The Hidden Cost of Misallocation: Low Markups, Low Welfare

Sandeep Achary* Nicolás Oviedo*

March 27, 2026

Latest version: Click here

Abstract

A large literature has shown that resource allocation in emerging economies is shaped by severe policy distortions, while a separate literature on market power has emphasized the role of markups for welfare, concentration, and labor shares, often treating aggregate markups as a useful summary statistic for the welfare cost of imperfect competition. This paper brings these two conversations together. We ask: when firms operate in distorted environments of the kind common in developing economies, what exactly do aggregate markups tell us about welfare? We show that, in the presence of distortions, aggregate markups are no longer a sufficient statistic for welfare. In particular, the low aggregate markups often estimated in emerging markets can be a symptom of severe misallocation.Our main result is that aggregate markups and welfare follow an inverted-U relationship in the presence of distortions. The mechanism operates through two channels. First, wedges change firm scale and therefore market concentration. Second, even when wedges are ex ante independent of productivity, survival is endogenous, so distortions and productivity become correlated among active firms. Large subsidy-like distortions keep many marginal, low-productivity firms alive, generating low aggregate markups together with low productivity and low welfare. Large tax-like distortions instead produce more concentrated markets with better selection, but also high aggregate markups and low welfare. In ongoing work, we extend the production-function approach to markup estimation to environments with distortions, with the aim of taking the model's decomposition to the data.

JEL Codes: D24, E23, D61, D43, L13, I31.

*Both authors are PhD students at the Department of Economics and Business of Universitat Pompeu Fabra. We thank Isaac Baley, Andrea Caggese, Jan Eeckhout, Marcela Eslava, Madalena Gaspar, Martí Mestieri, Dávid Nagy, Hugo Reichardt, Victoria Vanasco; and seminar participants at UPF and UAB.
Sandeep Achary: sandeep.pothedar@upf.edu
Nicolás Oviedo: nicolas.oviedo@upf.edu

I Introduction

The rise in aggregate markups and the fall in the labor share of income have attracted substantial attention in recent years. Much of this research, however, has centered on advanced market economies (Autor et al., 2020; De Loecker et al., 2020). In contrast, relatively little is known about the behavior and welfare implications of markups in emerging markets (Díez et al., 2018). There are at least two reasons to tailor the analysis to developing economies. First, both the level and evolution of aggregate markups differ from those observed in the United States and Europe (Díez et al., 2021). Second, emerging economies are often characterized by distortions that misallocate resources across firms (Banerjee and Moll, 2010; Eslava et al., 2024). This paper develops a framework to study how such distortions shape markups at the firm and aggregate levels, and how they alter the welfare interpretation of market power in economies with distortions.

The mechanism is easiest to understand by first considering a benchmark economy without misallocation frictions. A fixed resource is allocated among firms with heterogeneous productivity, each producing a differentiated variety and facing demand with variable price elasticity.[1] In this setup, the market efficiently allocates more resources to the most productive firms. As a result, these firms operate on a larger scale, set lower prices, and — because they face less elastic demand — charge higher markups than their less productive counterparts. With free entry, the only source of welfare loss stems from firms charging markups rather than passing the full benefits of higher productivity on to consumers.[2] Aggregate markups therefore have a clear welfare interpretation in this benchmark: higher aggregate markups are associated with lower welfare.

In the presence of distortions, however, this relationship becomes more complex. Consider an otherwise similar economy in which firms face revenue wedges. Resources are no longer allocated efficiently: the most productive firms no longer necessarily employ the largest share of resources, and the markup charged by a given firm depends not only on its productivity, but also on how distortions affect its scale and the market environment in which it operates. Welfare losses now stem from both firms charging markups and the inefficient allocation of resources. Aggregate markups therefore capture only one dimension of welfare loss. We show that aggregate markups are not necessarily higher in the presence of misallocation, nor do they always decline when misallocation is reduced. Their relationship with welfare depends on the magnitude and distribution of distortions

[1] Such a demand system arises under additively separable preferences, such as the translog expenditure function or the quadratic utility function used by Melitz and Ottaviano (2008).

[2] The benchmark economy is second-best in the spirit of [Dhingra and Morrow, 2019].

across firms.

To study this question, we build a tractable general-equilibrium model of heterogeneous firms with endogenous markups, free entry, and revenue wedges in the spirit of Hsieh and Klenow (2009), embedded in a variable-markup environment à la Melitz and Ottaviano (2008). The model combines two ingredients. First, firms differ in productivity and face variable residual demand, so markups are endogenous rather than constant. Second, firms face revenue wedges that are ex ante unrelated to fundamentals but distort scale and profitability. This setup allows us to study how pre-existing allocative frictions shape both the cross-sectional distribution of markups and the aggregate relationship between markups, productivity, and welfare. A useful way to summarize the mechanism is that firm-level markups reflect both distortions and residual demand: wedges affect effective costs and scale, while market power affects the elasticity firms face.

Our main theoretical result is that aggregate markups and welfare follow an inverted-U relationship in the presence of distortions. The mechanism operates through two channels. The first is a concentration channel: wedges shift effective costs and firm scale, changing firms' price–quantity choices and therefore their markups. The second is a selection channel: even when wedges are ex ante independent of productivity, survival is endogenous, so distortions and productivity become correlated among active firms. This endogenous wedge–productivity correlation is central, because the economy we observe is not the ex ante economy, but the selected set of firms that survive and produce.

These two channels imply that both low and high aggregate markups can be associated with low welfare. Subsidy-like distortions keep many marginal, low-productivity firms alive. These firms charge low markups but absorb resources inefficiently, generating low aggregate markups together with low productivity and low welfare. Tax-like distortions instead push the economy toward a more concentrated market structure with better selection, but surviving firms charge higher markups; the result is high aggregate markups together with low welfare. In this environment, policy improvements can raise welfare, productivity, and aggregate markups at the same time. The broader implication is that the welfare content of aggregate markups is fundamentally altered by misallocation: once distortions reshape composition and concentration, aggregate markups can no longer be read in the usual way.

The paper also develops an empirical strategy for taking the model to the data. In the simple benchmark studied here, the market-power and distortion channels are difficult to disentangle. In ongoing work, and as we outline in the empirical strategy section, we

extend the production-function approach to markup estimation to environments with distortions. The key idea is to move to a richer quantitative setting — for example, with a Kimball or Atkeson–Burstein demand system — in which the market-power and distortion components of firm-level markups can be separately identified and estimated. Under the assumption that distortions affect firms' scale but not their input mix, the production approach can still recover observed markups, allowing us to back out the market-power component separately. This, in turn, provides a path to quantifying aggregate efficiency losses through a decomposition into a markup-dispersion term, a distortion-dispersion term, and a covariance term between the two. Our theory suggests that this covariance term is especially important in distorted environments, since endogenous selection generates correlation between wedges and firm fundamentals even when wedges are ex ante unrelated to productivity.

Our paper contributes to two broad strands of literature: market power and misallocation. First, we contribute to the literature on market power by showing that, in distorted economies, aggregate markups need not move monotonically with welfare.[3] This qualifies a common interpretation in which higher aggregate markups are read as unambiguously signaling lower welfare. Our mechanism shows instead that policy improvements can raise welfare, productivity, and aggregate markups at the same time. Our paper also speaks to the literature on the measurement and evolution of markups at the firm level and in the aggregate [Hall, 1988; Roeger, 1995; De Loecker and Warzynski, 2012; De Loecker, Eeckhout and Unger, 2020] by extending the framework to economies with distortions.

Second, we contribute to the misallocation literature by endogenizing the covariance between wedges and firm fundamentals through selection. Much of the empirical and quantitative literature takes wedge–productivity covariance as an observed object; our model shows how such covariance can arise endogenously even when wedges are ex ante orthogonal to productivity. In this sense, we build directly on the literature that studies resource misallocation as dispersion in marginal products induced by distortions, and its consequences for aggregate productivity [Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Restuccia and Rogerson, 2017]. Our framework also connects this tradition to work on allocational efficiency under imperfect competition. In particular, with heterogeneous firms and variable-elasticity demand, markets may select the wrong distribution of firms and allocate the wrong quantities across them [Dhingra and Morrow,

[3]Relative to work that studies the allocative consequences of markup dispersion itself, our focus is different: we ask how orthogonal misallocation frictions change the mapping from aggregate markups to aggregate outcomes [Peters, 2020; Edmond, Midrigan and Xu, 2023; Weinberger, 2020; Baqaee and Farhi, 2020].

2019]. We take this variable-markup benchmark to economies with Hsieh–Klenow-style wedges and show that wedges interact with endogenous market power and endogenous survival, thereby altering the relationship between aggregate markups, productivity, and welfare.

The rest of the article is organized as follows. Section II presents the model. Section III presents the main theoretical results. Section IV discusses the empirical strategy and measurement agenda. Section V concludes.

II Model

II.A Consumption

The economy is populated by L consumers, each supplying a single unit of labor inelastically. Consumers have identical, quasilinear preferences over a continuum of differentiated varieties (q_i) and a homogeneous good (q_0), which serves as the numeraire. This preference structure ensures that the consumption decisions for the differentiated goods are not affected by income levels, simplifying the analysis. The utility function is given by:

$$U = q_0 + \alpha \int_{i\in\Omega} q_i di - \frac{1}{2}\gamma \int_{i\in\Omega} (q_i)^2 di - \frac{1}{2}\eta \left(\int_{i\in\Omega} q_i di \right)^2 \tag{1}$$

The parameters α and η shape the substitution patterns between the differentiated goods and the numeraire. Concretely, the relative demand for varieties versus the numeraire is increasing in α and decreasing in η. The parameter γ reflects the extent of love for varieties. In the extreme case where $\gamma = 0$, varieties are perfect substitutes as consumers only care about the total quantity consumed. These demand parameters (α, η, γ) are all positive.

By solving the consumer's optimization problem, we obtain the total demand for a variety i (q_i), which is the sum of individual demands for that variety. Total demand is a function of its own price p_i and an endogenous price bound (p_{max}), and is given by:

$$q_i = \frac{L}{\gamma} [p_{max} - p_i] \tag{2}$$

The price bound p_{max} is endogenously determined and reflects market conditions, such as the number of competing varieties (N), their average price ($\bar{p}$), and the demand parameters. This price bound can be thought of as a reservation or choke price, since it is the price at which demand for a variety is driven to zero. Formally, p_{max} is defined as:

$$p_{max} = \frac{\alpha\gamma + \eta N\bar{p}}{\eta N + \gamma} \tag{3}$$

The linear demand for each variety implies a variable price elasticity of demand:

$$\epsilon_i = \left(\frac{p_{max}}{p_i} - 1\right)^{-1} \tag{4}$$

Unlike the C.E.S case, the price elasticity of demand for each variety depends on its own price p_i and the price bound p_{max}. Given these factors, lower average prices ($\bar{p}$) or a larger number of competing varieties (N) induce a decrease in the price bound p_{max} and an increase in the price elasticity of demand ϵ_i at any given p_i. This leads to a "tougher" competitive environment, as proposed in Melitz and Ottaviano (2008). Thus, the price bound p_{max} serves to summarize the state of the market structure. The quasilinear preference structure and monopolistic competition for varieties are the source of the variable price elasticity of differentiated goods. This, in turn, generates endogenous and heterogeneous markups, a key feature of this environment. To demonstrate this, we now turn our attention to the production side of the economy.

II.B Production

Every good in this economy is produced using a single input: labor. The labor market is competitive, and labor can move freely across the production lines of different goods.

II.B.1 Homogeneous Good Producers

The producers of the homogeneous good use a unit labor technology and operate in a perfectly competitive, frictionless market. We normalize the price of the homogeneous good to 1. Since there are no labor market frictions and the product market is competitive,

the unit labor technology also implies a unit wage.

II.B.2 Differentiated Good Producers

Firms must incur upfront R&D expenses - an entry cost (f_e)- to develop a specific product variety. Research and development yields an uncertain outcome, allowing a firm to produce at a constant marginal cost of c[4]. Additionally, firms in the differentiated sector face misallocation frictions—modeled as a firm-specific revenue wedge (τ_i)—during the production process. Before incurring the entry costs, firms face uncertainty about two crucial factors: their productivity level ($1/c_i$) and the revenue wedge they will face (τ_i). Only those firms that find it profitable will pay the entry fee of f_e and proceed to enter the market.

Following the existing literature, we model misallocation frictions as firm-specific revenue wedges, in the spirit of Hsieh and Klenow (2009). These frictions can arise from various channels and are purely distortionary (Restuccia and Rogerson, 2017)[5]. We use taxes and subsidies as a running narrative for these misallocation wedges, as this simplifies the exposition and provides a clear path for thinking about the welfare costs of misallocation. The profit maximization problem for a firm that has entered the market is:

$$\begin{aligned} &\max_{p_i,q_i}(1-\tau_i)p_iq_i - c_iq_i \\ &\text{s.t: } q_i = \frac{L}{\gamma}\left[p_{max} - p_i\right] \end{aligned} \tag{5}$$

where c_i is the marginal cost of the firm and $\tau_i \in (-1, 1)$ is the tax (if $\tau_i > 0$) or subsidy (if $\tau_i < 0$) rate it faces.

The solution to the firm's profit maximization problem[6] is given by equations (6)-(8):

[4]The production function is $q = l$ and the total cost function is $c(q) = c \times w \times q$. Since the wage is equal to 1, the total cost function simplifies to $c(q) = c \times q$.

[5]Some examples are adjustment costs, financial frictions, political connections, access to transportation infrastructure, and size-dependent taxes or subsidies.

[6]See Appendix **??**.

$$p_i = \frac{1}{2}\left[p_{max} + \frac{c_i}{1-\tau_i}\right] \tag{6}$$

$$q_i = \frac{L}{2\gamma}\left[p_{max} - \frac{c_i}{1-\tau_i}\right] \tag{7}$$

$$\mu_i \equiv \frac{p_i}{c_i} = \frac{1}{2}\left[\frac{p_{max}}{c_i} + \frac{1}{1-\tau_i}\right] \tag{8}$$

In the presence of misallocation frictions, the firm's scale q_i, its price p_i, and its markup μ_i depend not only on its productivity $1/c$ but also on the tax rate it faces. To the extent that these tax rates are heterogeneous across firms, the fiscal policy in place breeds cross-sectional misallocation of resources[7]. Consequently, a firm's scale, its product price, and the markup it charges are not solely related to productivity—a deviation from the constrained efficient benchmark. Firms burdened with a tax ($\tau_i > 0$) charge a higher price, operate at a smaller scale, and set an even higher markup relative to a frictionless benchmark. Similarly, firms receiving a subsidy ($\tau_i < 0$) charge a smaller price, operate at a larger scale, and set a smaller markup.

It is important to note that our discussion up to this point has been a partial equilibrium assessment, as we have ignored the impact of misallocation on the price bound (p_{max}). To gain a comprehensive understanding, we need to explore how misallocation frictions influence firms' decisions to enter or exit the market, which in turn affects p_{max} and the overall market structure.

II.B.3 Free Entry Equilibrium

A potential entrant faces two sources of uncertainty when deciding whether to enter the market: its productivity level ($1/c_i$) and the firm-specific revenue wedge (τ_i). We assume that the firm is aware of the distribution of costs $G(c)$ and the distribution of revenue

[7]Recent literature on misallocation documents considerable heterogeneity in misallocation frictions. Eslava et al. (2024) documents that firms in Colombia face correlated wedges. To be precise, they find that less productive firms face higher negative wedges and more productive firms face higher positive wedges. On the tax front, Kaymak and Schott (2024) report substantial heterogeneity in effective corporate income tax rates faced by firms in the US.

wedges $\mathbb{G}(\tau)$ when making the entry decision, but their exact realization is only known after paying the fixed cost (f_e).

We introduce simplifying distributional assumptions to gain analytical tractability and derive closed-form solutions. The model thus yields simple solutions, which allow us to better underpin the forces behind our results[8].

ASSUMPTION A: We assume that a firm can potentially face two levels of wedges, so that $\mathbb{G}(\tau)$ is a Bernoulli distribution:

$$\tau_i = \begin{cases} \tau_l & \mathbb{P} \\ \tau_h & 1-\mathbb{P} \end{cases}$$

with $-1 < \tau_l \leq \tau_h < 1$. Draws from $\mathbb{G}(\tau)$ are independent across firms.

ASSUMPTION B: We assume that productivity draws are independent and follow a Pareto distribution with a lower bound of $1/c_m$ and a shape parameter $k \geq 1$. This implies that cost draws come from a distribution given by

$$G_c(c) = \left(\frac{c}{c_m}\right)^k, c \leq c_m$$

It is useful to define a wedge-adjusted effective cost measure (z). Casting everything in terms of z allows us to simplify the free entry condition and facilitates comparisons between different benchmarks.

DEFINITION 1: We define the wedge-adjusted cost of a firm with cost c_i and facing a wedge τ_i as z_i:

$$z_i = \frac{c_i}{1-\tau_i}$$

The joint distribution of z[9] is given by the following lemma:

LEMMA 1: Under assumptions (A) and (B), the distribution of wedge-adjusted cost is given by

$$G_z(z) = \mathbb{P}\left(\frac{z(1-\tau_l)}{c_m}\right)^k + (1-\mathbb{P})\left(\frac{z(1-\tau_h)}{c_m}\right)^k \tag{9}$$

[8]It is possible to solve for more general cases, albeit by reneging on tractability. See Autor et al. (2020) online appendix for more on this.

[9]Note that in our formulation each z_i corresponds to two levels of cost. One of which is $z_i(1-\tau_l)$ and the other $z_i(1-\tau_h)$.

DEFINITION 2: We define the wedge-adjusted cost of a firm that is indifferent to producing as z_D.

Firms with a combination of cost draw and tax rate (c_i, τ_i) such that $z_i = z_D$ will charge a price $p_i = p_{max}$ and are indifferent to producing[10]. We formalize this in the following lemma:

LEMMA 2: Define z_D as the wedge-adjusted cost of the firm that is indifferent to producing. Then

$$p_{max} = z_D \tag{10}$$

and solving for free-entry determines z_D.

Prior to entry, the expected firm profit is $\int_0^{z_D} \pi(z, z_D) dG_z(z) - f_E$. As long as some firms produce, the expected profit is driven to zero by the unrestricted entry of new firms. This yields the equilibrium free entry condition:

$$\int_0^{z_D} \pi(z, z_D) dG_z(z) = f_e \tag{11}$$

where $\pi(z, z_D)$ are the profits of a firm that faces a wedge-adjusted cost z. The free-entry condition (11) determines z_D, where only firms with $z_i < z_D$ will produce. Therefore, z_D is a market-level, wedge-adjusted cutoff that determines which firms survive and operate in the market. Under assumptions (A) and (B), we can solve for z_D in closed form, which yields:

$$z_D = c_D \times \left[\frac{1}{\mathbb{P}(1-\tau_l)^{k+1} + (1-\mathbb{P})(1-\tau_h)^{k+1}} \right]^{1/(k+2)} \tag{12}$$

Where c_D is the market cutoff in the absence of misallocation frictions ($\tau_l = \tau_h =$

[10]To see this we have to rewrite the profit maximization outcomes of a firm in terms of z_i. They are as following:

$$p(z_i) = \frac{1}{2}\left[z_i + p_{max}\right], \quad q(z_i) = \frac{L}{2\gamma}\left[p_{max} - z_i\right], \quad \mu(z_i, c_i) = \frac{1}{2}\left[\frac{z_i + p_{max}}{c_i}\right]$$

0):

$$c_D = \left[\frac{2\gamma c_M^k (k+1)(k+2) f_e}{L}\right]^{1/(k+2)}$$

Notice that c_D depends on market fundamentals $\{\gamma, k, c_M, L, f_e\}$ and is the same expression obtained by Melitz and Ottaviano (2008). Consequently, their benchmark model corresponds to the special case in our framework where there are no revenue wedges at all. Under the assumptions made, z_D is separable in misallocation and market fundamentals. These assumptions also allow us to clearly see the effect of market fundamentals and misallocation on the market structure.

II.B.4 Distribution of Surviving Firms

Since it operates as a cutoff, z_D determines the market structure. Once z_D is known, we can determine the number of surviving firms and their distribution. Using (10) and (3), we have:

$$z_D = \frac{\alpha\gamma + \eta N \bar{p}}{\eta N + \gamma}$$

This yields the number of surviving firms:

$$N = \frac{\gamma}{\eta}\frac{\alpha - z_D}{z_D - \bar{p}} \tag{13}$$

Despite making simplifying assumptions on how wedges are distributed, the distribution of surviving firms, their sizes and the prices they charge all follow a power-law with the same shape parameter as cost draws. The distribution of surviving firms can be derived using assumption B and Lemma 1, and is formalized in the following lemma:

LEMMA 3: The distribution of surviving firms follows a power law in wedge-adjusted cost.

$$\tilde{G}_z(z) = \left(\frac{z}{z_D}\right)^k, z \leq z_D \tag{14}$$

Finally, we can rewrite the firm decisions in terms of z_D and z:

$$p(z_i) = \frac{1}{2}\left[z_i + z_D\right] \tag{6'}$$

$$q(z_i) = \frac{L}{2\gamma}\left[z_D - z_i\right] \tag{7'}$$

$$\mu(z_i, c_i) = \frac{1}{2}\left[\frac{z_i + z_D}{c_i}\right] \tag{8'}$$

Thus, once we determine z_D, we have the decisions of the surviving firms (6′)-(8′), the number of surviving firms (13), and their distribution (14). We now have all the necessary components to analyze the role of misallocation in determining aggregate markups and its welfare consequences.

III Theoretical Results

We now analyze the impact of misallocation on firm decisions. The frictions we have introduced have two key effects: **partial equilibrium effects**, which are caused by cross-sectional resource misallocation, and **general equilibrium effects**, which result from inefficient entry and exit of firms.

III.A Misallocation in Partial Equilibrium

Equations (6′)- (8′) capture the partial equilibrium effects of misallocation—that is, the impact of the firm-specific revenue wedge (τ) on firm-level metrics while holding the market cutoff z_D constant. To understand these effects, consider the following lemma:

Lemma 4: Each level of wedge-adjusted cost, z_i, corresponds to two different levels of a firm's fundamental cost, c, and the probability of a firm having a high cost while receiving a lower wedge is ϵ. That is:

$$c_i = \begin{cases} (1-\tau_l)z_i & \epsilon \\ (1-\tau_h)z_i & 1-\epsilon \end{cases}$$

Note that $\epsilon \neq \mathbb{P}$, where $\mathbb{P}$ is the ex-ante probability of receiving the lower wedge. A useful way to think about ϵ is as a survival probability—it is the probability of a firm having received a low wedge conditional on successfully operating in the market. Our model's assumptions generate correlated wedges, meaning that for a given wedge-adjusted cost

(z), less productive firms that receive a lower wedge are more likely to survive.

As Lemma 4 points out, in the presence of misallocation, a single level of wedge-adjusted cost can correspond to two firms with different fundamental productivity levels. These two firms, despite their different productivities, will charge the same price and have the same scale of operations. This implies that the less productive firm is employing an inefficiently large amount of resources. Of these two firms, the more productive one (the one with lower fundamental cost) charges a higher markup.

How these partial equilibrium effects translate into aggregate markups is not straightforward. With two firms charging the same price and holding the same market share, the higher markup charged by the more productive firm may raise aggregate markups, while the lower markup of the less productive firm may reduce them. Therefore, the aggregate effect depends on the specific values and distribution of the wedges (τ_l and τ_h) and the survival probability (ϵ).

III.B Misallocation in General Equilibrium

Since firms account for the possibility of facing misallocation frictions when making their entry decisions, the presence of wedges alters the market structure, now fraught with inefficient entry and exit. This inefficiency can be summarized by the average level of misallocation, which is reflected in the market cutoff, z_D. The following theorem is central to our analysis:

THEOREM 1: There is a cutoff probability $\phi^*(\tau_l, \tau_h)$, such that whenever $\mathbb{P} \leq \phi^*(\tau_l, \tau_h)$, we have $z_D \geq c_D$, where

$$\phi^*(\tau_l, \tau_h) = \frac{1-(1-\tau_h)^{k+1}}{(1-\tau_l)^{k+1}-(1-\tau_h)^{k+1}} \approx \frac{\tau_h}{\tau_h - \tau_l}$$

.

We present two key corollaries that summarize the general equilibrium effects of misallocation:

COROLLARY 1.1: If the expected wedge is negative ($\mathbb{P}\tau_l + (1-\mathbb{P})\tau_h < 0$), then $z_D < c_D$. This results in an inefficiently large number of active firms relative to the frictionless benchmark ($N_\tau > N_0$).

COROLLARY 1.2: If the expected wedge is non-negative ($\mathbb{P}\tau_l + (1-\mathbb{P})\tau_h \geq 0$), then $z_D \geq c_D$. This results in an inefficiently small number of active firms in comparison to the

benchmark without any misallocation frictions ($N_\tau \leq N_0$).

The reason for the inefficient number of firms is as follows. In the first case, some firms that would not have entered in a frictionless environment now do so because there is a high chance of receiving a subsidy or a lower tax. In the second case, some firms that would have entered are discouraged from doing so because the probability of being taxed is high, or the tax level itself is very high.

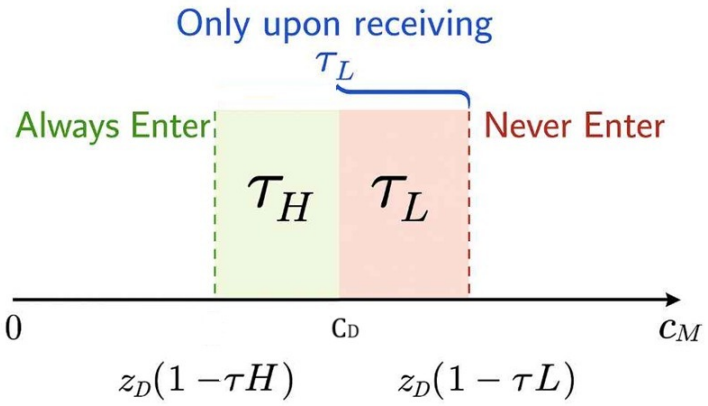


Figure 1: Schematic illustration of inefficiencies in the entry margin.

As the Figure 1 illustrates, misallocation leads to the inefficient entry of less productive firms (those to the right of c_D indicated in red) and the exit of otherwise productive firms (those to the left of c_D who exit upon receiving a higher tax).

III.C Market Aggregates and Welfare

In this simple model, we can express market aggregates in terms of the model's parameters and the endogenous objects (z_D, ϵ).

III.C.1 Aggregate Markups

The aggregate revenue-weighted markups can be computed using the distribution from Lemma 3 and are given by:

$$\bar{\mu_R} = \frac{1}{2N_\tau}\left[\frac{\epsilon}{1-\tau_l} + \frac{1-\epsilon}{1-\tau_h}\right]\left[\frac{2k^2+2k-1}{(k-1)(k+1)}\right]$$

As can be seen, the effect of misallocation on aggregate markups is mediated by N_τ, τ_h, τ_l, and ϵ.

III.C.2 Welfare

Throughout the model, we have used taxes and subsidies as a narrative for misallocation frictions, even though they do not have to be. We use the same narrative to evaluate the welfare costs of misallocation by assuming a government in the background that rebates the tax revenues back to consumers in a lump-sum fashion[11]. This approach is consistent with the literature[12].

Welfare in this economy can be expressed as:

$$U = L + \frac{L}{2}(\eta + \gamma/N_\tau)^{-1}(\alpha - \bar{p}(\tau))^2 + \frac{L}{2}\frac{N_\tau}{\gamma}\sigma_p^2(\tau) + N_\tau T$$

where $T = [(1-\epsilon)\tau_h + \epsilon\tau_l]\,\frac{L}{2\gamma}\frac{1}{k+2}z_D^2$ represents the total tax revenue. The last term ($N_\tau T$) captures the direct welfare losses due to misallocation, while the rest of the expression captures the welfare effects of having a lower average price and a larger number of varieties. This expression pinpoints that considering welfare without including the misallocation costs can be very misleading. For example, a situation where everyone receives a subsidy will lead to lower average prices and a higher number of firms, which might appear as higher welfare. However, when the true cost of such a program is considered, the overall effect on welfare can be reversed.

III.D Comparative Statics

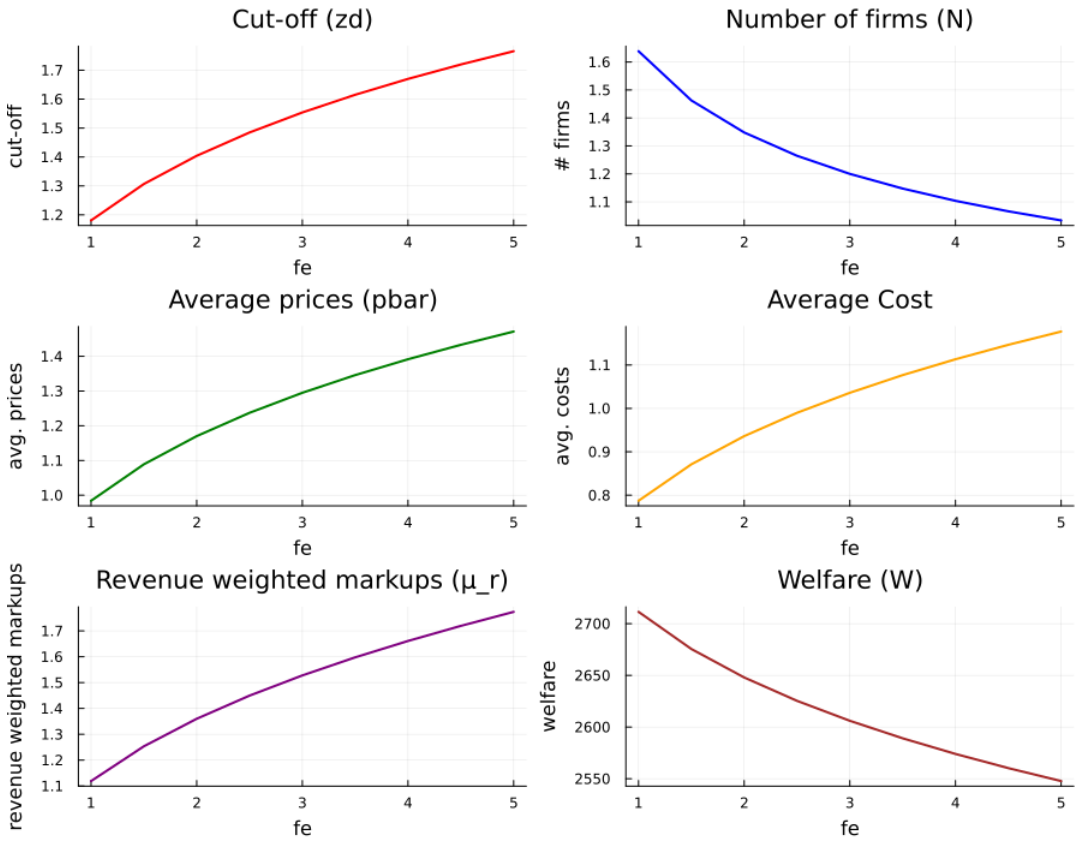
Even though our model provides closed-form solutions and is analytically tractable, we provide results from a comparative static exercise to illustrate the framework's inner workings.

III.D.1 No Misallocation and the Role of Market Fundamentals

First, we establish a benchmark with no misallocation by setting $\tau_l = \tau_h = 0$ to examine whether our model can replicate the role of fixed costs as highlighted by the literature (De Loecker et al. (2020)).

[11] The way these taxes are levied are indeed random, giving them more a flavor of market frictions than a fiscal policy.

[12] See for example Liu (2019).



As the figure shows, an increase in fixed costs leads to an increase in the cutoff z_D (which equals c_D in the benchmark case), a decrease in the number of active firms (N), an increase in aggregate markups and average prices, and a decline in welfare. Thus, our model aligns with the current literature on markups and market power.

III.D.2 Misallocation and Aggregate Markups

In the presence of misallocation, welfare is always lower than in the benchmark, but the level of aggregate markups depends on the nature of the misallocation. We present two cases to illustrate this point.

Low Markups and Low Welfare. We maintain the benchmark's parametrization but introduce misallocation frictions where firms face potential subsidies or lower taxes. Specifically, we set $\tau_h = 0$, vary the size of τ_l, and compute aggregate metrics for such different configurations of our laboratory economy. Figure 2 reflects the results.

At first glance, it is clear that the inverse relationship between aggregate markups and welfare is muted in the presence of misallocation. This is because subsidies favor the entry of less productive firms, which take resources away from productive units and

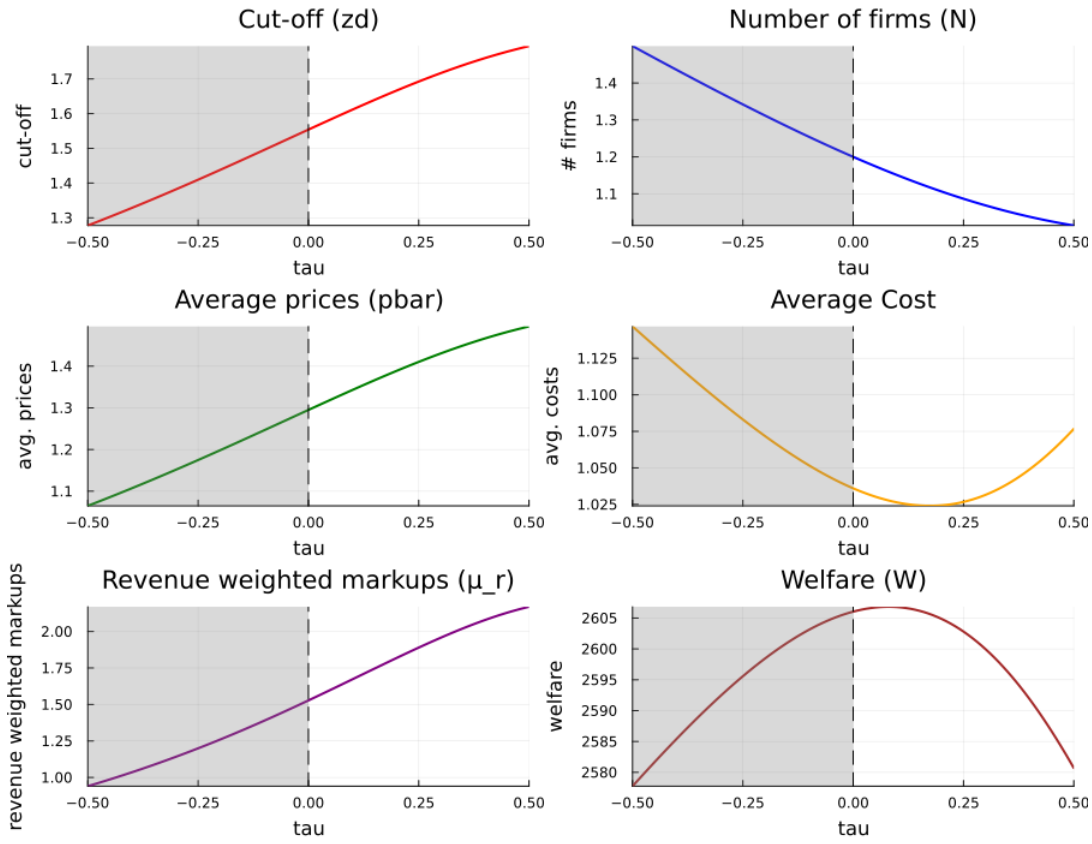


Figure 2: Comparative statics for τ_l fixing $\tau_h = 0$ in the gray shaded area. The dash line demarcates the frictionless benchmark ($\tau_l = \tau_h = 0$).

charge lower markups. As the probability of receiving a subsidy increases, z_D falls (in line with Corollary 1.1), more firms enter the market, and both average prices and aggregate markups decline. However, the welfare costs of this misallocation friction are so significant that they offset the positive effects of lower average prices and a larger number of product varieties. In this situation, a policy that reduces misallocation can simultaneously increase both aggregate markups and welfare.

High Markups and Low Welfare. We now present a different case in which the presence of misallocation can lead to high markups and low welfare. We assume that $\tau_l = 0$, $\mathbb{P} = 0.5$, and increase the value of τ_h.

As Figure 3 shows, the outcome is very different from the previous case. As τ_h increases (holding $\mathbb{P}$ constant), z_D increases (in line with Corollary 1.2), more firms exit the market, and both average prices and aggregate markups rise. This is because these frictions lead to inefficient exits, concentrating resources in the hands of more productive firms. While these firms are efficient, a lack of competition allows them to not fully pass these benefits

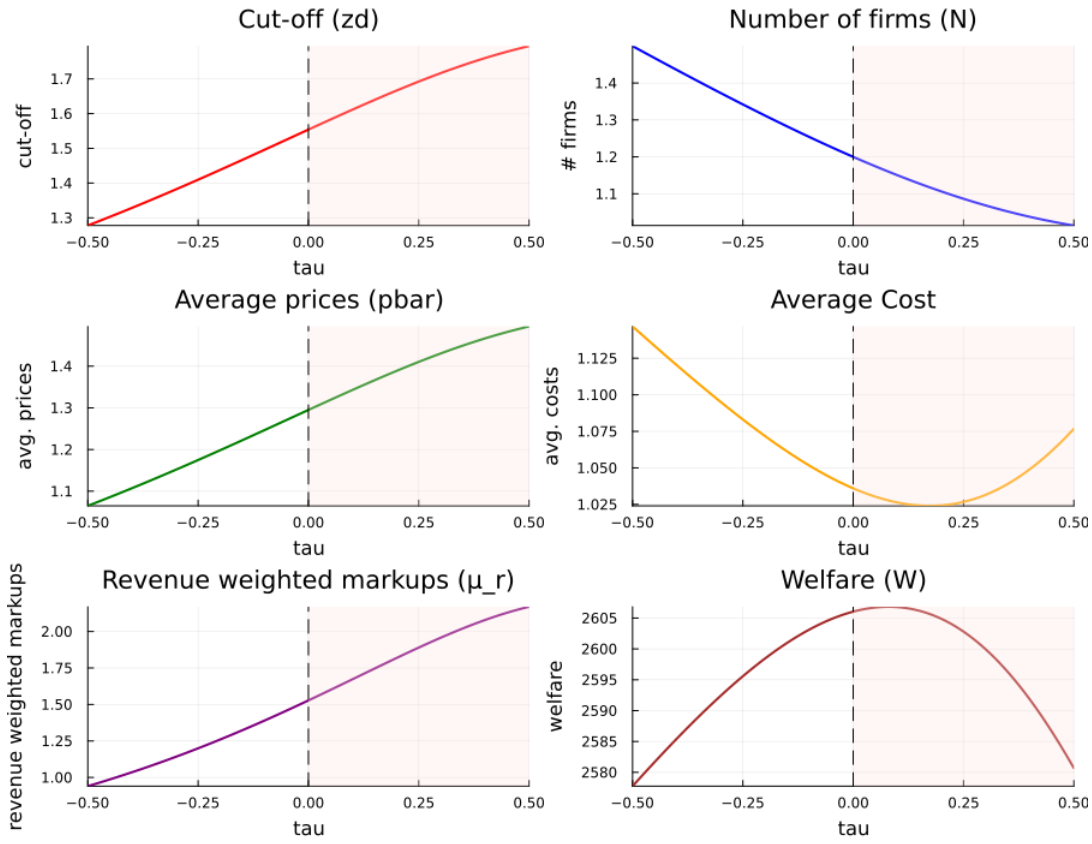


Figure 3: Comparative statics for τ_h fixing $\tau_l = 0$ in the pink shaded area. The dash line demarcates the frictionless benchmark ($\tau_l = \tau_h = 0$).

on to consumers. Welfare declines because this incomplete pass-through dominates the benefits of efficient resource use, and any positive effect from tax rebates may not be enough to offset it. In this situation, a policy that reduces misallocation would reduce aggregate markups and increase welfare.

Thus, in the presence of misallocation, the level and change in aggregate markups are not reliable indicators of welfare. Therefore, one should be cautious when making welfare judgments based solely on aggregate markups.

IV Empirical Strategy

This section outlines the empirical strategy suggested by the model and the measurement agenda that follows from it. The simple benchmark developed in the paper is useful for isolating the theoretical mechanism, but it is less well suited for separately identifying the market-power and distortion components of observed markups. Our empirical strat-

egy therefore takes the model as a guide for moving to a richer quantitative environment in which these two components can be separately identified and estimated.

The key idea is to embed the analysis in a demand system with variable elasticities, such as a Kimball aggregator or an Atkeson–Burstein environment [Kimball, 1995; Atkeson and Burstein, 2008]. In such a setting, firm-level markups can be written as the product of a distortion component and a market-power component:

$$\mu_i(\tilde{y}_i) = \left(\frac{1}{1-\tau_i}\right)\frac{\sigma(\tilde{y}_i)}{\sigma(\tilde{y}_i)-1} \equiv \left(\frac{1}{1-\tau_i}\right)\mu_i^K(\tilde{y}_i),$$

where τ_i denotes the revenue wedge and $\sigma(\tilde{y}_i)$ is the elasticity faced by firm i as a function of its market share or scale. Under this representation, the observed markup combines a wedge term and a residual-demand term. The advantage of this formulation is that it preserves the central insight of the theory — markups reflect both distortions and market power — while making the two sources of variation separately measurable.

The production-function approach remains useful in this environment under a key identifying assumption: distortions affect firms' scale decisions, but do not directly distort their input mix. Under this condition, the standard production approach can still recover the overall firm-level markup from the gap between output elasticities and expenditure shares. Once this object is recovered, the richer demand system can be used to back out the market-power component, and the residual can be interpreted as the distortion component. In this sense, the empirical strategy extends the production-function approach to markup estimation to environments with distortions, rather than treating observed markups as pure measures of market power. The empirical section therefore should be understood not as a standalone reduced-form exercise, but as a measurement strategy disciplined by the model.

A central implication of this approach is that aggregate efficiency losses can be decomposed into distinct components. Following the logic of [Li et al., 2025], a second-order approximation yields a decomposition of TFP losses of the form

$$\log Z^P - \log Z \approx \frac{\sigma}{2}\operatorname{Var}(\log\mu) + \frac{\sigma\alpha^2+\alpha(1-\alpha)}{2}\operatorname{Var}\left(\log\frac{1}{1-\tau}\right) + \sigma\alpha\operatorname{Cov}\left(\log\mu, \log\frac{1}{1-\tau}\right).$$

This decomposition is useful because it separates a component due to markup dispersion, a component due to distortion dispersion, and a covariance term between the two. The theory developed in this paper suggests that the covariance term is especially important in distorted environments, since endogenous selection generates correlation between

wedges and firm fundamentals even when wedges are ex ante unrelated to productivity. Quantifying the size and sign of this covariance term is therefore the main empirical object of interest.

In ongoing work, we implement this strategy in a richer quantitative setting and examine how important this covariance channel is in emerging-market environments. More broadly, the empirical agenda is intended to complement the paper's theoretical contribution: the model shows how distortions reshape the welfare content of aggregate markups, while the empirical strategy provides a path for measuring the relative importance of market power, wedges, and their interaction in the data.

V Conclusion

This paper develops a parsimonious framework to study the interaction between resource misallocation and market power, and to analyze its welfare consequences. The central message is that, in economies with distortions, aggregate markups are no longer a sufficient statistic for welfare. When firms face revenue wedges, distortions shape both the distribution of markups and the aggregate markup level through their effects on firm scale, market concentration, and equilibrium selection. As a result, the relationship between aggregate markups and welfare becomes non-monotonic: low aggregate markups need not signal healthy competition, and reductions in distortions can raise welfare, productivity, and aggregate markups at the same time. We have argued that this mechanism is especially relevant for emerging economies, where resource allocation is often shaped by substantial distortions.

Beyond the theoretical results, the paper also outlines an empirical strategy for taking this framework to the data. The model suggests that observed markups in distorted environments reflect both market power and wedges, and therefore should not be interpreted as pure measures of market power. In ongoing work, we extend the production-function approach to markup estimation to environments with distortions, with the aim of separating the market-power and distortion components of firm-level markups. This, in turn, provides a path toward quantifying aggregate efficiency losses through markup dispersion, distortion dispersion, and, most importantly, the covariance between the two.

A few limitations should be clear. The model is deliberately stylized and static, and the frictionless allocation it considers remains second-best . We nevertheless believe that the simplicity of the framework is a strength: it isolates a mechanism that is transparent,

tractable, and useful for understanding the joint behavior of markups, misallocation, and welfare.

REFERENCES

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms*. *The Quarterly Journal of Economics*, 135(2):645–709.

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

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications*. *The Quarterly Journal of Economics*, 135(2):561–644.

Díez, F. J., Fan, J., and Villegas-Sánchez, C. (2021). Global declining competition? *Journal of International Economics*, 132:103492.

Díez, F. J., Leigh, D., and Tambunlertchai, S. (2018). Global market power and its macroeconomic implications. Working paper, International Monetary Fund.

Eslava, M., Haltiwanger, J., and Urdaneta, N. (2024). The size and life-cycle growth of plants: The role of productivity, demand, and wedges. *The Review of Economic Studies*, 91(1):259–300.

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

Kaymak, B. and Schott, I. (2024). Tax Heterogeneity and Misallocation. *International Finance Discussion Papers*, (1393):1–47.

Liu, E. (2019). Industrial policies in production networks. *The Quarterly Journal of Economics*, 134(4):1883–1948.

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *The review of economic studies*, 75(1):295–316.

Restuccia, D. and Rogerson, R. (2017). The theory of misallocation and an evaluation of its empirical applications. *Journal of Economic Perspectives*, 31(3):151–174.