit ‘subsidies’ in that input.

Based on suggestive evidence from the knitted garment industry, Banerjee and Duflo (2005) emphasizes the role of differences in marginal products of capital in explaining India's low manufacturing TFP. Indeed, in the current setup, differences in interest rates for plants of different sizes can lead to substantial variation in the marginal product of capital. In particular, if the smaller plants face higher rates than their larger counterparts, we can expect the relationship between the scaled capital wedge to be decreasing in plant size, with the capital distortion stemming from imperfect financial markets. Figure 7 shows that the the smallest plants seem to face implicit barriers in renting capital, while the larger plants tend to benefit from reduced rates. This pattern is more evident for 2014-15 and 2018-19, and less so for 2010-11. Overall, the barriers to renting capital for small plants appear to have slightly decreased in 2014-15, followed by a rise in 2018-19. On the other hand, the relatively larger plants (including mid-sized plants) seem to benefit from implicit subsidies in renting capital. However, it is important to note that this conclusion is highly sensitive to the way capital is measured from the data i.e., this conclusion will only hold if our measure of the average net book value of fixed assets indeed captures plant-level capital well.

Recent work by Boehm and Oberfield (2020) uses plant-level data from ASI to investigate the distortions in intermediate inputs. In their setup, hold up problems and weak enforcement of contracts (characterised by poor courts) on intermediate inputs lead to a decline in state-level productivity. Sen (2020) reports that small sized plants face issues in obtaining licences while mid-sized plants report issues in access to energy, transport and telecommunications as the strongest obstacles to growth. In this context, it is important to ask how the intermediate input wedge behaves with plant size. Figure 7 illustrates that the implicit barriers in employing intermediate inputs are faced by majority of the plants, except those at the very top of the plant-size distribution. Notably, most of this bite is felt by the smallest plants. Moreover, from 2010-11 to 2018-19, the level of the intermediate input wedge appears to have shifted significantly upwards for the medium and medium to large sized plants, implying that these plants are increasingly facing stronger distortions to intermediate input use.

The last panel in Figure 7 shows the relationship between the absolute informal labor wedge (inverse of the output wedge) and plant size. There are two important observations. First, majority of the plants benefit from implicit subsidies in hiring informal labor, as we would expect. Second, the larger plants benefit from relatively higher implicit subsidies associated with informal labor, compared to the smaller plants. Thus, the largest plants who are 'constrained' by size-based formal labor regulations seem to be benefiting more from the reduced cost of hiring informal labor. This could potentially reflect the idea that larger plants might have better access to contracting agencies and a stronger bargaining power when dealing with them, something smaller plants might lack. Reassuringly, given the fact that there has been no new size-based labor regulations for informal workers, the observed relationship appears to have remained stable overtime.

Overall, we can conclude that although the small and mid-sized plants benefit from implicit subsidies in hiring formal workers, they face strong constraints in other inputs. In particular, the small sized plants are constrained by strong distortions in capital and intermediate input. On the other hand, the mid-sized plants seem to be increasingly facing strong constraints in intermediate inputs. The mid to large and large-sized plants are constrained by the high regulatory costs of formal workers, induced by size-based regulations. However, at the same time, these plants seem to benefit from a lower costs of informal hiring. Given the legal 'gray area' in the use of contract workers (ref. Section III.1), we can expect larger plants to easily substitute formal workers with more informal hiring.[48] Thus, labor informality might reduce the relevance of formal labor distortions for larger plants, while they benefit from relatively lower costs in other inputs.

V.2 Relative Importance of Distortions

V.2.1 Output Gains from Isolated Reallocation

Following the documentation of the potential sources of misallocation, it is important to investigate the contribution of each individual distortion towards aggregate misallocation. As noted before in Section II.6, it is difficult to break down the variance of TFPR_{si} into its individual components due to the assumed functional form. Rather than examining the dispersion of each individual distortion, we can instead compute the reallocation gains generated by eliminating one distortion at a time, while keeping the others operational. This amounts to imposing an average wedge for all plants within a given industry, and finding the efficient allocation of the given factor. Subsequently, we can compute the new output under the efficient allocation of only the given factor and re-purpose equation (31) to compute the reallocation gains.[49]

Table 7 reports the resulting aggregate gross output gains, by eliminating one wedge at a time. On an average, the formal labor input wedge and the intermediate input wedge seem to be driving majority of the misallocation over the years. The reallocation gains implied by removing variations in the intermediate input wedge appear to increase in 2014-15, before falling back to the level observed in 2010-11. While the capital input wedge appears to have become more problematic in 2018-19, after a decline in its contribution to aggregate misallocation in 2014-15, compared to 2010-11. From the previous section, we know that the formal labor wedge traces the impact of size-based labor regulations. Interestingly, we observe that gains from an efficient allocation of formal labor decrease in 2014-15 and 2018-19. Before discussing this observation further, it is instructive to first explore how the static plant size distribution behaves when size-based labor regulations are removed.

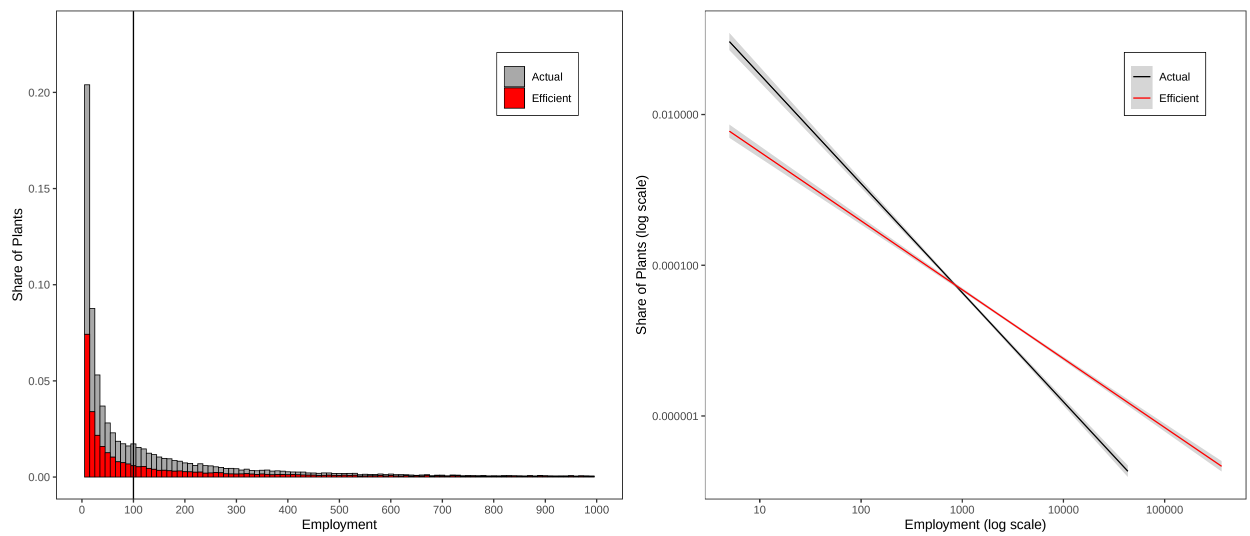
[48] As estimated by Padmakumar (2022), the estimated elasticity of substitution between both labor inputs

Table 7: Aggregate Gross Output Gains, Isolated Factor Reallocation

Eliminated Wedge	2010-11	2014-15	2018-19
Formal Labor Input	41.02	15.28	15.20
Capital Input	10.44	6.06	15.12
Intermediate Inputs	30.72	51.06	31.42
Overall	85.29	73.44	63.78

Note: The reported statistics are for $(\frac{\widetilde{Y}^Z}{Y}) - 1) * 100$, where $\frac{\widetilde{Y}^Z}{Y}$ $(\forall Z = f, K, Q)$ is computed using equations (107) to (109) in Appendix B.8.

Figure 8: Actual Size v/s Size Absent Formal Labor Distortion, 2018-19



Note: The left panel illustrates the plant size distribution (based on the number of employees) using a bin-width of 10, limited to plants employing less than or equal to 1000 workers. The right panel illustrates the OLS line based on the log-log plot of the same distribution, for all plants in the distribution.

This is illustrated in the first panel Figure 8.[50] Notably, the density of both small and medium sized plants would decrease in the counterfactual economy with no idiosyncratic variations in the formal labor wedge. In particular, the log-log plot in the second panel of Figure 8 clearly shows that the density of plants hiring 1000 workers or less will shrink, while those hiring more will increase in density. Overall, the counterfactual plant size distribution will exhibit a more dense right tail, favoring the existence of more large sized plants. This is in line with what policymakers can expect by removing size-based labor regulations – there will be less small-sized plants and more large-sized plants. However, Table 7 reports that the reallocation gains

stands at 2.7, reflecting high substitutability.

[49] The algebra behind removing the formal labor wedge was reported in Section II.6. The derivations for removing the capital and intermediate input wedge is shown in Appendix B.8.

[50] A qualitatively similar figure is found for both 2010-11 and 2014-15.

implied by eliminating variation in the formal labor wedge are substantially lower in 2014-15 and 2018-19, compared to 2010-11. If the variation in this wedge is indeed attributable to size-based labor regulations, this implies – eliminating variations in the labor regulations, such that plants of different sizes face the same formal labor regulations, would have lead to substantially smaller output gains in the 2014-15 and 2018-19, compared to 2010-11. Since larger plants are more productive (ref. Section VI.3) and the counterfactual economy indeed exhibits an increase in their density, why do the output gains from this policy decline?

V.2.2 Model-Based Reconciliation

This finding can be rationalized by the observed increase in informal worker usage in 2014-15 and 2018-19, relative to 2010-11 (ref. Figure 1 and Table 9 in Appendix A). In essence, with medium to large plants increasingly hiring informally, reallocating formal labor towards these plants[51] will lead to lower output gains compared to the case when informal labor usage is low. The intuition is straightforward – since these plants are able to sidestep regulations and meet their optimal labor input requirements by hiring informal workers, allocating these units more formal workers while keeping the other inputs fixed, will lead to limited increase in gross output of these plants.[52] The mechanism behind sidestepping regulations by hiring more informal workers can be seen by manipulating the first-order conditions of a given plant's profit maximization problem with respect to both labor inputs. The equilibrium informal to formal labor input ratio writes:

$$\frac{H_{c,si}}{H_{f,si}} = \left[\left(\frac{w_{f,s}(1+\tau_{f,si})}{w_{c,s}} \right) \left(\frac{\beta_{c,s}}{\beta_{f,s}} \right) \right]^{\nu} \quad ; \forall i \in s \tag{43}$$

Evidently, if the regulation induced distortion to the cost of hiring formal labor increases, the informal to formal labor ratio of the given plant also increases. Moreover, this effect is further augmented by the elasticity of substitution across both labor inputs (ν) i.e., the incremental increase in informal to formal labor input ratio is higher if formal labor can be easily substituted by their informal counterparts. As noted before, size-based labor regulations can be eliminated from the model by imposing an average formal labor input wedge that does not vary across plants within a given industry, this would alter equation (43) as follows:

$$\frac{H_{c,s}}{H_{f,s}} = \left[\left(\frac{w_{f,s}(1+\widetilde{\tau}_{f,s})}{w_{c,s}} \right) \left(\frac{\beta_{c,s}}{\beta_{f,s}} \right) \right]^{\nu} \quad ; \forall i \in s \tag{44}$$

[51] As highlighted before in Figure 7, on an average, the medium to large sized plants have a positive scaled formal labor wedge, meaning the reallocation exercise will increase their employment of formal workers.

[52] Note that the model assumes that all labor resources are fully utilised such that there is no unemployment. This means that informal workers are not fired in response to reallocation of formal workers, thus, the reallocation exercise will give medium to large sized plants greater amounts of formal labor, while keeping the amount of informal labor constant.

This implies that eliminating size-based labor regulations essentially amounts to a constant informal to formal labor input ratio for all plants within a given industry. In the data, this ratio will be higher for medium to large sized plants who face higher values of $\tau_{f,si}$, and thus the counterfactual exercise will reallocate large amounts of formal labor to these plants in order to maintain the constant ratio in equation (44), within each industry. Subsequently, although there will be an increase in gross output, the increase will be limited. To see this formally, we can rewrite the first-order condition with respect to formal labor reported in equation (94) in Appendix B.7, as follows:

$$\widetilde{P_{si}Y_{si}} = (\mathrm{TFPQ}_{si})^{\psi}\left(\frac{\sigma}{\sigma-1}\right)\left(\frac{\mathrm{MRPH}_{f,si}}{\alpha_{L,s}\beta_{f,s}}\right)\left(\frac{\widetilde{H}_{f,si}^{1-\theta}}{\widetilde{X_{si}}^{\psi-\theta}}\right) \tag{45}$$

Where $\widetilde{P_{si}Y_{si}}$ denotes the new gross output when size-based labor regulations are eliminated, and $\widetilde{X}_{si}$ denotes the aggregate labor input. There are several components that induce an increase in counterfactual gross output for plants constrained by the regulations - First, since larger plants are also more physically productive (high TFPQ_{si}), they will produce more in the counterfactual.[53] Second, a higher markup also induces higher production. Third, since regulations increase the marginal cost of employing formal labor, the constrained plants will have high $\mathrm{MRPH}_{f,si}$ (ref. equation (94)). This further implies that the additional revenue generated from an incremental increase in formal labor will be high for such plants. Finally, we observe that the increase in gross output will be limited by the fact that these plants are already hiring large amounts of informal labor i.e., the aggregate labor input (X_{si}) will already be high even before reallocating more formal labor to these plants. This last effect appears to be negating much of the positive effects on gross output, thus leading to a small increase in output from plants that are constrained by size-based labor regulations.

In other words, since the constrained plants meet their labor requirements by hiring informally, giving these plants more formal labor while keeping other factors constant, will not lead to substantial output gains. This happens because 'constrained' plants will hit diminishing returns to its total labor input faster. Thus, due to labor informality, output gains from removing the formal wedge (equivalently size-based labor regulations) are not high. Hence, we can conclude that formal labor misallocation will contribute less towards to aggregate misallocation, when labor informality is taken into account. Ignoring this phenomenon will thus lead to an incorrect vision of aggregate gains from such policy.

Overall, there are three main findings in this section that can be linked to the previously reported decrease in aggregate misallocation or equivalently the decrease in aggregate gains from reallocation for 2010-11 to 2018-19. First, majority of the aggregate misallocation in all

[53] The relationship between plant size and physical productivity is shown in Section VI.3.

three years is attributable to the formal labor and intermediate input wedges, with the capital input wedge playing a relatively smaller role. Second, although the formal labor input wedge was behind the high aggregate misallocation observed for 2010-11, its contribution declined substantially in 2014-15 and 2018-19 due to the increase in informal labor usage. Third, the decline in the contribution of the formal labor input wedge was behind the large decline in aggregate misallocation for 2014-15, while the decline in contribution of the intermediate input wedge drove majority of the decline in aggregate misallocation for 2018-19.

It is important to note that although aggregate misallocation has declined overtime, the intermediate input distortions have become as problematic as they were in 2010-11, while distortions in the capital input seems to have somewhat increased relative to 2010-11. Thus, the observed decline in aggregate misallocation does not seem to coming from policies that favored a more efficient allocation of resources, rather it is mostly a construct of high informal labor usage which subsequently reduced the importance of formal labor misallocation. Thus, if the recent 'ease of business' narrative behind the easing of the size-based labor regulations in India is seen from the perspective of potential increase in aggregate output, the gains appear to be small. Instead more recently, for 2018-19, the estimates in Table 7 suggests that most of potential output gains could have been realised by focusing on the distortions that generate misallocation in intermediate inputs.

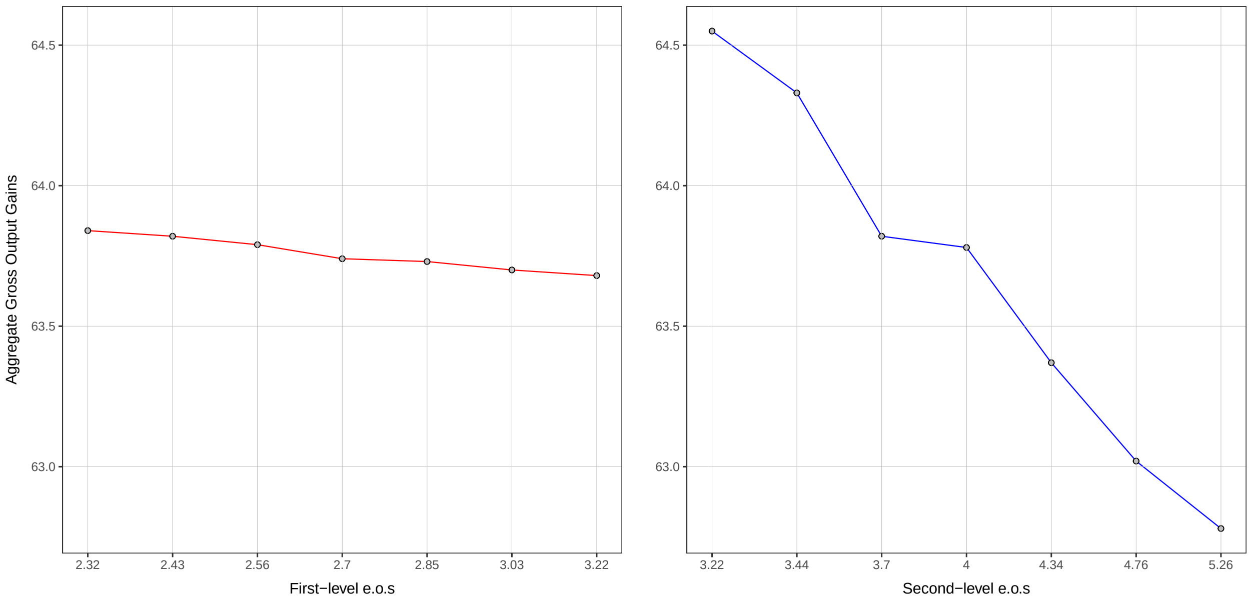
VI Robustness Checks

VI.1 Varying Factor Elasticity of Substitution

As previously highlighted in Section II.3, the dispersion of TFPR_{si} and hence the corresponding reallocation gains could be sensitive to the choice of the elasticity of substitution across both labor inputs (first-level e.o.s, henceforth) and across combined labor, capital and intermediate inputs (second-level e.o.s, henceforth). Figure 9 illustrates the sensitivity of the computed gross output reallocation gains as the elasticity parameters change. The left panel clearly evinces that although the reallocation gains decline in response to an increasing first-level e.o.s, the rate of decline is slow. On the other hand, the right panel reports that the reallocation gains decline much more rapidly when the second-level e.o.s increases.

The intuition behind the observed difference in sensitivity can be understood by a simple example. Suppose that a plant faces stronger distortions in intermediate inputs, but the second-level e.o.s. is high. Clearly, this plant can circumvent constraints on employing intermediate inputs by easily substituting towards more capital or labor, in effect reducing the impact of the intermediate input distortion. Alternatively, assume that the first-level e.o.s. is high. In this case, the plant facing strong distortions in intermediate inputs will not be able to easily shift towards more labor or capital to meet its desired scale of operation, and thus the intermediate

Figure 9: Aggregate Gross-Output Reallocation Gains, Varying E.O.S, 2018-19



Note: Aggregate reallocation gains are computed using $(\frac{Y^*}{Y} - 1) * 100$, where $\frac{Y^*}{Y}$ comes from equation (31). The baseline computations were based on $\nu = 2.7$ and $\eta = 4$, which is equivalent to assuming $\theta = 0.63$ and $\psi = 0.75$, respectively.

input distortion maintains its relevance regardless of a high first-level e.o.s. In other words, a higher second-level e.o.s is helpful in dodging distortions to each 'second-level' input, while a higher first-level e.o.s is only useful when the formal labor input distortion is high (ref. equation (43)).

Overall, we can conclude that a non-unitary elasticity of substitution across inputs is extremely relevant in explaining the observed misallocation in the economy. The misallocation literature based on the Hsieh and Klenow (2009) framework mostly operates on the canonical Cobb-Douglas specification for plant-level production, essentially assuming a unitary elasticity of substitution across inputs. Although a Cobb-Dogulas specification is much more tractable and easier to work with algebraically, I show that this assumption has strong implications for the measure of misallocation and the computation of hypothetical gains from reallocation. Working with a CES specification is a theoretically richer approach which can evidently capture this sensitivity and thus produces more robust estimates of misallocation.

VI.2 Computation of Labor Input and Elasticity across Goods

The baseline computations were done by measuring both types of labor input using their respective wage bills. This was based on the assumption that differences in worker skills and hours worked are captured by wages, such that the estimates of TFPR_{si} and TFPQ_{si} accounts for these aspects, rather than simply measuring labor input as the number of employees. However, if we refer to the literature on non-competitive labor markets, wages can potentially reflect rent sharing between the plant and the workers. If this is indeed the case, more productive plants might be paying higher wages and thus hire less than they otherwise would, amplifying the gaps in TFPR and hence increasing misallocation. This implies that measuring labor input as their respective wage bill might understate the gaps in TFPR and hence the extent of misallocation. Interestingly, reallocation gains appear to slightly increase for 2010-11 and decrease for 2014-15 and 2018-19, compared to the baseline estimates. This implies that wage differences from rent sharing indeed amplified TFPR differences in 2010-11, however, rent sharing does not seem to be much of an issue for 2014-15 and 2018-19.

Table 8: Aggregate Reallocation Gains, Number of Employees as Labor Input

Output Measure	2010-11	2014-15	2018-19
Gross Output	86.01	73.24	64.03
Value Added	215.02	193.18	165.70

Note: The reported statistics are for $(\frac{Y^*}{Y} - 1) * 100$, where $\frac{Y^*}{Y}$ is computed from equation (31) using both gross output and value added measures of plant-level output. Plants with negative value-added are dropped in the computation.

The value of the elasticity of substitution across plant-level differentiated goods (σ) was chosen to be 3 in the baseline computations. As shown in Hsieh and Klenow (2009), the reallocation gains are highly sensitive to the calibration of this elasticity parameter. Intuitively, for a given productivity distribution and given distortions, when σ rises or equivalently when markups fall for all plants within a given industry, we know that each plant will produce more. This implies that plants that are potentially benefiting from implicit subsidies will also produce more i.e., utilise more resources than they otherwise would in an efficient allocation. This clearly implies that the extent of misallocation will increase and thus the reallocation gains will be higher.[54] Following Hsieh and Klenow (2009), I raise the elasticity of substitution to 5 to check the sensitivity of reallocation results. As expected, reallocation gains appear to rise. In particular, on an average across all three years, economy-wide value added gains are around 195 percent compared to the baseline (with $\sigma = 3$) average of 190 percent. Thus, the reallocation gains reported in Table 3 should be seen as lower bound estimates, with a conservative choice for the elasticity of substitution across plant-level output.

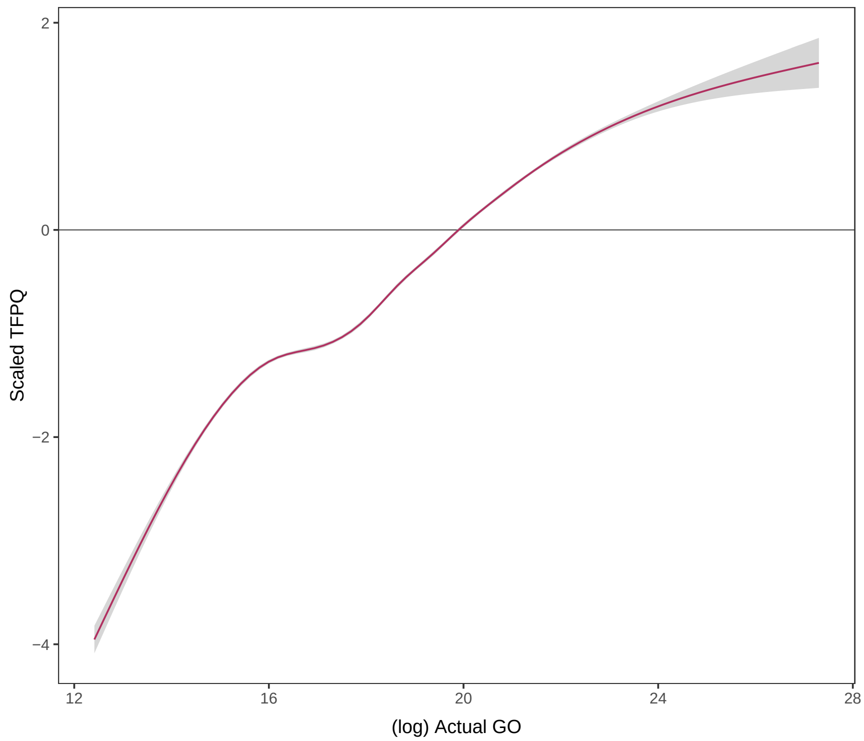
VI.3 Relationship between Plant-Size and Productivity

Conventional productivity theory suggests that larger plants should be more productive owing to the benefits of economies of scale. If the larger plants in turn face relatively higher

[54]Consistent with my argument, Hsieh and Klenow (2009) remark that a higher σ will mean that TFPR gaps will be closed more slowly in response to a reallocation of factors, hence leading to higher reallocation gains.

distortions, the economy loses more potential output or equivalently the potential gains from reallocation towards these plants is higher. Indeed, this was the main intuition behind equation (27), and an explanatory component of equation (45). As a robustness check, I investigate whether this relationship holds. Figure 10 illustrates this by looking at the association between plant size (gross output) and scaled TFPQ.

Figure 10: Plant-Size and Productivity



Note: Scaled TFPQ is measured by $ln(A_{si}W_s^{\frac{1}{\sigma-1}}/\text{TFPQ}_s^*)$ and made robust to outliers by removing the top and bottom 1% tails. Sample weights are applied to the gross output of each plant.

Evidently, we observe that plants that are technically more efficient i.e., plants with TFPQ_{si}, relative to the industry efficient TFPQ_s^*, tend to be larger in size. While those below the industry efficient TFPQ_s^* are comparatively smaller.

VII Conclusion

The problem of resource misallocation in developing economies is increasingly gaining more attention from researchers. This dissertation aims to improve our understanding of resource misallocation, by accounting for an important feature of developing economies, namely - the issue of labor informality. To understand the consequences of labor informality on the observed misallocation, I extend the widely used framework of Hsieh and Klenow (2009) by introducing heterogeneity in the labor input encompassed in a more general production technology. The model is based on heterogeneous firms in a monopolistic competition within industries, facing idiosyncratic distortions to factor prices. The firms are assumed to produce output based on a CRS two-level CES specification, using two types of labor – formal and informal in the first-level, combined with capital and intermediate inputs in the second. This modelling choice allows me capture three important elements.

First, due to the CES specification, the equilibrium ratio of informal to formal labor only depends on their relative factor prices. Since the distortion to formal labor prices comes from labor regulations, this allows me to parsimoniously capture the effect of regulatory costs on the equilibrium ratio of informal to formal labor. In line with intuition, as these regulatory costs increase, firms would sidestep them and hire more informal labor. This effect is further augmented by a higher elasticity of substitution across both types of labor input. The model is thus able to inculcate the idea of labor informality and the mechanism of sidestepping labor regulations.

Second, the separate introduction of both types of labor allows me to estimate a specific formal labor wedge from the data. Since the majority of constraining labor regulations only apply to formal workers, this wedge is able to capture the incidence of the regulatory costs across firms of different sizes. Looking at the relationship between this wedge and firm size, one can decipher whether labor regulations seem to bite the largest firms more than the smaller units. Further, capturing this wedge allows me to run a counterfactual exercise where size-based labor regulations are essentially removed from the economy.[55] We can subsequently look at the output gains from this exercise and compare it to the gains from removing distortions in other inputs. This essentially allows me to compare the contribution of each distortion towards aggregate misallocation.

Third, the generality of the two-level CES specification allows me to explicitly account for the elasticity of substitution across all inputs. This is crucial since it allows me to evaluate how the observed misallocation or gains from reallocation behave as the ease of substitution across inputs increases. The misallocation literature stemming from the work of Hsieh and Klenow (2009) often works on the assumption of unitary elasticity of substitution across inputs, by working with a Cobb-Douglas technology. Although this specification is much more tractable, it misses the idea that firms might shift to different inputs if the price distortions of a given input are particularly high. As a theoretical contribution, I show that reallocation gains are indeed declining when the ease of substitution across inputs increases.

The model is estimated using cross-sectional plant-level data (2010-11, 2014-15 and 2018-19) from the registered segment of Indian manufacturing, well-known for its increasing reliance on informal labor (Bertrand et al., 2021). I find that aggregate misallocation in formal

[55] This is done by removing the variation in the formal labor wedge, such that all units face the same formal labor wedge. In other words, if size-based regulations lead to different prices of formal labor across firms of different sizes, this exercise essentially eliminates such regulations.

manufacturing has declined over the 2010-11 to 2018-19 period. However, in absolute terms, misallocation appears to severely hinder the output of manufacturing plants. In particular, for 2018-19, aggregate manufacturing gross output output could have been higher by 63 percent under an efficient allocation of resources within each industry. On an average over 2010-11 to 2018-19, several industries with high output share seem to be suffering from an inefficient allocation of resources. These include – manufacturing of refined petroleum products, basic iron and steel, automobiles, textiles and pharmaceuticals. More recently, in 2018-19, the refined petroleum and the basic iron and steel industry appear to be suffering high misallocation. Thus, the relative decline in aggregate misallocation should not be taken at face value without addressing the high misallocation in such important industries.

Dissecting the observed decline in aggregate misallocation at the plant-level, several important findings emerge. First, the bite of factor misallocation appears to be strongly hindering the output of small and mid-sized plants. In particular, the smallest plants suffer from high distortions in capital prices, while the mid-sized plants face barriers in employing intermediate inputs. Notably, the latter issue appears to be increasing overtime. Second, although the largest plants suffer from distortions in formal labor, they benefit from implicit 'subsidies' in hiring capital and intermediate inputs. Importantly, although majority of the plants benefit from implicit subsidies in hiring informal labor, the larger plants benefit relatively more than the small and mid-sized plants. Third, the plant size distribution under an intra-industry efficient allocation of resources shows that the small and mid-sized plants would increase their output, while the largest plants would shrink. According to the model, the counterfactual output of plants would be higher if they face strong distortions overall, while the plants benefiting from implicit subsidies in factor prices would shrink in size. This leads me to suspect whether the highly debated size-based regulations are no longer constraining the output of the largest plants.

A counterfactual exercise which removes the size-based labor regulations indeed exhibits a higher density of large-sized plants, while the density of smaller plants declines. This is in line with the recent 'ease of business' narrative of Indian policymakers,[56] such that there will be more large-sized plants and less small-sized plants, benefiting the economy on an aggregate. In this context, I ask – First, what are the gains in removing size-based regulations in the context of labor informality. Second, could it be that distortions to capital and intermediate inputs constrain output more than these regulations. To investigate the relative importance of each distortion towards aggregate misallocation, I remove each factor distortion one at a time, and evaluate the aggregate gross output gains. Importantly, I find that removing size-based labor regulations would have benefited aggregate manufacturing output by a towering 41 percent

[56] *"..the pace of reforms in enabling ease of doing business need to be enhanced so that India can be ranked within the top 50 economies on this metric.* – Chapter 1 (p.7), Economic Survey of India (2019-20)
"..As India leapfrogs towards a five trillion-dollar economy by 2024-25, simplifying and maintaining a business-friendly regulatory environment is essential." – Chapter 6 (p.1), Economic Survey of India (2019-20)

in 2010-11, followed by a only a 15 percent increase in both 2014-15 and 2018-19. Rather, removing distortions in intermediate inputs leads to a 51 percent and 31 percent increase in manufacturing output for 2014-15 and 2018-19, respectively. Why does the relative gains from eliminating size-based labor regulations decline?

Reconciliation from the model suggests that the equilibrium ratio of informal to formal labor rises unambiguously in the presence of regulation induced costs of hiring formal labor. This effect is further augmented by a high elasticity of substitution between the two, estimated at around 2.7 by Padmakumar (2022). Given the rise in informal worker usage by those 'constrained' via labor regulations,[57] such plants seem to be readily meeting their labor requirements by hiring informally. Hence, the exercise of giving these plants more formal workers in the hope of generating higher output will run into diminishing returns to their total labor input quicker, since the total labor input is already high due to informal hiring. Thus, the policy instrument of easing size-based labor regulations must take into account labor informality, which has been evidently affording flexibility to the 'constrained' plants. Importantly, for 2018-19, I find that eliminating misallocation in intermediate inputs would have increased aggregate manufacturing output by 31 percent, much higher than the exercise of removing size-based labor regulations which leads to a 15 percent rise. Overall, if the aim of policymakers is to improve aggregate output, I argue that an isolated focus on promoting 'ease of business' through deregulating labor laws is a misguided trend in recent policy discussions.

Finally, and rather concerningly so, it appears that the contribution of capital and intermediate input distortions towards aggregate misallocation is as high in 2018-19, as for 2010-11.[58] Thus, the decline in aggregate misallocation appears to be majorly coming from a decline in the relevance of formal labor misallocation. This implies that a decade's worth of policies aimed at reducing resource misallocation have not benefited India's formal manufacturing sector. The decline in aggregate misallocation is instead a construct of the increasing incidence of labor informality.[59]

[57] Refer to Figure 1, 9 and also Bertrand et al. (2021).

[58] Capital distortions are actually more constraining in 2018-19, relative to 2010-11.

[59] Note that although labor informality appears to benefit aggregate output, we cannot simply focus on the aspect of efficiency. My framework is silent about the welfare outcomes of both types of workers, which should be taken into account when we speak of policies affecting informality. Unfortunately, this cannot be studied in the current framework.

Appendices

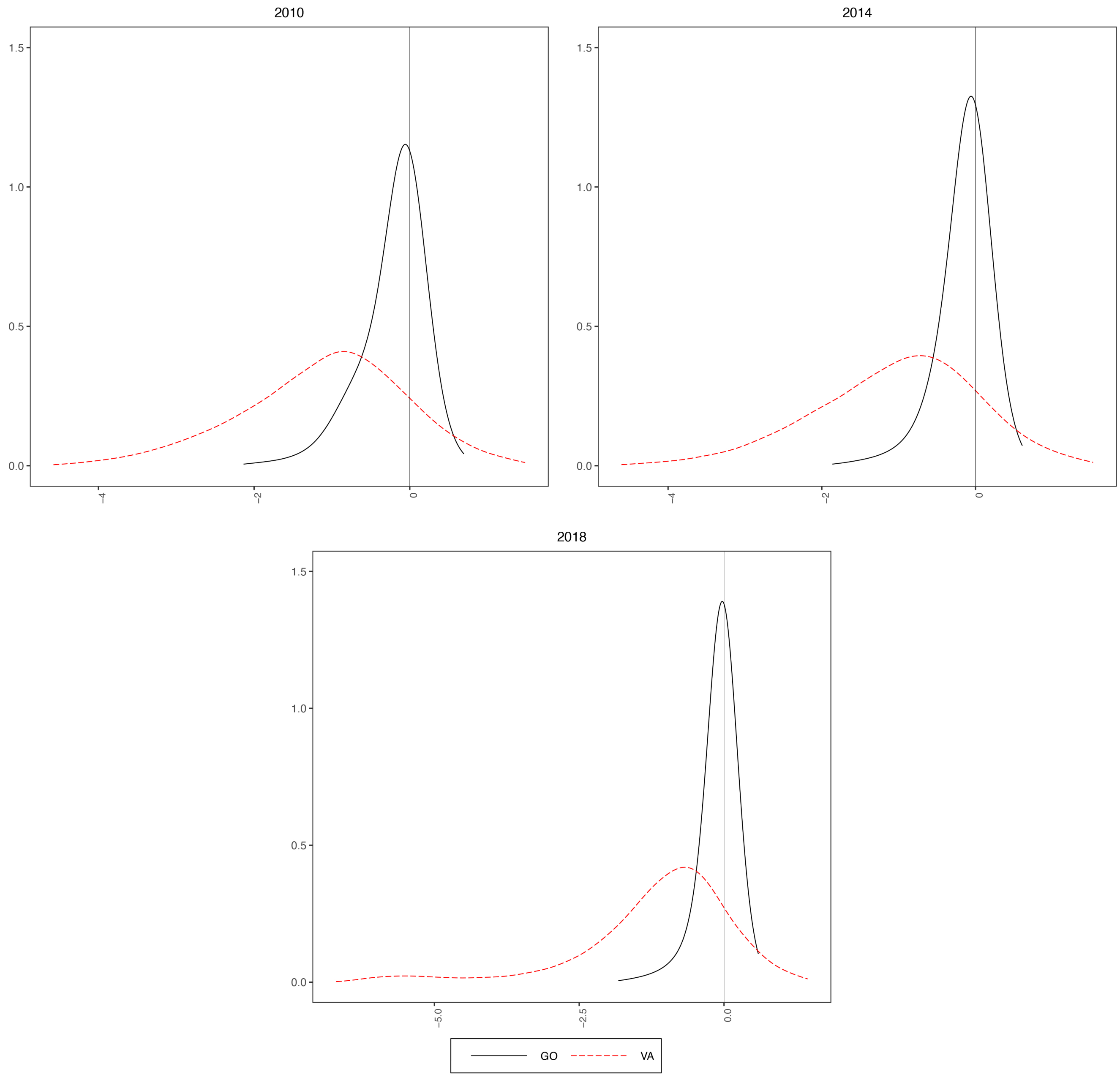
A Empirical Annexes

Table 9: Average Contract Worker Usage by Plant Size Bins

Plant Size	No. of Plants	Average Contract Worker Usage
	2010-11	
1-10	7,963	1
11-20	5,186	2
21-50	5,335	8
51-100	3,795	28
101-500	8,865	78
Above 500	2,198	487
	2014-15	
Plant Size	No. of Plants	Average Contract Worker Usage
1-10	11,119	1
11-20	5,899	2
21-50	5,874	8
51-100	4,786	26
101-500	10,402	85
Above 500	2,568	486
	2018-19	
Plant Size	No. of Plants	Average Contract Worker Usage
1-10	8,673	1
11-20	5,145	2
21-50	6,083	7
51-100	5,896	26
101-500	12,640	89
Above 500	3,177	547

Note: Due to missing values for relevant calculations later in this study, a number of plants will be dropped from the computation. To maintain consistency, the statistics reported in this table are also based on the same data. The average values are rounded to their nearest integer.

Figure 11: Dispersion in (log) Scaled TFPR, Gross Output v/s Value-Added



Note: The top and bottom 1% tails of the $ln(\text{TFPR}_{si}/\text{TFPR}^*_s)$ are removed to deal with outliers, for each output measure. Plants with negative value-added are removed from the computation.

Table 10: State-wise Reallocation Gains, Gross Output Approach

S.No.	State	2010-11	2014-15	2018-19	State Average
1.	Andaman & Nicobar	150.08	60.50	57.66	89.41
2.	Andhra Pradesh	84.25	68.74	70.91	74.63
3.	Arunachal Pradesh	-	92.90	127.37	110.35
4.	Assam	144.09	87.47	106.37	112.64
5.	Bihar	120.31	102.02	109.39	110.57
6.	Chandigarh	100.90	85.94	87.55	91.46
7.	Chattisgarh	106.98	92.73	96.64	98.78
8.	Dadra & Nagar Haveli	85.79	70.77	53.46	70.00
9.	Daman & Diu	75.11	63.67	71.39	70.05
10.	Delhi	74.88	62.02	57.41	64.77
11.	Goa	74.28	66.82	79.35	73.48
12.	Gujarat	103.51	87.77	64.31	85.19
13.	Harayana	60.08	58.97	54.65	57.90
14.	Himachal Pradesh	75.82	68.54	69.74	71.36
15.	Jammu & Kashmir	98.66	87.08	60.20	81.98
16.	Jharkhand	75.82	80.45	63.07	73.11
17.	Karnataka	87.01	73.42	67.02	75.81
18.	Kerala	97.23	80.11	49.23	75.52
19.	Madhya Pradesh	65.83	57.21	67.15	63.39
20.	Manipur	135.41	97.79	74.04	102.41
21.	Meghalaya	62.17	61.77	52.14	58.69
22.	Nagaland	123.96	79.64	64.61	89.40
23.	Odisha	112.98	100.31	115.65	109.64
24.	Puducherry	77.15	50.10	67.93	65.06
25.	Punjab	87.62	65.70	58.08	70.46
26.	Rajasthan	69.36	64.78	51.98	62.04
27.	Sikkim	88.01	107.51	105.58	100.36
28.	Tamil Nadu	68.75	59.83	49.03	59.20
29.	Telangana	-	70.89	70.12	47.00
30.	Tripura	164.97	100.83	61.61	109.13
31.	Uttar Pradesh	72.25	67.69	59.77	66.57
32.	Uttarakhand	72.25	59.71	58.87	63.61
33.	West Bengal	106.17	87.73	85.67	93.19
	Across State Average	93.99	76.32	71.95	

Note: The reported statistics are for $(\frac{Y_Z^*}{Y_Z} - 1) * 100$, where $\frac{Y_Z^*}{Y_Z}$ is computed from equation (46) using the gross output approach. The two missing blanks in 2010-11 are because – First, Telangana was carved out of the Andhra Pradesh in 2014, and hence does not have any data points for the 2010-11 cross-section. Second, the 2010-11 ASI cross-section does not provide data for the state of Arunachal Pradesh.

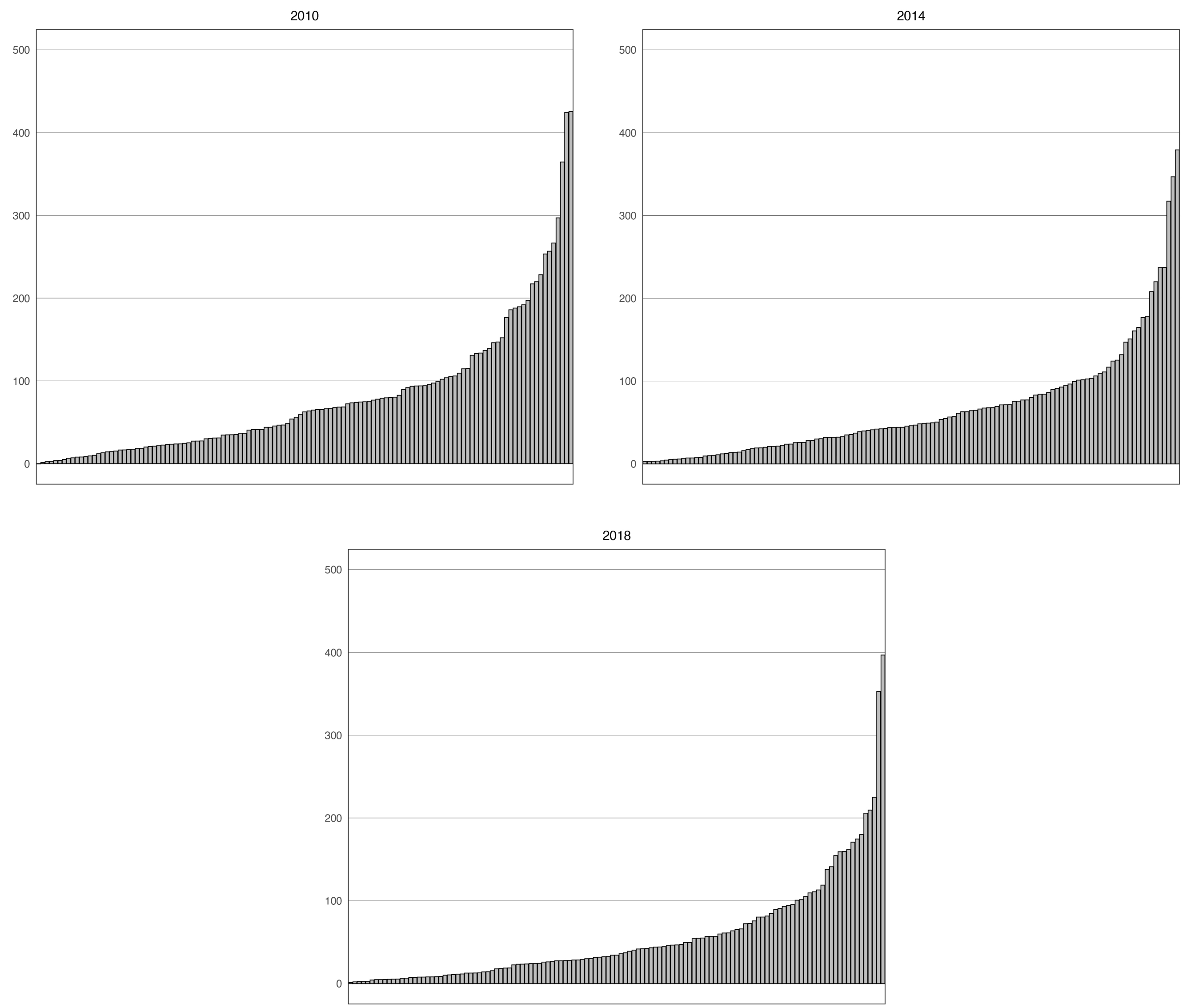
This exercise is done by simply altering the industry gross output shares in equation (31), such that instead of computing shares as a fraction of economy-wide output, the shares are computed relative to each state's aggregate output. Mathematically, using Z to index the regions, this amounts to the following:

$$\frac{Y_Z^*}{Y_Z} = \prod_{s=1}^{S\in Z} \left\{ \frac{Y_{s,Z}^*}{Y_{s,Z}} \right\}^{\mu_{s,Z}} = \prod_{s=1}^{S\in Z} \left\{ \left[\frac{\sum_{i=1}^{N_s\in S\in Z} \lambda_i (A_{si})^{\sigma-1}}{\sum_{i=1}^{N_s\in S\in Z} \lambda_i \left(A_{si} \frac{\mathrm{TFPR}_s^*}{\mathrm{TFPR}_{si}} \right)^{\sigma-1}} \right]^{\frac{\sigma}{\sigma-1}} \right\}^{\mu_{s,Z}} \quad ; \forall Z \in \{1, .., 33\} \quad (46)$$

Table 10 reports the findings. The issue of resource misallocation in formal manufacturing seems to be be particularly high in the North-Eastern and Eastern parts of India. The former is well-known to be isolated from the rest of the country, with rough terrain and weaker transportation lines. The high misallocation in these regions could thus be stemming from high intermediate input distortions, which I indeed find to be the most important source of misallocation in the aggregate. On an average over 2010-11 to 2018-19, several states appear to be suffering from high misallocation, including – Assam, Bihar, Arunachal Pradesh, Odisha, Tripura and Sikkim.

Several states seem to have consistently improved their allocative efficiency over the 2010-11 to 2018-19 period. In particular, those with substantial decline in misallocation include – Andaman & Nicobar Islands, Gujarat, Jammu & Kashmir, Kerala, Manipur, Nagaland and Tripura. Notably, The latter three states belong to the North-Eastern part of the country and have shown substantial improvements in their allocative efficiency in 2018-19. Explaining this across state variation is beyond the scope and focus of my study. However, reassuringly, comparing the across state average in each year shows the declining pattern in aggregate misallocation, as we would expect from the results of Sections IV.1 and IV.2. Thus, this exercise can be understood as another detailed robustness check of the results in Sections IV.1 and IV.2. Explaining the variation in misallocation across states although remains a relevant exercise from policy perspective, and can be taken up in subsequent research projects.

Figure 12: Industry Gross-Output Reallocation Gains, Ranked in Ascending Order



Note: Industry reallocation gains are computed using $(\frac{Y_s^*}{Y_s} - 1) * 100$, where $\frac{Y_s^*}{Y_s}$ is computed from equation (27).

Table 11: Industry Reallocation Gains, Ranked by Gains

S.No	4-Digit Industry (NIC, 2008)	Gross Output Reallocation Gains	Gross Output Share
	2010-11		
1.	Manufacture of bicycles and invalid carriages	425.62	0.26
2.	Manufacture of clay building materials	424.38	0.15
3.	Manufacture of soft drinks; production of mineral waters and other bottled waters	364.44	0.33
4.	Manufacture of tanks, reservoirs and containers of metal	296.97	0.26
5.	Saw milling and planing of wood	266.56	0.01
6.	Manufacture of plastics and synthetic rubber in primary forms	256.69	0.75
7.	Manufacture of other food products n.e.c.	253.33	1.28
8.	Manufacture of other general-purpose machinery	228.31	0.69
9.	Manufacture of basic precious and other non-ferrous metals	219.91	1.78
10.	Manufacture of bakery products	217.22	0.24
	2014-15		
1.	Manufacture of articles of fur	379.16	0.002
2.	Manufacture of plastics and synthetic rubber in primary forms	346.79	0.56
3.	Manufacture of basic precious and other non-ferrous metals	317.27	1.46
4.	Manufacture of metal-forming machinery and machine tools	237.15	0.15
5.	Manufacture of agricultural and forestry machinery	236.98	0.38
6.	Manufacture of tobacco products	219.98	0.67
7.	Distilling, rectifying and blending of spirits; ethyl alcohol production from fermented materials	207.98	0.85
8.	Manufacture of soft drinks; production of mineral waters and other bottled waters	177.80	0.46
9.	Manufacture of clay building materials	176.72	0.13
10.	Manufacture of other fabricated metal products n.e.c.	164.76	0.90
	2018-19		
1.	Manufacture of tobacco products	396.88	0.58
2.	Manufacture of communication equipment	352.81	0.61
3.	Manufacture of consumer electronics	225.08	0.10
4.	Manufacture of other non-metallic mineral products n.e.c.	209.54	0.19
5.	Manufacture of soft drinks; production of mineral waters and other bottled waters	205.85	0.38
6.	Distilling, rectifying and blending of spirits; ethyl alcohol production from fermented materials	180.04	1.09
7.	Manufacture of computers and peripheral equipment	174.61	0.16
8.	Manufacture of watches and clocks	170.82	0.04
9.	Manufacture of plastics and synthetic rubber in primary forms	161.95	0.94
10.	Manufacture of refined petroleum products	159.62	10.21

Note: The reported statistics for a given industry s are from $(\frac{Y_s^*}{Y_s} - 1) * 100$, where $\frac{Y_s^*}{Y_s}$ is computed using equation (27). Gross output share corresponds to $\mu_s = \frac{P_s Y_s}{PY}$ in equation (6), where $P_s Y_s = \sum_{i=1}^{N_s} \lambda_i P_{si} Y_{si}$. All correlation estimates are siginificant at the 1% level.

B Mathematical Derivations

B.1 Final Good Producer Optimization Problem

The profit maximization problem of the representative final good producing firm is given by:

$$\max_{Y_s} \Pi_F = P \prod_{s=1}^{S} Y_s^{\mu_s} - \sum_{s=1}^{S} P_s Y_s \tag{47}$$

Using the final good's price P as the numeraire, the first order condition gives:

$$\frac{\partial \Pi_F}{\partial Y_s} = 0 \implies P_s Y_s = \mu_s P Y \tag{48}$$

B.2 Industry Optimization Problem

A given industry combines the differentiated products of all plants within that industry, the demand for the plant's output is then derived from the industry profit maximization problem:

$$\max_{Y_{si}} \Pi_s = P_s \left(\sum_{i}^{N_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} - \sum_{i=1}^{N_s} P_{si} Y_{si} \tag{49}$$

The first order condition gives:

$$\frac{\partial \Pi_s}{\partial Y_{si}} = 0 \implies P_s \left(\sum_{i}^{N_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} Y_{si}^{-\frac{1}{\sigma}} = P_{si} \tag{50}$$

Using the expression for the industry production function, we get:

$$P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} = P_{si} \tag{51}$$

$$P_s Y_s^{\frac{1}{\sigma}} = P_{si} Y_{si}^{\frac{1}{\sigma}} \tag{52}$$

B.3 Plant Optimization Problem

B.3.1 Profit Minimization

Since the plants compete in a monopolistic environment within each industry, each plant internalizes it's market power in variety i by taking into account the demand curve it faces from the industry (s) from equation (52). Using the dual problem and solving for the cost function (denoted as TC_{si} for the time being), a given plants's profit maximization problem writes:

$$\max_{Y_{si}} \Pi_{si} = (1-\tau_{Y_{si}})P_{si}Y_{si} - TC_{si} \tag{53}$$

$$\text{s.t. } P_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} \implies P_{si}Y_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{\frac{\sigma-1}{\sigma}} \tag{54}$$

Note that although a given plants recognizes its market power in variety i, it takes economy wide aggregates as given - the essence of monopolistic competition. This implies that although Y_s clearly depends on Y_{si}, it is take as given, and can be thus thought of as a constant in the above maximization problem. Plugging the constraint in equation (54) in the objective function, we get:

$$\max_{Y_{si}} \Pi_{si} = (1-\tau_{Y_{si}})P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{\frac{\sigma-1}{\sigma}} - C_{si}(.) \tag{55}$$

The first order condition gives:

$$\frac{\partial \Pi_{si}}{\partial Y_{si}} = 0 \implies (1-\tau_{Y_{si}})P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} \left(\frac{\sigma-1}{\sigma}\right) = MC_{si}(.) \tag{56}$$

We know that $P_{si} = P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}}$, using this in the above equation, we get the standard condition of price equalling a markup over the marginal cost:

$$P_{si} = \left(\frac{\sigma}{\sigma-1}\right)\frac{MC_{si}(.)}{(1-\tau_{Y_{si}})} \tag{57}$$

B.3.2 Two-Stage Cost Minimization

The first level cost minimization problem writes:[60]

$$\min_{H_{c,si},H_{f,si}} \quad w_{c,s}H_{c,si} + (1+\tau_{f_si})w_{f,s}H_{f,si} \tag{58}$$

$$\text{s.t. } (\beta_{c,s}H_{c,si}^{\theta} + \beta_{f,s}H_{f,si}^{\theta})^{\frac{1}{\theta}} \geq X_{si} \implies \beta_{c,s}H_{c,si}^{\theta} + \beta_{f,s}H_{f,si}^{\theta} \geq X_{si}^{\theta} \tag{59}$$

The Lagrangian of the above problem writes:

$$\mathcal{L}_{si} = w_{c,s}H_{c,si} + (1+\tau_{f_si})w_{f,s}H_{f,si} + \lambda_{si}\left\{X_{si}^{\theta} - H_{c,si}^{\theta} + \beta_{f,s}H_{f,si}^{\theta}\right\} \tag{60}$$

The first order conditions with respect to $H_{c,si}$ and $H_{f,si}$, respectively, gives:

$$w_{c,s} = \lambda_{si}\beta_{c,s}\theta(H_{c,si})^{\theta-1} \tag{61}$$

$$(1+\tau_{f_{si}})w_{f,s} = \lambda_{si}\beta_{f,s}\theta(H_{f,si})^{\theta-1} \tag{62}$$

[60]Here, X_{si} can be thought of as a the output produced solely based on the two types of labor inputs.

These first-order conditions can be re-expressed as:

$$H_{c,si}^{\theta} = (w_{c,s})^{\frac{\theta}{\theta-1}} \beta_{c,s}^{-\frac{\theta}{\theta-1}} (\lambda_{si}\theta)^{-\frac{\theta}{\theta-1}} \tag{63}$$

$$H_{f,si}^{\theta} = [(1+\tau_{f_{si}})w_f]^{\frac{\theta}{\theta-1}} \beta_{f,s}^{-\frac{\theta}{\theta-1}} (\lambda_{si}\theta)^{-\frac{\theta}{\theta-1}} \tag{64}$$

Plugging equations (63) and (64) in constraint (59) with equality, we get:

$$(\lambda_{si}\theta)^{-\frac{\theta}{\theta-1}} = \frac{X_{si}^{\theta}}{\left[w_c^{\frac{\theta}{\theta-1}} \beta_1^{-\frac{1}{\theta-1}} + [(1+\tau_{L_{si}})w_f]^{\frac{\theta}{\theta-1}} \beta_2^{-\frac{1}{\theta-1}} \right]} \tag{65}$$

Plugging equation (65) in equations (63) and (64), we get the following conditional factor demands for each type of labor input:

$$H_{c,si} = w_{c,s}^{\frac{1}{\theta-1}} \beta_{c,s}^{-\frac{1}{\theta-1}} \left[w_{c,s}^{\frac{\theta}{\theta-1}} \beta_{c,s}^{-\frac{1}{\theta-1}} + [(1+\tau_{f_{si}})w_{f,s}]^{\frac{\theta}{\theta-1}} \beta_{f,s}^{-\frac{1}{\theta-1}} \right]^{-\frac{1}{\theta}} X_{si} \tag{66}$$

$$H_{f,si} = [(1+\tau_{f_{si}})w_{f,s}]^{\frac{1}{\theta-1}} \beta_{f,s}^{-\frac{1}{\theta-1}} \left[w_{c,s}^{\frac{\theta}{\theta-1}} \beta_{c,s}^{-\frac{1}{\theta-1}} + [(1+\tau_{f_{si}})w_{f,s}]^{\frac{\theta}{\theta-1}} \beta_{f,s}^{-\frac{1}{\theta-1}} \right]^{-\frac{1}{\theta}} X_{si} \tag{67}$$

Plugging in the conditional factor demands in the objective function, we get the first level cost function, denoted by $TC_{si}^{(1)}$, as follows:

$$TC_{si}^{(1)} = w_{c,s}H_{c,si} + (1+\tau_{f_{si}})w_{f,s}H_{f,si} \tag{68}$$

$$TC_{si}^{(1)} = X_{si} \left[w_{c,s}^{\frac{\theta}{\theta-1}} \beta_{c,s}^{-\frac{1}{\theta-1}} + [(1+\tau_{f_{si}})w_{f,s}]^{\frac{\theta}{\theta-1}} \beta_{f,s}^{-\frac{1}{\theta-1}} \right]^{\frac{\theta-1}{\theta}} \tag{69}$$

Using the expression for $TC_{si}^{(1)}$, the second level cost minimization problem writes:

$$\min_{X_{si},K_{si},Q_{si}} \quad TC_{si}^{(1)}(w_{c,s}, w_{f,s}, \beta_{c,s}, \beta_{f,s}, \tau_{f_{si}}, X_{si}, \nu) + r_s K_{si} + q_s Q_{si} \tag{70}$$

$$\text{s.t. } A_{si}\left[\alpha_{L,s}(\beta_{c,s}H_{c,si}^{\theta} + \beta_{f,s}H_{f,si}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s}K_{si}^{\psi} + \alpha_{Q,s}Q_{si}^{\psi} \right]^{\frac{1}{\psi}} \geq Y_{si}$$

Using the same steps as done for the first-level, we arrive at the following second-level cost function:

$$TC^{(2)}_{si} = \frac{Y_{si}\left(\frac{\sigma}{\sigma-1}\right)}{A_{si}(1-\tau_{Y_{si}})}\Bigg\{\left[\left(\frac{w_{c,s}^{\frac{\theta}{\theta-1}}}{\beta_{c,s}^{\frac{1}{\theta-1}}} + \frac{(1+\tau_{f_{si}})^{\frac{\theta}{\theta-1}} w_{f,s}^{\frac{\theta}{\theta-1}}}{\beta_{f,s}^{\frac{1}{\theta-1}}}\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\psi}{\psi-1}} \frac{1}{\alpha_{L,s}^{\frac{1}{\psi-1}}} + \frac{(1+\tau_{K_{si}})^{\frac{\psi}{\psi-1}} r_s^{\frac{\psi}{\psi-1}}}{\alpha_{K,s}^{\frac{1}{\psi-1}}} + \frac{(1+\tau_{Q_{si}})^{\frac{\psi}{\psi-1}} q_s^{\frac{\psi}{\psi-1}}}{\alpha_{Q,s}^{\frac{1}{\psi-1}}}\Bigg\}^{\frac{\psi-1}{\psi}} \quad ;\theta,\psi<1,\sigma>1 \quad (71)$$

Rewriting equation (71) using $\nu = 1/(1-\theta)$ and $\eta = 1/(1-\psi)$ gives:

$$TC^{(2)}_{si} = \frac{Y_{si}\left(\frac{\sigma}{\sigma-1}\right)}{A_{si}(1-\tau_{Y_{si}})}\Bigg\{(\alpha_{L,s})^{\nu}\left[\beta_{c,s}^{\nu} w_{c,s}^{1-\nu} + \beta_{f,s}^{\nu} w_{f,s}^{1-\nu}(1+\tau_{f_{si}})^{1-\nu}\right]^{\frac{1-\eta}{1-\nu}} + (\alpha_{K,s})^{\eta} r_s^{1-\eta}(1+\tau_{K_{si}})^{1-\eta} + (\alpha_{Q,s})^{\eta} q_s^{1-\eta}(1+\tau_{Q_{si}})^{1-\eta}\Bigg\}^{\frac{1}{1-\eta}} \quad (72)$$

B.4 Expression for TFPR$_{si}$

Taking a partial derivative w.r.t to Y_{si} and plugging the marginal cost in the plant's pricing equation (57), gives:

$$P_{si} = \frac{\left(\frac{\sigma}{\sigma-1}\right)}{A_{si}(1-\tau_{Y_{si}})}\Bigg\{(\alpha_{L,s})^{\nu}\left[\beta_{c,s}^{\nu} w_{c,s}^{1-\nu} + \beta_{f,s}^{\nu} w_{f,s}^{1-\nu}(1+\tau_{f_{si}})^{1-\nu}\right]^{\frac{1-\eta}{1-\nu}} + (\alpha_{K,s})^{\eta} r_s^{1-\eta}(1+\tau_{K_{si}})^{1-\eta} + (\alpha_{Q,s})^{\eta} q_s^{1-\eta}(1+\tau_{Q_{si}})^{1-\eta}\Bigg\}^{\frac{1}{1-\eta}} \quad (73)$$

Subsequently, since TFPR$_{si} = P_{si}A_{si}$, we can simply manipulate equation (73) to arrive at the final expression for TFPR$_{si}$ reported in the main text:

$$\text{TFPR}_{si} = \frac{\left(\frac{\sigma}{\sigma-1}\right)}{(1-\tau_{Y_{si}})}\Bigg\{(\alpha_{L,s})^{\nu}\left[\beta_{c,s}^{\nu} w_{c,s}^{1-\nu} + \beta_{f,s}^{\nu} w_{f,s}^{1-\nu}(1+\tau_{f_{si}})^{1-\nu}\right]^{\frac{1-\eta}{1-\nu}} + (\alpha_{K,s})^{\eta} r_s^{1-\eta}(1+\tau_{K_{si}})^{1-\eta} + (\alpha_{Q,s})^{\eta} q_s^{1-\eta}(1+\tau_{Q_{si}})^{1-\eta}\Bigg\}^{\frac{1}{1-\eta}} \quad (74)$$

B.5 Building the Efficient Counterfactual[61]

B.5.1 Efficient level of Real Output and Nominal Output

Computing the efficient level of output is the first step in arriving at the potential gains from a counterfactual reallocation exercise. To do so, I first exploit the industry (s) demand for plant-level output given in equation (52), setting $Y_s^{\frac{1}{\sigma}} P_s = 1$.[62] This gives:

$$P_{si} = Y_s^{\frac{1}{\sigma}} P_s Y_{si}^{-\frac{1}{\sigma}} \implies P_{si} = Y_{si}^{-\frac{1}{\sigma}} \tag{75}$$

Next, using the fact that $\text{TFPR}_{si} = P_{si} A_{si}$, we get:

$$Y_{si}^{\frac{1}{\sigma}} = \frac{A_{si}}{\text{TFPR}_{si}} \implies Y_{si} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma} \tag{76}$$

Note that this gives an expression for the real units of output which is otherwise not available when there is no separate data on plant-level prices. Thus, I start from the fact that prices must be lower for consumers to demand more plant-leve output. Subsequently, I raise $P_{si} Y_{si}$ to the power of $\sigma/(\sigma - 1)$ to arrive at Y_{si}. That is equivalent to inferring the separation in price and quantity from revenue and an assumed elasticity of demand, σ (Hsieh and Klenow, 2009).[63]Multiplying equation (76) by P_{si} and using the fact that $P_{si} = Y_{si}^{-\frac{1}{\sigma}}$, we also get an expression for nominal output:

$$P_{si} Y_{si} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma-1} \tag{77}$$

Let TFPR_s^* denote the efficient level of TFPR in a given industry s, under an efficient allocation of resources, such that there is no variation in TFPR_{si}. Imposing TFPR_s^* and using equations (76) and (77), the efficient level of real and nominal output at the plant-level write:

$$Y_{si}^* = \left(\frac{A_{si}}{\text{TFPR}_s^*}\right)^{\sigma} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma} \left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma} = Y_{si} \left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma} \tag{78}$$

$$(P_{si} Y_{si})^* = \left(\frac{A_{si}}{\text{TFPR}_s^*}\right)^{\sigma-1} = \left(\frac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma-1} \left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma-1} = P_{si} Y_{si} \left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s^*}\right)^{\sigma-1} \tag{79}$$

[61]This section is inspired from the work of Dias et al. (2016, 2018).

[62]I argue why we can set $Y_s^{\frac{1}{\sigma}} P_s = 1$ without any consequence to my counterfactual exercise, in section B.6.

[63]Note that $Y_{si} = \dfrac{A_{si}}{\text{TFPR}_{si}} \implies Y_{si} = \left(\dfrac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma} = (P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}}$

B.5.2 Expression for TFPR$_s^*$

By definition, the expression for the efficient level of industry TFPR writes:

$$\text{TFPR}_s^* = \frac{(P_s Y_s)^*}{\left[\alpha_{L,s}(\beta_c H_{c,s}^{\theta} + \beta_f H_{f,s}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s} K_s^{\psi} + \alpha_{Q,s} Q_s^{\psi}\right]^{\frac{1}{\psi}}} \tag{80}$$

Using a sample-weighted sum of efficient plant-level output, we can write $P_s Y_s = \sum_i^{N_s} (P_{si} Y_{si})^*$. Further we can use the expression of $(P_{si} Y_{si})^*$ from equation (79). This gives:

$$\text{TFPR}_s^* = \frac{\sum_{i=1}^{N_s} \lambda_i (P_{si} Y_{si})^*}{\left[\alpha_{L,s}(\beta_c H_{c,s}^{\theta} + \beta_f H_{f,s}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s} K_s^{\psi} + \alpha_{Q,s} Q_s^{\psi}\right]^{\frac{1}{\psi}}} \tag{81}$$

$$\text{TFPR}_s^* = \frac{\dfrac{\sum\limits_{i=1}^{N_s} A_{si}^{\sigma-1}}{(\text{TFPR}_s^*)^{\sigma-1}}}{\left[\alpha_{L,s}(\beta_c H_{c,s}^{\theta} + \beta_f H_{f,s}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s} K_s^{\psi} + \alpha_{Q,s} Q_s^{\psi}\right]^{\frac{1}{\psi}}} \tag{82}$$

Rearranging the expression for TFPR$_s^*$ in equation (82), we get the final expression mentioned in the text:

$$\text{TFPR}_s^* = \left\{ \frac{\sum\limits_{i=1}^{N_s} A_{si}^{\sigma-1}}{\left[\alpha_{L,s}(\beta_c H_{c,s}^{\theta} + \beta_f H_{f,s}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s} K_s^{\psi} + \alpha_{Q,s} Q_s^{\psi}\right]^{\frac{1}{\psi}}} \right\} \tag{83}$$

B.5.3 Reallocation Gains

Using the expression for efficient plant-level output from equation (78) and applying the CES aggregator at the industry-level (with sample weights), we can take a ratio of efficient and observed industry output.

$$\frac{Y_s^*}{Y_s} = \frac{\left[\sum\limits_{i=1}^{N_s} \lambda_i (Y_{si})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}}{\left[\sum\limits_{i=1}^{N_s} \lambda_i (Y_{si}^*)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}} = \frac{\left[\sum\limits_{i=1}^{N_s} \lambda_i \left\{\left(\dfrac{A_{si}}{\text{TFPR}_s^*}\right)^{\sigma}\right\}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}}{\left[\sum\limits_{i=1}^{N_s} \lambda_i \left\{\left(\dfrac{A_{si}}{\text{TFPR}_{si}}\right)^{\sigma}\right\}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}} \tag{84}$$

$$\frac{Y_s^*}{Y_s} = \frac{\left[\sum_{i=1}^{N_s} \lambda_i \left\{\frac{A_{si}}{\text{TFPR}_s^*}\right\}^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}}}{\left[\sum_{i=1}^{N_s} \lambda_i \left\{\frac{A_{si}}{\text{TFPR}_{si}}\right\}^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}}} = \left[\frac{\sum_{i=1}^{N_s} \lambda_i (A_{si})^{\sigma-1}}{\sum_{i=1}^{N_s} \lambda_i \left(A_{si} \frac{\text{TFPR}_s^*}{\text{TFPR}_{si}}\right)^{\sigma-1}}\right]^{\frac{\sigma}{\sigma-1}} \tag{85}$$

Applying the Cobb-Douglas aggregator for the final goods producer from equation (5), the aggregate output gains can be expressed as:

$$\frac{Y^*}{Y} = \prod_{s=1}^{S} \left\{\frac{Y_s^*}{Y_s}\right\}^{\mu_s} \tag{86}$$

$$\frac{Y^*}{Y} = \prod_{s=1}^{S} \left\{\left[\frac{\sum_{i}^{N_s} (A_{si})^{\sigma-1}}{\left[\sum_{i}^{N_s} \left(A_{si} \frac{\text{TFPR}_s^*}{\text{TFPR}_{si}}\right)^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}}}\right]^{\frac{\sigma}{\sigma-1}}\right\}^{\mu_s} \tag{87}$$

B.6 Backing out TFPQ$_{si}$ and Ignoring Industry Constant

I deploy the following "trick" to back out an expression for A_{si} using the model's equations. First, starting with the solution to the industry (s) profit maximization problem in equation (10):

$$\begin{aligned}
P_{si} Y_{si} &= P_s Y_s^{\frac{1}{\sigma}} Y_{si}^{\frac{\sigma-1}{\sigma}} \\
(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}} \left(\frac{Y_s^{-\frac{1}{\sigma}}}{P_s}\right)^{\frac{\sigma}{\sigma-1}} &= Y_{si} \\
(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}} \frac{(Y_s^{\frac{1}{\sigma}})^{\frac{\sigma}{1-\sigma}}}{P_s^{\frac{\sigma}{\sigma-1}}} &= Y_{si} \\
(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}} \left(P_s Y_s^{\frac{1}{\sigma}}\right)^{\frac{\sigma}{1-\sigma}} &= Y_{si}
\end{aligned}$$

Now, using the definition of A_{si}, we get:

$$A_{si} = \frac{Y_{si}}{\left[\alpha_{L,s} X_{si}^{\psi} + \alpha_{K,s} K_{si}^{\psi} + \alpha_{Q,s} Q_{si}^{\psi}\right]^{\frac{1}{\psi}}} = \frac{(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}} \left(P_s Y_s^{\frac{1}{\sigma}}\right)^{\frac{\sigma}{1-\sigma}}}{\left[\alpha_{L,s} X_{si}^{\psi} + \alpha_{K,s} K_{si}^{\psi} + \alpha_{Q,s} Q_{si}^{\psi}\right]^{\frac{1}{\psi}}} \tag{88}$$

Then, denoting $\kappa_s = \left(P_s Y_s^{\frac{1}{\sigma}}\right)^{\frac{\sigma}{1-\sigma}}$ as an industry-level constant, we get:

$$A_{si} = \kappa_s \frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{\left[\alpha_{L,s}X_{si}^{\psi} + \alpha_{K,s}K_{si}^{\psi} + \alpha_{Q,s}Q_{si}^{\psi}\right]^{\frac{1}{\psi}}} \tag{89}$$

Since I care about an intra-industry reallocation exercise, the dispersion of TFPR_{si} matters. We already know that variation in prices only come from the variation in idiosyncratic distortions (ref. equation (73)). Further, if we can show that the dispersion in A_{si} does not depend on the industry constant, the dispersion of TFPR_{si} will also not depend on this constant. The empirical exercise is based on the dispersion of the logarithm of TFPR_{si}, thus we must show that the variation in the logarithm of A_{si} does not depend on this industry constant.

$$ln(A_{si}) = ln(\kappa_s) + ln\left[\frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{\left[\alpha_{L,s}X_{si}^{\psi} + \alpha_{K,s}K_{si}^{\psi} + \alpha_{Q,s}Q_{si}^{\psi}\right]^{\frac{1}{\psi}}}\right] \tag{90}$$

$$\text{Var}[ln(A_{si})] = \text{Var}\left\{ ln\left[\frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{\left[\alpha_{L,s}X_{si}^{\psi} + \alpha_{K,s}K_{si}^{\psi} + \alpha_{Q,s}Q_{si}^{\psi}\right]^{\frac{1}{\psi}}}\right]\right\} \tag{91}$$

Thus, the industry level constant can be ignored, or equivalently set equal to one in my case i.e., $\kappa_s = 1$.

B.7 Inferring Distortions from the Data

Using the first-order conditions from the profit maximization problem of a given plant, we can infer the "wedges" from the data. Plugging the constraint in the problem outlined in equation (9), the problem simplifies to:

$$\begin{aligned}\Pi_{si} = \max_{L_{si}^c, L_{si}^f, K_{si}} & (1-\tau_{Y\,si})A_{si}^{\frac{\sigma-1}{\sigma}}\left[\alpha_{L,s}(\beta_{c,s}H_{c,si}^{\theta} + \beta_{f,s}H_{f,si}^{\theta})^{\frac{\psi}{\theta}} + \alpha_{K,s}K_{si}^{\psi} + \alpha_{Q,s}Q_{si}^{\psi}\right]^{\frac{\sigma-1}{\sigma\psi}} \\ & - \ w_{c,s}H_{c,si} - w_{f,s}(1+\tau_{f_{si}})H_{f,si} - r_s(1+\tau_{K_{si}})K_{si} - q_s(1+\tau_{Q_{si}})Q_{si} \end{aligned} \tag{92}$$

After some manipulation, the first-order conditions w.r.t $H_{c,si}$, $H_{f,si}$ and K_{si} write:

$$\alpha_{L,s}\beta_{c,s}\left(\frac{\sigma-1}{\sigma}\right)\frac{P_{si}Y_{si}}{\left[\alpha_{L,s}X_{si}^{\psi}+\alpha_{K,s}K_{si}^{\psi}+\alpha_{Q,s}Q_{si}^{\psi}\right]}X_{si}^{\psi-\theta}H_{c,si}^{\theta-1}=\frac{w_{c,s}}{(1-\tau_{Y_{si}})}=\mathrm{MRPH}_{c,si} \tag{93}$$

$$\alpha_{L,s}\beta_{f,s}\left(\frac{\sigma-1}{\sigma}\right)\frac{P_{si}Y_{si}}{\left[\alpha_{L,s}X_{si}^{\psi}+\alpha_{K,s}K_{si}^{\psi}+\alpha_{Q,s}Q_{si}^{\psi}\right]}X_{si}^{\psi-\theta}H_{f,si}^{\theta-1}=\frac{w_{f,s}(1+\tau_{f_{si}})}{(1-\tau_{Y_{si}})}=\mathrm{MRPH}_{f,si} \tag{94}$$

$$\alpha_{K,s}\left(\frac{\sigma-1}{\sigma}\right)\frac{P_{si}Y_{si}}{\left[\alpha_{L,s}X_{si}^{\psi}+\alpha_{K,s}K_{si}^{\psi}+\alpha_{Q,s}Q_{si}^{\psi}\right]}K_{si}^{\psi-1}=\frac{r_s(1+\tau_{K_{si}})}{(1-\tau_{Y_{si}})}=\mathrm{MRPK}_{si} \tag{95}$$

$$\alpha_{Q,s}\left(\frac{\sigma-1}{\sigma}\right)\frac{P_{si}Y_{si}}{\left[\alpha_{L,s}X_{si}^{\psi}+\alpha_{K,s}Q_{si}^{\psi}+\alpha_{Q,s}Q_{si}^{\psi}\right]}Q_{si}^{\psi-1}=\frac{q_s(1+\tau_{Q_{si}})}{(1-\tau_{Y_{si}})}=\mathrm{MRPQ}_{si} \tag{96}$$

I arrive at the formal labor, capital and intermediate input wedges by dividing equation (93) by equations (94), (95) and (96), respectively. The resulting expressions are given by:

$$(1+\tau_{f_{si}})=\left(\frac{\beta_{f,s}}{\beta_{c,s}}\right)\left(\frac{w_{c,s}}{w_{f,s}}\right)\left(\frac{H_{c,si}}{H_{f,si}}\right)^{1-\theta} \tag{97}$$

$$(1+\tau_{K_{si}})=\left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}}\right)\left(\frac{w_{c,s}}{r_s}\right)\left(\frac{H_{c,si}^{1-\theta}}{K_{si}^{1-\psi}}\right)X_{si}^{\theta-\psi} \tag{98}$$

$$(1+\tau_{Q_{si}})=\left(\frac{\alpha_{Q,s}}{\alpha_{L,s}\beta_{c,s}}\right)\left(\frac{w_{c,s}}{q_s}\right)\left(\frac{H_{c,si}^{1-\theta}}{Q_{si}^{1-\psi}}\right)X_{si}^{\theta-\psi} \tag{99}$$

The output distortion can be arrived at by simply rewriting equation (93), as follows:

$$(1-\tau_{Y_{si}})=\left(\frac{\sigma}{\sigma-1}\right)\frac{w_{c,s}H_{c,si}^{1-\theta}}{(\alpha_{L,s}\beta_{c,s})P_{si}Y_{si}}\frac{\left[\alpha_{L,s}X_{si}^{\psi}+\alpha_{K,s}Q_{si}^{\psi}+\alpha_{Q,s}Q_{si}^{\psi}\right]}{X_{si}^{\psi-\theta}} \tag{100}$$

B.8 Shutting down Individual Wedges

The relative importance of each distortion is found out by computing the reallocation gains by turning off each input specific wedge, one at a time. As noted in the text, this is equivalent to imposing an average industry wedge while maintaining the same industry demand for the given input, as observed from the data. This section outlines how this is done for the capital.

From equation (98), we can get an expression for the observed capital input, which writes:

$$K_{si} = \left[\left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}} \right) \left(\frac{w_{c,s} H_{c,si}^{1-\theta}}{r_s(1+\tau_{K_{si}})} \right) (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}} \tag{101}$$

Imposing an industry average wedge $\widetilde{\tau}_{K_s}$ for capital gives:

$$\widetilde{K}_{si} = \left[\left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}} \right) \left(\frac{w_{c,s} H_{c,si}^{1-\theta}}{r_s(1+\widetilde{\tau}_{K_s})} \right) (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}} \tag{102}$$

Making sure the aggregate industry demand for capital remains the same, such that we carry out a intra-industry reallocation, we can impose the following:

$$K_s = \sum_{i=1}^{N_s} \lambda_i \widetilde{K}_{si} = \sum_{i=1}^{N_s} \lambda_i \left[\left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}} \right) \left(\frac{w_{c,s} H_{c,si}^{1-\theta}}{r_s(1+\widetilde{\tau}_{K_s})} \right) (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}} \tag{103}$$

Allowing the summation operator to pass through the constants, we have:

$$K_s = \left[\left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}} \right) \left(\frac{w_{c,s}}{r_s(1+\widetilde{\tau}_{K_s})} \right) \right]^{\frac{1}{1-\psi}} \sum_{i=1}^{N_s} \lambda_i \left[H_{c,si}^{1-\theta} (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}} \tag{104}$$

Isolating $(1 + \widetilde{\tau}_{K_s})$ gives:

$$(1 + \widetilde{\tau}_{K_s}) = \left[\left(\frac{\alpha_{K,s}}{\alpha_{L,s}\beta_{c,s}} \right) \left(\frac{w_{c,s}}{r_s} \right) \right] \left\{ \frac{\sum_{i=1}^{N_s} \lambda_i \left[H_{c,si}^{1-\theta} (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}}}{K_s} \right\}^{1-\psi} \tag{105}$$

Plugging back this expression for the industry capital wedge in equation (102), we get the final expression for the new allocation of capital:

$$\widetilde{K}_{si} = \frac{\left[H_{c,si}^{1-\theta} (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}}}{\frac{\sum_{i=1}^{N_s} \lambda_i \left[H_{c,si}^{1-\theta} (\beta_{c,s} H_{c,si}^{\theta} + \beta_{f,s} H_{f,si}^{\theta})^{\frac{\theta-\psi}{\psi}} \right]^{\frac{1}{1-\psi}}}{K_s}} \tag{106}$$

We can subsequently recompute the output using this new allocation of capital. The derivation is virtually the same for intermediate inputs and is hence omitted, while the main text already reports this derivation for formal labor. With the new allocations of formal labor, capital and intermediate inputs and the implied output, the statistics in Table 7 use the following expressions for individual reallocation gains:

$$\frac{\widetilde{Y}^f}{Y} = \prod_{s=1}^{S}\left\{\frac{\widetilde{Y}_s^f}{Y_s}\right\}^{\mu_s} = \prod_{s=1}^{S}\left\{\left[\frac{\sum_{i=1}^{N_s}\lambda_i(\widetilde{Y_{si}}^f)^{\frac{\sigma-1}{\sigma}}}{\sum_{i=1}^{N_s}\lambda_i(Y_{si})^{\frac{\sigma-1}{\sigma}}}\right]^{\frac{\sigma}{\sigma-1}}\right\}^{\mu_s} \tag{107}$$

$$\frac{\widetilde{Y}^K}{Y} = \prod_{s=1}^{S}\left\{\frac{\widetilde{Y}_s^K}{Y_s}\right\}^{\mu_s} = \prod_{s=1}^{S}\left\{\left[\frac{\sum_{i=1}^{N_s}\lambda_i(\widetilde{Y_{si}}^K)^{\frac{\sigma-1}{\sigma}}}{\sum_{i=1}^{N_s}\lambda_i(Y_{si})^{\frac{\sigma-1}{\sigma}}}\right]^{\frac{\sigma}{\sigma-1}}\right\}^{\mu_s} \tag{108}$$

$$\frac{\widetilde{Y}^Q}{Y} = \prod_{s=1}^{S}\left\{\frac{\widetilde{Y}_s^Q}{Y_s}\right\}^{\mu_s} = \prod_{s=1}^{S}\left\{\left[\frac{\sum_{i=1}^{N_s}\lambda_i(\widetilde{Y_{si}}^Q)^{\frac{\sigma-1}{\sigma}}}{\sum_{i=1}^{N_s}\lambda_i(Y_{si})^{\frac{\sigma-1}{\sigma}}}\right]^{\frac{\sigma}{\sigma-1}}\right\}^{\mu_s} \tag{109}$$

B.9 Expression for Log Scaled TFPQ

The idea behind this expression is to gauge the ratio of TFPQ$_{si}$ ($=A_{si}$) of each plant $i \in s$, relative to the industry efficient TFPQ$_s^*$. Since we have probability sampling for certain plants, this ratio needs to be augmented with plant-level sample weights, denoted by λ_i. Let $W_s = \sum_{i=1}^{N_s}\lambda_i$ denote the sum of weights for all plants $i \in s$. The unweighted expression for log scaled TFPQ is simply the ratio of TFPQ$_{si}$ ($=A_{si}$) by TFPQ$_s^*$:

$$ln\left[\frac{A_{si}}{(\sum_{i=1}^{N_s}A_{si})^{\frac{1}{\sigma-1}}}\right] \tag{110}$$

We need to apply sample weights to the sum in the denominator, this gives:

$$ln\left[\frac{A_{si}(\sum_{i=1}^{N_s}\lambda_i)^{\frac{1}{\sigma-1}}}{(\sum_{i=1}^{N_s}\lambda_i A_{si})^{\frac{1}{\sigma-1}}}\right] = ln\left[\frac{A_{si}(W_s)^{\frac{1}{\sigma-1}}}{(\sum_{i=1}^{N_s}\lambda_i A_{si})^{\frac{1}{\sigma-1}}}\right] \tag{111}$$

This is the final expression used in the main text.

References

Amirapu, A. and Gechter, M. (2020). Labor Regulations and the Cost of Corruption: Evidence from the Indian Firm Size Distribution. *The Review of Economics and Statistics*, 102:1–48.

Banerjee, A. and Duflo, E. (2005). Growth Theory through the Lens of Development Economics. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 1, Part A, chapter 07, pages 473–552. Elsevier, 1 edition.

Bertrand, M., Hsieh, C.-T., and Tsivanidis, N. (2021). Contract Labor and Firm Growth in India. Working Paper 29151, National Bureau of Economic Research.

Boehm, J. and Oberfield, E. (2020). Misallocation in the Market for Inputs: Enforcement and the Organization of Production. *The Quarterly Journal of Economics*, 135(4):2007–2058.

Chatterjee, U. (2011). Resource Allocation and Efficiency in Developing Countries. Phd thesis, University of California Berkeley.

Chen, K. and Irarrazabal, A. (2013). Misallocation and the recovery of manufacturing TFP after a financial crisis. Working Paper 2013/01, Norges Bank.

Cobbold, T. S. (2003). A Comparison of Gross Output and Value-Added Methods of Productivity Estimation.

Debroy, B. and Bhandari, L. (2008). India Labour Report. Teamlease Services.

Dias, D., Marques, C., and Richmond, C. (2016). Misallocation and productivity in the lead up to the Eurozone crisis. *Journal of Macroeconomics*, 49.

Dias, D. A., Marques, C. R., and Richmond, C. (2018). A Tale of Two Sectors : Why is Misallocation Higher in Services than in Manufacturing? International finance discussion papers, Board of Governors of the Federal Reserve System (U.S.).

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3):297–308.

Economic Survey of India (2019-20). Government of India, Ministry of Finance, Department of Economic Affairs. Technical report.

Farrell, D. (2004). The Hidden Dangers of the Informal Economy. *McKinsey Quarterly.*

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425.

Gollop, F. and Roberts, M. (1979). Firm interdependence in oligopolistic markets. *Journal of Econometrics*, 10(3):313–331.

Gustavo, M. C. and Cristobal, B. J. (2012). Misallocation and Manufacturing TFP in Bolivia during the Market Liberalization Period. *The B.E. Journal of Macroeconomics*, 12(1):1–45.

Hicks, J. (1981). *Wealth and Welfare: Collected Essays on Economic Theory*, volume 1. Cambridge: Harvard University Press.

Hoff, A. (2004). The Linear Approximation of the CES Function with n Input Variables. *Marine Resource Economics*, 19(3):295–306.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hulten, C. R. (1991). The Measurement of Capital. In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, pages 119–158. National Bureau of Economic Research, Inc.

India Productivity Report (2022). Centre for Development Economics, Delhi School of Economics, University of Delhi. Technical report.

Jorgenson, D. and Griliches, Z. (1967). The Explanation of Productivity Change. *Review of Economic Studies*, 34(3):249–283.

Jorgenson, D. W., Gollop, F., and Fraumeni, B. (1987). *Productivity and U.S. Economic Growth*. Harvard University Press.

Jorgenson, D. W. and Stiroh, K. J. (2001). Raising the Speed Limit: US Economic Growth in the Information Age. OECD Economics Department Working Papers 261, OECD Publishing.

Kabiraj, S. (2020). Analyzing the Sources of Misallocation in Indian Manufacturing: A Gross-Output Approach. *Asian Development Review*, 37(2):134–166.

Kmenta, J. (1967). On Estimation of the CES Production Function. *International Economic Review*, 8(2):180–189.

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*. Washington, DC: Brookings Institution Press.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725.

Misch, F. and Saborowski, C. (2018). Resource Misallocation and Productivity: Evidence from Mexico. *IMF Working Papers*, 18:1.

Mohommad, A., Sandoz, C., and Sodsriwiboon, P. (2021). Resource Misallocation in India: The Role of Cross-State Labor Market Reform. *IMF Working Papers*, 21.

Oulton, N. and O'Mahony, M. (1994). *Productivity and Growth*. Cambridge University Press.

Padmakumar, K. (2022). Small by Choice? Reassessing the Aggregate Implications of Size-Based Regulations. Phd thesis (job market paper), The Pennsylvania State University.

Peters, M. (2020). Heterogeneous Markups, Growth, and Endogenous Misallocation. *Econometrica*, 88(5):2037–2073.

Sato, K. (1967). A Two-Level Constant-Elasticity-of-Substitution Production Function. *Review of Economic Studies*, 34(2):201–218.

Sen, K. (2020). The 'Missing Middle' Problem in Indian Manufacturing. *Economic and Political Weekly*, 55(16).

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, 70(1):65–94.

Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8):2015–47.