Concentrated Risk: Misallocation and Granular Business Cycles*

Tomer Ifergane†

September 2, 2024

Abstract

This paper uncovers a novel interaction between production efficiency and economic stability. Using a tractable heterogeneous firms model, I prove the existence of an efficiency-stability trade-off in granular economies. Specifically, reducing misallocation increases business cycle volatility. This trade-off originates because firms choose their optimal size without internalizing their effect on aggregate consumption risk. Utilizing approximations and results on order statistics, I propose a tractable method to quantify this effect and show that commonly studied misallocation counterfactuals involve a sizeable increase in business cycle volatility. I discuss how different assumptions on the nature of misallocation and factor mobility influence this result.

JEL: E32 D24 O47

Keywords: Business cycles, Misallocation, Granularity, Stabilization policies, Size-dependent policies.

*I am especially grateful to Ben Moll for his extensive feedback and continued support during my work on this paper. I also thank Daniel Albuquerque, Nadav Ben Zeev, Vasco Carvalho, Wouter den Haan, Maarten De Ridder, Xavier Gabaix, Lukas Freund, Jonas Gathen, Basile Grassi, David Lagziel, Ethan Ilzetzki, Dmitry Mukhin, Hugo Reichardt, Jane Olmstead Rumsey, Aditya Soenarjo, Yannick Schindler, Roy (JJ) Shmueli, and Ricardo Reis for their helpful feedback, comments and suggestions. I am also grateful to Matan Levintov for outstanding research assistance. This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No.51/22)

†Ben-Gurion University of the Negev, Department of Economics & London School of Economics and Political Science, Centre for Macroeconomics. iftomer@gmail.com

1 Introduction

Improving the efficiency of the production process and stabilizing economic activity are two crucial goals of economic policy. Indeed, understanding both the sources of misallocation and aggregate fluctuations has long been central to issues for economic research. Recently, the 'granular cycles hypothesis' emerged as a likely theory of business cycle shocks. The granular cycles hypothesis posits that firm-level shocks can manifest as aggregate shocks if the firm-size distribution is sufficiently skewed. This theoretical view can account for a sizable portion of total aggregate fluctuations and is supported by the empirical evidence on the firm-size distribution.

This paper embraces the granular view of the business cycle and formalizes the following idea. Any policy or fundamental of the economic environment that affects the firm-size distribution alters the degree to which granular shocks manifest as aggregate shocks. This simple observation has several implications which imply a tight link between misallocation and stability. First, misallocation is likely to affect macroeconomic stability because it affects the firm's optimal choice of size and the resulting firm-size distribution. Second, the firm chooses its optimal size without internalizing its effect on the degree of aggregate risk in the economy. If the representative household is risk averse, it might be beneficial to mitigate this risk. Thus, the granular cycles hypothesis gives rise to an efficiency stability trade-off where a central planner might be willing to have an efficiency loss to gain more certainty about output and consumption.

The first contribution of this paper is to formalize the relationship between misallocation and stability in granular economies. To do so, I employ a canonical heterogeneous firms model that allows for misallocation, as commonly modeled in the literature, to study the interaction between misallocation and business cycle volatility. I establish the following analytical result. Misallocation dampens business cycle volatility if distortions positively correlate with productivity, the empirically likely case. Alternatively stated, if misallocation disproportionately harms the more productive firms, it reduces aggregate volatility. Conversely, misallocation increases business cycle volatility if distortions are negatively correlated with productivity.

Obtaining a clean characterization of this result is challenging since one cannot allude to the law of large numbers. Therefore, even for a large number of firms, there's still uncertainty about the results that emerge in a finite sample. To overcome this, I must make stark assumptions regarding short-term factor mobility. I also discuss the implications of deviations from these assumptions on the resulting economy and explore this quantitatively.

In a second contribution, I derive an expression for welfare in the model that relates to the welfare gains from stability and the degree of risk aversion in a way that connects to Lucas' canonical cost of the business cycle analysis. I demonstrate how the level and volatility of aggregate productivity shape welfare. Welfare is increasing in the expected level of productivity and decreasing in its volatility. Anything that reduces misallocation increases productivity and its volatility, simultaneously giving rise to a trade-off between improving production efficiency and reducing stability.

These two contributions, taken together, have important implications for research and policy. First, it is common in the misallocation literature to compute counterfactual gains from alleviating misallocation. Such gains may be overstated if one ignores the value of stability. Second, size-dependent policies and misallocation more broadly can act as ex-ante automatic stabilizers. By allocating relatively fewer resources to the most productive firms than in the efficient production case, such policies may reduce the economy's sensitivity to shocks affecting large firms and allow for better risk diversification. Third, this paper demonstrates that any rise in concentration has the adverse consequence of increasing business cycle volatility.

Computing the effects of granular shocks on aggregate volatility is challenging, and doing so for a counterfactual scenario involving the same set of N firms is doubly so. The third contribution of this paper is to propose a general method to compute the effects of granular shocks on aggregate volatility. The method builds on approximations and the statistical properties of order statistics in a random sample and is parsimonious in data requirements. Its only requirements are the number of firms and a statistical description of the firm-size distribution as inputs. Thus, the method is applicable even when data is limited or when one wishes to compute the effects of granular shocks in a model-implied firm-size distribution. I validate the method by replicating the headline result from Carvalho and Grassi (2019), finding that both approaches attribute a similar magnitude of business cycle volatility to granular shocks.

Finally, this paper's last contribution is to quantify the implied change in business cycle volatility resulting from a counterfactual reform that improves production efficiency. I find that reducing misallocation in the model increases the implied aggregate volatility by between 10% to 80% depending on the assumptions about short-term factor mobility. This sizable and economically meaningful change supports the importance of the proposed mechanism.

This paper is first and foremost connected to the literature on granular business cycles, e.g., Gabaix (2011); Carvalho and Gabaix (2013); di Giovanni et al. (2018); Carvalho and Grassi (2019). The most closely related papers are Gabaix (2011), who was the first to propose the granular hypothesis, and Carvalho and Grassi (2019), who demonstrated that this mechanism is quantitatively meaningful within the context of the canonical heterogeneous firms model. Compared with Gabaix (2011), the present work illuminates novel implications of the granular hypothesis by linking it with the misallocation literature. The methods employed in this paper build on the tools and insights gleaned from the work of Carvalho and Grassi (2019). Conceptually, this paper differs from Carvalho and Grassi (2019) by comparing the effects of granular shocks in economies with misallocation and without it, whereas Carvalho and Grassi (2019) only consider the efficient benchmark. Additionally, this paper proposes a simplified framework compared with the one in Carvalho and Grassi (2019), allowing for tractable characterizations and reducing the computational burden, thus making the techniques introduced here more easily portable for future work.

Second, this paper connects the above literature on granular business cycles to the misallocation literature. The modern literature on misallocation is vast, starting from the works of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Within it, the most closely related works are those concerned with the interaction between firm or establishment size distribution and the allocation of production factors such as the works of Bartelsman et al. (2013); Hsieh and Klenow (2014); Bento and Restuccia (2017); Buera and Fattal Jaef (2018), and Poschke (2018a).[1] I contribute to this literature by demon-

[1] For a comprehensive review of this literature see the excellent review in Hopenhayn (2014).

strating the previously unexplored consequences of commonly studied counterfactuals. Namely, I show that granular shocks are amplified differently in an economy with and without misallocation. By leveraging the consensus in this empirical literature regarding the correlated nature of distortions, I conclude that reducing misallocation would also make granular shocks more pronounced and increase business cycle volatility.

Additionally, this paper is conceptually and methodologically related to the literature concerning Hulten's theorem (Hulten, 1978) and the transmission of shocks through production networks Acemoglu et al. (2012); Grassi (2018); Baqaee and Farhi (2019) and Baqaee and Farhi (2020). While the present paper abstracts from the presence of production networks, the analysis conducted here can be extended by Hulten's original theorem to be the first-order effect in any arbitrary production structure. My baseline is a model in which Hulten's theorem holds exactly. Doing so allows me to draw sharp predictions from a model involving a discrete number of firms without alluding to the law of large numbers, which would nullify the granular mechanism. However, I employ the insights gleaned by Baqaee and Farhi (2019) to demonstrate how and to what extent potential deviations from Hulten's theorem will affect the result and offer some guidance for future empirical investigations in assessing their importance to the proposed mechanism.

This paper is organized as follows. Section 2 presents the benchmark model and reiterates known results on granular fluctuations in a novel form. Section 3 presents the main theoretical result of the paper concerning the effect of misallocation on granular business cycles. Additionally, I derive the welfare effect of granular shocks in the environment and illustrate the tradeoff implied by misallocation counterfactuals. I also discuss the effects of assumptions on factor mobility on the previously derived results. Section 4 proposes an approximation method particularly suited to quantify the impact of granular shocks on aggregate volatility. To validate the method, I demonstrate that it can replicate the headline result of Carvalho and Grassi (2019) regarding the severity of granular cycles. Finally, I demonstrate how to perform misallocation counterfactuals on aggregate volatility and show that the effects might be economically meaningful. The final section concludes.

2 Benchmark Environment - Efficient Production

Technology There are N firms in the economy, each with a decreasing returns to scale technology that produce a single homogeneous consumption good y_i. Each firm i produces its output using labor hired at the beginning of the period[2] using the production function

$$y_i = z_i l_i^{\gamma}, \quad \log(z_i) = \log(a_i) + \tilde{x}_i, \tag{1}$$

where $0 < \gamma < 1$ denotes the degree of decreasing returns in the economy, and z_i is the productive ability of firm i. z_i is composed of a firm's ability a_i which is deterministic and a unit-mean state-dependent component $e^{\tilde{x}_i}$, such that its log $\tilde{x}_i$ is a mean-zero random variable with volatility σ_x, common to all N firms. I assume that the realizations of $\tilde{x}_i$ are iid. $\mathbf{s}$ will denote the aggregate state of the economy consisting of the N realized idiosyncratic values of $\tilde{x}_i$, thus, formally, $\mathbf{s} \in R^N$. Throughout this paper, I adopt the convention that tilde denotes stochastic variables, bold face letters denote column vectors, and aggregates are denoted in capital letters.

Under this environment there are no true aggregate shock in the model economy. All volatility arises purely from firm-level shocks and will be viewed as volatility along the state-space and not the time dimension. Aggregate volatility will be discussed with respect to the stochastic aggregates of a static economy. The form $\log(z_i) = \log(a_i) + \tilde{x}_i$ is particularly convenient since it allows one to consider aggregate fluctuations around a stable ergodic firm productivity distribution, i.e., the constant dispersion of a_i which under most conventional firm heterogeneity model would also generate, in equilibrium, the ergodic firm-size distribution and the measures of concentration in the economy. Thus, fluctuations $\tilde{x}_i$ can be viewed as the fluctuations of firms around that ergodic firm-size distribution. If one contemplates the business-cycle as fluctuations around a balanced growth path, then one can still interpret the values of a_i as pertaining to the de-trended productivity distribution, and to the shocks $\tilde{x}_i$ as the deviations from trend growth of each individual firm. By so doing, the model is well suited to study short-run fluctuations but not long-run dynamics and industry specific trends.

[2] Everything below carries through with capital and additional factors, a single factor model serves a presentational purpose.

Decision problem of the firm The firm choose how much labor to hire so as to maximize expected output. The firm's problem is as follows

$$\max_{l_j} \; E\left[y_j\right] - w \times l_j \tag{2}$$

The wage rate w will be determined in equilibrium by the market clearing condition

$$\sum_{i=1}^{N} l_i \; = L \tag{3}$$

Thus stated, production is ex-ante efficient with the expected marginal revenue product of labor equalized across all production units. The resulting equilibrium would yield the maximum *expected* output in the economy. It is equivalent to a situation whereby a central planner allocates labor across production units with full information of a_i but with only expectations of $\tilde{x}_i$. This problem is standard and the first order conditions and step by step derivation of the solution are given in appendix A.1 and results in the following allocation of labor

$$l_j = \frac{a_j^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}} L, \tag{4}$$

and in the following expected aggregate production function representation of the economy

$$E[Y_{\mathbf{s}}] = \sum_{i=1}^{N} E\left[y_i\right] \; = L^{\gamma} \left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right]^{1-\gamma}. \tag{5}$$

Few comments are in order. First, observe that the allocation rule for labor is scale invariant. I.e., scaling up or down the productivity of each firm by a constant factor $(1+g)$ leaves the allocation of labor identical and *relative* productive ability is all that matters. Second, as is standard in the literature, the model generates a non-degenerate firm-size distribution since $l_i > 0, \forall i$. Third, *aggregate expected total factor productivity* (TFP) is given by $E[Z_{\mathbf{s}}] = \left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right]^{1-\gamma}$ which is not the same as the realized TFP. To demonstrate, observed that the realized firm-level output is given by

$$y_j = \frac{a_j^{\frac{1}{1-\gamma}} e^{\tilde{x}_j}}{\left(\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}, \tag{6}$$

and thus the realized aggregate production function is given by

$$Y_{\mathbf{s}} = \sum_{i=1}^{N} y_i = \frac{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}} e^{\tilde{x}_i}}{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right]^{\gamma}} L^{\gamma} = Z_{\mathbf{s}} \times L^{\gamma}, \tag{7}$$

and thus aggregate TFP is given by $Z_{\mathbf{s}} = \frac{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}} e^{\tilde{x}_i}}{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right]^{\gamma}}$. Finally, observe that this aggregation result implies that expected TFP is strictly increasing in the number of firms. I will continue to relate to this expression as TFP rather than use the term to describe $Z_{\mathbf{s}}/N$ as is sometimes is done in the literature since $Z_{\mathbf{s}}$ is the correct theoretical counterpart in my environment to the classical Solow residual. This economy is a static description of a variant of the industry dynamics model of Hopenhayn (1992).

Business cycle volatility Let us now proceed to study the effect of a one percent granular shock to firm j. To do so I define $\eta_{j,\mathbf{s}}$ as the elasticity of aggregate TFP with respect to a one percent shock to $\tilde{x}_j$ and thus to its relized productive ability $z_{j,\mathbf{s}}$

$$\eta_{j,\mathbf{s}} = \frac{\partial \log(Z_{\mathbf{s}})}{\partial \tilde{x}_j} = \frac{1}{Z_{\mathbf{s}}} \frac{a_j^{\frac{1}{1-\gamma}} e^{\tilde{x}_j}}{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right]^{\gamma}} = \frac{a_j^{\frac{1}{1-\gamma}} e^{\tilde{x}_j}}{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}} e^{\tilde{x}_i}}. \tag{8}$$

$\eta_{j,\mathbf{s}}$ is also the Domar weight of the jth firm since from Equation (7) we obtain that $\frac{y_j}{Y_{\mathbf{s}}} = \frac{a_j^{\frac{1}{1-\gamma}} e^{\tilde{x}_j}}{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}} e^{\tilde{x}_i}} = \eta_{j,\mathbf{s}}$. We have thus varified that Hulten's theorem (Hulten, 1978), which relates the Domar weight, or the sale's share, of a firm or a sector to the effect an idiosyncratic shock to it will have on the aggregate holds exactly in the presented benchmark model. Hulten's theorem relays on the envelope theorem and thus only holds up to a first order[3] Unlike Hulten's theorem, the above derivation involves no envelope condition and is the exact solution to the aggregate representation of the model economy. Thus innocent of the critique against Hulten's theorem. However, by verifying that Hulten's theorem indeed holds in this environment, I can generalize the results obtained in this model economy as the first order representation of TFP

[3] See Baqaee and Farhi (2019) for a comprehensive treatment of this.

volatility in any economy with an arbitrary production network. Quantitatively, the shocks described throughout this paper should be interpreted as the usual churn of economic activity "graden-variety fluctuations" if you will, and emphatically not as catastrophic events like the 2007-2008 financial crisis or the COVID19 recession which are arguably "true" aggregate shocks.

The above derivation hold regardless of the distributional assumptions on a_i and $\tilde{x}_i$, however these distributions would vastly affect the resulting amplification. The distribution of $\eta_{j,\mathbf{s}}$ in the economy is crucial for its aggregate volatility. This key insight originates in the seminal work of Gabaix (2011) which illustrates that if the sales shares are distributed according to a heavy-tailed distribution, granular shocks of a reasonable magnitude can generate business cycles of a quantitatively likely scale even in the presence of a large number of firms. To understand the significance of the TFP elasticities note first that $\eta_{j,\mathbf{s}}$ is the relative size of the jth firm compared to the total size of all the firms in the economy. Second, given this interpretation, we see that a one percent shock to a large firm will affect aggregate TFP more than a shock to a small firm would. Third, in this simple model of a horizontal economy, all of these elasticities sum up to unity, $\sum_{i=1}^{N}\eta_{i,\mathbf{s}} = 1$ which simply means that reducing the productive ability of all firms by 1% also reduces TFP by 1%. Using these, we can conclude that a key feature that the environment needs to have to generate amplification from granular shocks is a very skewed firm-size distribution. To illustrate, suppose that all firms are of the same ability, the above expression would simply collapse into $\eta_{j,\mathbf{s}} = \frac{1}{N}$. However, if the firm-size distribution is sufficiently skewed, we could obtain large amplification, at the extreme when one firm controls the entire market or $\eta_{j,\mathbf{s}} \to 1$ we obtain a perfect pass-through from one granular shock to the aggregate.

Using these elasticities, it is possible to characterize the volatility of log TFP, σ_{TFP} as follows

Proposition 1. *The volatility of log TFP is given by*

$$\sigma_Z \approx \sigma_x \sqrt{\sum_{i=1}^{N} \overline{\eta_i}^2} = \underbrace{\sigma_x}_{\textit{micro volatility}} \times \underbrace{\Psi}_{\textit{amplification term}}, \tag{9}$$

where σ_x *is the volatility of* $\tilde{x}_i$ *and* $\overline{\eta_i} = \frac{\partial \log[Z(\mathbf{a},\mathbf{x}=0)]}{\partial \tilde{x}_i}$.

For proof see Appendix A.2.[4] Simply put, Proposition 1, states that the volatility of aggregate TFP is proportional to the volatility of the micro-level shock and to an amplification term Ψ that is a function of concentration in the economy without micro-level shocks, or the innate skewness of the ability distribution when the idiosyncratic shocks are shut down, $\Psi = \sqrt{\sum_{r=1}^{N} \overline{\eta_i}^2}$. This amplification term is also the square root of the Herfindahl-Hirschman Index (HHI) in the economy without shocks since $\overline{\eta_i}$ is the market share of firm i. Since this is an horizontal economy without downstream production, the amplification is bounded above by unity, as an economy with a single firm that controls the entire market will have an HHI of unity. The amplification term is also bounded below by $\frac{1}{\sqrt{N}}$ in the case of an economy with N firms each having a market share of $\frac{1}{N}$.

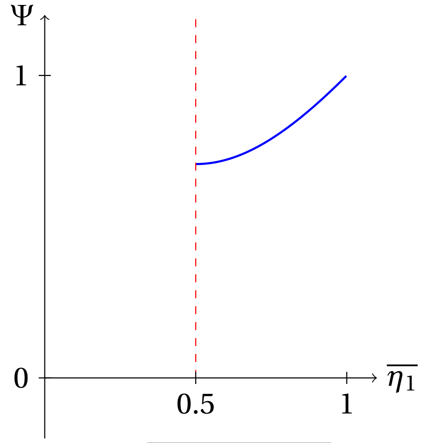
To demonstrate this transmission from the micro-level shock to the aggregate, suppose that there are only two firms in the economy $N = 2$. Further suppose, without loss of generality, that $\overline{\eta_1} \geq \overline{\eta_2}$. Aggregate volatility in this economy is given by $\Psi = \sqrt{\overline{\eta_1}^2 + \left(1 - \overline{\eta_1}\right)^2}$. Figure 1 reports the value of the amplification term for different levels of the market share of the largest firm $\overline{\eta_1}$. Intuitively, aggregate volatility is lowest if both firms are of the same size and highest if one firm controls the entire market. This intuition generalizes to the general case as the more skewed the firm size distribution the more volatile the economy will be all else equals.

In this stylized economy, the firm's choice of size, i.e. how much labor to use in production, is ex-ante efficient. That is, it maximize the expected output of the economy. If we were to measure the dispersion of the marginal product of labor as a measure of efficiency, we would get that it is constant up to some noise component arising from the shock $\tilde{x}_i$.

To build intuition for the next section, one can ask how an economy with inefficient production would differ from the efficient benchmark. Through the lens of Figure 1, the question amounts to what would be the influence of production inefficiency on $\overline{\eta_1}$? We would expect an increase in volatility as a result of introducing misallocation into the model if results in the first firm's market share increases. Conversely, if the first firms market share decreases as a result of misallocation, then volatility would decrease.

[4] This result is similar in flavor to Equations (4) and (5) in Gabaix (2011).

Figure 1: The Amplification Term Ψ in a Two-Firm Economy



$$\Psi = \sqrt{\overline{\eta_1}^2 + \left(1 - \overline{\eta_1}\right)^2}$$

3 Misallocation and Aggregate Volatility

In this section I introduce misallocation into the benchmark model and demonstrate its effects on the resulting economy, its aggregate representation and the implied volatility. The headline result of this section is to show how misallocation influences volatility compared to the efficient benchmark. Specifically, I will prove that the correlation of distortions and underlying ability pins down the direction of influence - positively correlated distortion dampen volatility while negatively correlated distortion amplify it. The former case is favoured by the literature as the likely scenario. The size of the effect would ultimately depend on the skewness of the firm-size distribution and the dispersion of the underlying distortions themselves. I proceed by discussing the influences of my assumptions regarding short-run factor mobility.

3.1 Aggregate Volatility in a Distorted Economy

Compared with the efficient benchmark, the production technology remains unchanged, there are still N firms producing with the same decreasing returns to scale technology. However, the firm faces an implicit output tax τ_i which distorts its decision problem as follows:

$$\max_{l_i} \; E\left[y_i(1-\tau_i)\right] - w l_i \tag{10}$$

This modified decision problem would result in a distorted labour allocation as follows

$$l_i = L\frac{\left(a_i(1-\tau_i)\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}\left(a_i(1-\tau_i)\right)^{\frac{1}{1-\gamma}}}, \tag{11}$$

which is derived in an analogues fashion to the one in the efficient case. [5]

Note that having unequal wedges distorts the allocation of labor between firm i and j in a way that does not directly depend on their productive ability. However, each firm would ultimately produce $y_i = a_i e^{\tilde{x}_i} l_i{}^{\gamma}$ units of output. Thus, the marginal product of labor would not be equated across firms and the production process will be inefficient ex-ante. In the case with equal implicit taxes, i.e., $(1-\tau_i) = \left(1-\tau_j\right), \forall i, j$, the above collapses into the allocation of labor expressed in Equation (4). Having the implicit taxes change the allocation of labor compared to the efficient case gives rise to misallocation as modeled by Hsieh and Klenow (2009) and Restuccia and Rogerson (2008). I assume throughout that the implicit taxes are strictly positive, i.e., $(1-\tau_i) > 0, \forall i$.

Aggregating the Distorted Economy By summing output across all firms we can obtain that realized aggregate output in the economy is now given by

$$Y_s = \sum_{i=1}^{N} y_i = L^{\gamma}\frac{\sum_{i=1}^{N} a_i{}^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_i}}{\left[\sum_{i=1}^{N}\left(a_i(1-\tau_i)\right)^{\frac{1}{1-\gamma}}\right]^{\gamma}}, \tag{12}$$

where TFP in the distorted economy is now given by $Z_{\mathbf{s}}{}^{d} = \frac{\sum_{i=1}^{N} a_i{}^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_i}}{\left[\sum_{i=1}^{N}\left(a_i(1-\tau_i)\right)^{\frac{1}{1-\gamma}}\right]^{\gamma}}$, where the superscript d is intended to distinguish between the efficient case and the distorted case. Observe that TFP is homogeneous of degree zero with respect to changes in the scale of the weights. In particular, multiplying all $(1-\tau_i)$ in some positive constant would leave the value of TFP unaltered. It is only the relative values of these weights that affect the allocation of labor in the resulting economy.

[5] for a step by step derivation for this see Appendix A.3.

Misallocation and amplification We can use the same concepts of TFP elasticities and amplification term we defined in the efficient benchmark economy to study the volatility of the distorted economy. The, now distorted, TFP elasticities are given by

$$\delta_{j,\mathbf{s}} = \frac{\partial \log\left(Z_{\mathbf{s}}{}^{d}\right)}{\partial \tilde{x}_j} = \frac{1}{Z_{\mathbf{s}}{}^{d}} \frac{a_j{}^{\frac{1}{1-\gamma}}\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_j}}{\left[\sum_{i=1}^{N}\left(a_i(1-\tau_i)\right)^{\frac{1}{1-\gamma}}\right]^{\gamma}} = \frac{a_j{}^{\frac{1}{1-\gamma}}\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_j}}{\sum_{i=1}^{N} a_i{}^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_i}} \tag{13}$$

Note that, viewed through these elasticities, misallocation changes the sensitivity of aggregate TFP to firm-level shocks. The sum of the elasticities doesn't change, i.e., $\sum_{i=1}^{N}\delta_{i,\mathbf{s}} = \sum_{i=1}^{N}\eta_{i,\mathbf{s}} = 1$ and a shock that reduces the productive ability of all firms by one percent would still lower aggregate TFP by one percent. However, the relative sizes of these elasticities does change in the distorted economy compare with the efficient benchmark. Intuitively, since aggregate volatility depends disproportionately on the elasticities of TFP to the large firms' shock, if misallocation increases the elasticity of TFP with respect to larger firms' shocks at the expense of TFP's elasticity to smaller firms, it will lead to more aggregate volatility. However, if misallocation increases the elasticity of TFP with respect to smaller firms' shocks at the expense of TFP's elasticity to shocks to larger firms, then misallocation will lead to lower aggregate volatility. Observe also that in the distorted economy presented above, Hulten's theorem still holds exactly.[6] Thus, sales shares are still sufficient to understand the amplification due to granular fluctuations.

We can now use the rationale of Proposition 1, to characterize aggregate volatility of TFP in the distorted economy as

$$\sigma_Z^d \approx \sigma_x \times \sqrt{\sum_{i=1}^{N} \overline{\delta i}^2} = \sigma_x \times \Psi^d, \tag{14}$$

where Ψ^d is the amplification term in the distorted economy and $\overline{\delta_i} = \frac{\partial \log Z_{\mathbf{s}}^d}{\partial \tilde{x}_i}|_{(\tilde{\mathbf{x}}=0)}$ is the TFP elasticity with respect to a shock to the i^{th} firm when all other shocks are shut down or $z_i = a_i$. Any difference in aggregate volatility be-

[6] To verify that statement we can see that the sales share of firm i and the i^{th} TFP elasticity are identical $\delta_{i,\mathbf{s}} = \frac{y_{is}}{Y_s} = \frac{a_i{}^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}} e^{x_{is}}}{\sum_{i=1}^{N} a_i{}^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}} e^{x_{is}}}$.

tween the distorted economy and the efficient benchmark would amount to a difference between the amplification term in each economy Ψ^d and Ψ correspondingly. In what follows, I analyze the relationship between the two.

To relate the efficient and inefficient case it is useful to define the distortion d_j as the change in the elasticity of TFP with respect to firm j when all firm-level shocks are at their expected level as follows

$$\overline{\delta_j} = \frac{a_j^{\frac{1}{1-\gamma}}\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{i=1}^N a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}} = \frac{a_j^{\frac{1}{1-\gamma}}}{\sum_{i=1}^N a_i^{\frac{1}{1-\gamma}}} \frac{\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}}{(1-\hat{\tau})^{\frac{1}{1-\gamma}}} = \overline{\eta_j}\sqrt{1-d_j}, \tag{15}$$

where the value of $\hat{\tau}$ is defined as the following weighted average

$$(1-\hat{\tau})^{\frac{\gamma}{1-\gamma}} = \sum_{i=1}^N \overline{\eta_i}(1-\tau_i)^{\frac{\gamma}{1-\gamma}} .$$

Given the assumptions on a_j and τ_j we can conclude that $\sqrt{1-d_j} > 0$.

In the misallocation literature broadly, the term distortion is usually applied to the values of $1-\tau_j$, and the concept of positively correlated distortions implies that misallocation disproportionally harms the high ability firms by more than it does the low ability firms. Negatively correlated distortions imply the converse, low ability firms are harmed more by misallocation.

My newly defined distortion d_j is similar in essence but defined in terms of the elasticities which are incidently the sales shares here. If misallocation increases the market share of the j^{th} firm, we will say that the distortion is negative, whereas if misallocation reduced the market share of j, the distortion is said to be positive. Note also that $\overline{\eta_j}$ is the relative ability of the firm. The most able firm has the highest value of $\overline{\eta_j}$. Thus, if the size of the distortion is positively correlated with $\overline{\eta_j}$, it is also positively correlated with the ability of the firm making my notion of positively correlate distortions consistent with the one in the literature. In section 4.5, I will also demonstrate how the two notions coincide using a parametric form used widely in the misallocation literature.

3.2 Comparing TFP Volatility in a Distorted Economy with the Efficient Benchmark

Using Equation (15) now to relate the values of $\overline{\eta_i}$, and $\overline{\delta_i}$ using the concept of the distortions we can state the amplification term in the distorted economy as follows:

$$\Psi_d = \sqrt{\sum_{i=1}^{N} \overline{\delta_i}^2} = \sqrt{\sum_{i=1}^{N} \overline{\eta_i}^2 (1 - d_i)} \,. \tag{16}$$

The above equation demonstrates that amplification in the distorted economy is governed by the elasticities in the efficient benchmark $\overline{\eta_i}$ and the distortion with respect to it given by the d_i. By examining the squares of the above equation and using the definition in Equation (9) we can obtain the following:

$$\Psi^2 - \Psi_d{}^2 = \sum_{i=1}^{N} \overline{\eta_i}^2 d_i = ||\mathbf{d}|| \times \left|\left|\overline{\boldsymbol{\eta}}^{\mathbf{2}}\right|\right| \times \cos(\theta) \tag{17}$$

where $||\cdot||$ denotes the Euclidian norm, vector powers denoted the elementwise operation, and since $\sum_{i=1}^{N} \overline{\eta_i}^2 d_i$ is the dot product of two vectors we can use cosine similarity where θ is the angle between the two as follows $\cos(\theta) = \frac{\sum_{i=1}^{N} \overline{\eta_i}^2 d_i}{||\mathbf{d}|| \times \left|\left|\overline{\boldsymbol{\eta}}^{\mathbf{2}}\right|\right|}$.

Several clarifications are in order. Since the values of $\overline{\eta_i}^2$ are strictly positive and the values of d_i are not the sign of the above expression depends on the association between the two. If distortions are positively correlated with ability, we will have that a positive value of d_i is associated with a high value of $\overline{\eta_i}$ and thus $\overline{\eta_i}^2$, and the cosine similarity between the two is positive. However, if a high value of $\overline{\eta_i}$ is associated with negative distortions, the value of the cosine similarity is negative. Additionally, the expressions $||\mathbf{d}||$, and $\left|\left|\overline{\boldsymbol{\eta_i}}^{\mathbf{2}}\right|\right|$ are strictly positive and relate to the overall dispersion of the distortions and the dispersion of the elasticities $\overline{\eta_i}$ in the benchmark case which is related to the skewness of the firm-size distribution. Thus, we can state the following theorem

Theorem 2. *Misallocation affects the amplification of granular shocks as follows*

1. *Positively correlated distortions dampen granular volatility compared to the efficient benchmark.*

2. *Negatively correlated distortions amplify granular volatility compared to the efficient benchmark.*

3. *The strength of the above differences is more pronounced when:*

   (a) *The cosine similarity* $\cos(\theta)$ *is higher in absolute value.*

   (b) *The efficient benchmark contains a more skewed distribution of the elasticities* $\overline{\eta_i}$.

   (c) *The dispersion of the distortions themselves is higher.*

Positive or negative correlation? The bulk of the misallocation literature, starting from Hsieh and Klenow (2009), and Restuccia and Rogerson (2008) argues that the positive correlation of distortions and ability is an important feature. Hsieh and Klenow (2014) find a positive correlation between distortions and ability using data and from India and Mexico concerning the life-cycle of plants and relating it to the United States data. They also report model implied positive elasticities for the United States. Similar positive elasticities are used to match establishment sizes and firm sizes in various countries, e.g., Bartelsman et al. (2013), Buera and Fattal Jaef (2018); Bento and Restuccia (2017); Poschke (2018b), and David and Venkateswaran (2019).

An alternative approach to thinking about misallocation is to consider its sources directly. Two compelling cases are market power and financial frictions. In the case of market power, consider for example the model of Atkeson and Burstein (2008), in this type of model, more productive firms within a sector are able to charge higher markups and are under-producing compared to an efficient production benchmark. This would manifest as a higher τ_i for the high ability firms or as a positively correlated distortions.[7] In the case financial frictions a-la Kiyotaki and Moore (1997), as applied in the models of Buera and Shin (2013), and Moll (2014), financial frictions limit the amount of inputs a producer can use as a function of their wealth. The higher is their production ability, the more capital they wish to employ and thus the higher is the implicit tax imposed on them by the financial friction. Alternatively stated, high ability individuals are harmed more by the existence of a constraint on their input structure, giving rise to positively correlated distortions.

[7] For more on this logic see Edmond et al. (2015).

3.3 The Stabilizing or Destabilizing Role of Policies: A Concrete Policy Example

To provide better intuition for the result in Theorem 2, it is instructive to examine the effects of real tax policies on the efficient benchmark economy. Consider an economy whereby all firms are uniformly subject to a revenue tax at a rate of t_0, which is non-distortionary, and either (i) the government decided to subsidize or offer tax breaks to small businesses or (ii) large firms get tax credit due to their size and disproportionate political influence, or (ii). The former case will be referred to as 'SME subsidies'[8] and the latter as 'corruption'. I use size as measured by employment here for tractability. Note that SME subsidies are an instance of positively correlated distortion and corruption of negatively correlated ones. However, unlike the above case, these taxes and subsidies are now made explicit. Both of these examples are stylized representations of size-dependent policies more broadly.

In the case of corruption, the policy is such that after tax revenues are a fraction $1 - t(l_i) = (1 - t_0)\left(\frac{l_i}{L}\right)^{\nu}$ of actual revenues. Observe that this fraction is bounded between zero and $1 - t_0$ and is increasing in l_i, and aggregate labor L serves as a normalizing constant. In the case of SME subsides after tax revenues are similarly given by $1 - t(l_i) = (1 - t_0)(1 + s_0)\left(\frac{l_i}{L}\right)^{-\nu}$ that is now decreasing in l_i and s_0 is a policy parameter for the subsidy. In both cases ν denotes the policy's elasticity with respect to size.[9] In what follows I derive how firm behavior is affected by SME subsides and TFP volatility as a result. The results hold for the case of corruption with the opposite sign. For a step-by-step derivation see Appendix A.4

3.4 Efficiency Stability Tradeoff

Gains from reducing misallocation These examples clarify that the likely scenario is that of positively correlated distortions. Thus, any counterfactual analysis that involves alleviating misallocation, necessarily involves increasing output volatility. To demonstrate the tradeoff inherent in such an exer-

[8] Subsidies geared towards small and medium enterprises (SMEs) are common in the development context and are in place in many places. See XXX

[9] For technical reasons it is necessary to impose in the case of SME subsidies that $0 \leq \nu < \gamma$ and for the case of corruption that $0 \leq \nu < 1 - \gamma$ otherwise there is no interior solution to the firm's problem.

cise, consider now that welfare in the economy is given by $u(Y_{\mathbf{s}})$, i.e., there is a representative household in the economy that consumes all output produced. The household has a concave utility function with a constant relative risk aversion (CRRA) parameter of χ. Suppose without loss of generality that aggregate labor supply is normalized to unity. Therefore, we can express utility simply as $u(Z_{\mathbf{s}})$.

Consider the following second-order Taylor series approximation for welfare where the approximation is taken around $\overline{Z} = Z_{\mathbf{s}}\,(\tilde{\mathbf{x}} = 0)$

$$\mathrm{u}(Z_{\mathbf{s}}) \approx \tag{18}$$
$$\mathrm{u}\left(\overline{Z}\right) + \left[\mathrm{u}'\left(\overline{Z}\right)\sum_{i=1}^{N}\frac{\partial Z_{\mathbf{s}}}{\partial \tilde{x}_i}\right]\tilde{x}_i + \frac{1}{2}\left[\sum_{i=1}^{N}\left(\mathrm{u}''\left(\overline{Z}\right)\left[\frac{\partial Z_{\mathbf{s}}}{\partial \tilde{x}_i}\right]^2 + \mathrm{u}'\left(\overline{Z}\right)\frac{\partial^2 Z_{\mathbf{s}}}{(\partial \tilde{x}_i)^2}\right)\right]\tilde{x}_i^2.$$

We can take expectations around the above and use the fact that ${\sigma_x}^2 = E\left[\tilde{x}_i^2\right]$ to obtain

$$E[\mathrm{u}(Z_{\mathbf{s}})] \approx \mathrm{u}\left(\overline{Z}\right) + \frac{{\sigma_x}^2}{2}\left[\sum_{i=1}^{N}\left(\mathrm{u}''\left(\overline{Z}\right)\left[\frac{\partial Z_{\mathbf{s}}}{\partial \tilde{x}_i}\right]^2 + \mathrm{u}'\left(\overline{Z}\right)\frac{\partial^2 Z_{\mathbf{s}}}{(\partial \tilde{x}_i)^2}\right)\right] \tag{19}$$

We can exploit the elasticities $\delta_{i,\mathbf{s}}$ to express the derivatives as $\frac{\partial Z_{\mathbf{s}}}{\partial \tilde{x}_i} = \delta_{i,\mathbf{s}} \times Z_{\mathbf{s}}$. This expression allows one to derive that $\frac{\partial^2 Z_{\mathbf{s}}}{(\partial \tilde{x}_i)^2} = \frac{\partial \delta_{i,\mathbf{s}}}{\partial \tilde{x}_i} \times Z_{\mathbf{s}} + \delta_{i,\mathbf{s}}\frac{\partial Z_{\mathbf{s}}}{\partial \tilde{x}_i}$. Additionally, one can compute $\frac{\partial \delta_{i,\mathbf{s}}}{\partial \tilde{x}_i}$ and obtain that $\frac{\partial^2 Z_{\mathbf{s}}}{(\partial \tilde{x}_i)^2} = \delta_{i,\mathbf{s}} Z_{\mathbf{s}}$.[10] We can combine those derivatives and exploit the definition of the amplification term to obtain that $E[\mathrm{u}(Z_{\mathbf{s}})] = \mathrm{u}\left(\overline{Z}\right) + \frac{{\sigma_x}^2}{2}\overline{Z}^2\,\mathrm{u}''\left(\overline{Z}\right){\Psi_d}^2 + \frac{{\sigma_x}^2}{2}\overline{Z}\,\mathrm{u}'\left(\overline{Z}\right)$. Finally, by ex-

[10] This can be shown directly from

$$\frac{\partial \delta_{j,\mathbf{s}}}{\partial \bar{x}_j} = \frac{a_j^{\frac{1}{1-\gamma}}(1-\tau_j)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_j}\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_i}\right] - \left[a_j^{\frac{1}{1-\gamma}}(1-\tau_j)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_j}\right]^2}{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_i}\right]^2} =$$
$$\frac{\left(a_j^{\frac{1}{1-\gamma}}(1-\tau_j)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_j}\right)}{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_i}\right]}\,\frac{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_i}\right] - \left[a_j^{\frac{1}{1-\gamma}}(1-\tau_j)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_j}\right]}{\left[\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}e^{\bar{x}_i}\right]} = \delta_{j,\mathbf{s}}\left(1-\delta_{j,\mathbf{s}}\right),$$

which implies $\frac{\partial^2 Z_{\mathbf{s}}}{(\partial \bar{x}_i)^2} = \frac{\partial \delta_{i,\mathbf{s}}}{\partial \bar{x}_i} \times Z_{\mathbf{s}} + \delta_{i,\mathbf{s}}\frac{\partial Z_{\mathbf{s}}}{\partial \bar{x}_i} = \frac{\partial \delta_{i,\mathbf{s}}}{\partial \bar{x}_i} \times Z_{\mathbf{s}} + {\delta_{i,\mathbf{s}}}^2 \times Z_{\mathbf{s}} = Z_{\mathbf{s}} \times \left[\delta_{j,\mathbf{s}}\left(1-\delta_{j,\mathbf{s}}\right) + {\delta_{i,\mathbf{s}}}^2\right] = \delta_{j,\mathbf{s}} Z_{\mathbf{s}}$.

ploiting the CRRA utility specificiation, i.e., $\chi = -\frac{\overline{Z}\,\mathrm{u}''\left(\overline{Z}\right)}{\mathrm{u}'\left(\overline{Z}\right)}$, we have that

$$E[\mathrm{u}(Z_{\mathbf{s}})] = \mathrm{u}\left(\overline{Z}\right) + \frac{{\sigma_x}^2}{2}\overline{Z}\,\mathrm{u}'\left(\overline{Z}\right) - \chi {\Psi_d}^2 \frac{{\sigma_x}^2}{2}\overline{Z}\,\mathrm{u}'\left(\overline{Z}\right). \tag{20}$$

Any counterfactual that alleviates misallocation in this economy increases $\overline{Z}$, which is the standard welfare gain associated with improving production efficiency. However, if distortions are positively correlated, we would have that the increase reduces the welfare in Ψ_d that would ensue. To further demonstrate this point, suppose we were to evaluate the welfare gain from a policy counterfactual that moves the economy from being distorted, i.e., having TFP $Z^d_{\mathbf{s}}$, to its efficient production counterpart with TFP $Z_{\mathbf{s}}$. Suppose that preferences take log form for tractability reasons only, thus $\chi = 1$. Using Equation (20) for each case and differentiating the two, we can obtain that

$$\underbrace{E[\mathrm{u}(Z_{\mathbf{s}})] - E\left[\mathrm{u}\left(Z^d_{\mathbf{s}}\right)\right]}_{\text{Counterfactual welfare gain}} = \underbrace{\mathrm{u}\left(\overline{Z}\right) - \mathrm{u}\left(\overline{Z}^d\right)}_{\substack{\text{Gain from}\\ \text{production efficiency}}} \underbrace{- \frac{{\sigma_x}^2}{2}\left[\Psi^2 - {\Psi_d}^2\right]}_{\text{Stability loss}}. \tag{21}$$

The first term in the left-hand side of the above equation is the standard welfare gain we would obtain from a misallocation counterfactual without aggregate volatility. Consumption in the efficient production case will be, on average, higher; therefore, this term is positive. The second term corresponds to the difference in Equation (17), and its sign is given by Theorem 2. Given a positive correlation between distortion and ability, we obtain that the second term is negative. Thus, the welfare gain described by the first term overstates the total effect, which should also internalize the stability value.

Observe also that while market clearing conditions and labor market equilibrium dictate the level of output and consumption in this economy. From an ex-ante perspective, it might be the case that if the social planner could alter the allocation of labor, we would obtain a different allocation due to the *granular externality*. Namely, the firm which chooses it size by demanding labor in a market, does so without internalizing its own decision's influence on the aggregate consumption risk. It is possible that given sufficient risk aversion the social planner would allocate ex-ante less labor than would seem optimal from a strict output maximization perspective in order to attenuate consumption risk

3.5 Short-Term Factor Mobility and the Amplification of Granular Shocks

An assumption that underlays the entire analysis thus far is that the allocation of labor is done without knowledge of the realized values of $\tilde{x}_i$. Alternatively, I assume that factors are allocated ex-ante and after learning on the realisations of the shocks, it is impossible to reallocate them. From the perspective of a short run analysis I believe this is a likely scenario in practice.[11] However, the stylized nature of my model allows me to try and gauge the importance of deviations from that assumption.

Suppose we were to solve a problem similar to (10) but with full knowledge of the values of $\tilde{x}_i$ to illustrate the key differences that would arise compare to my analysis in this section thus far. For ease of comparison I will use the same notations when possible to facilitate an easy comparison between the two cases, for completeness the full derivation is included in Appendix A.5. In both the efficient case studied in Section 2 and the inefficient case described in this section Hulten's theorem held exactly. Thus, letting the output share of firm j in the efficient case by $s_{Y,j}$, for the inefficient case by $s^d_{Y,j}$, and their corresponding values when $\tilde{\boldsymbol{x}} = 0$ by $\bar{s}^d_{Y,j}$ and $\bar{s}_{Y,j}$. I compared the volatility in both cases by exploiting the following relationship

$$\overline{\delta_j} = \bar{s}^d_{Y,j} = \bar{s}_{Y,j} \times \sqrt{1-d_j} = \overline{\eta_j} \times \sqrt{1-d_j}. \tag{22}$$

When we allow the choice of labor to be done with full knowledge of the firm level shocks we obtain that

$$\overline{\delta_j} = \bar{s}^d_{Y,j} + \frac{\gamma}{1-\gamma}\left(\bar{s}^d_{Y,j} - \bar{s}^d_{L,j}\right), \tag{23}$$

,where $\bar{s}^d_{L,j}$ denotes the input share of firm j in the ineffiecent case where $\tilde{\boldsymbol{x}} =$

[11]For example, Google's CEO in announcing major layoffs in January 2023 stated that: "Over the past two years we've seen periods of dramatic growth. To match and fuel that growth, we hired for a different economic reality than the one we face today."The full statement is available at https://blog.google/inside-google/message-ceo/january-update/. Thus, hinting that hiring decisions had been made in the absence of information about the present state of the economy.

0. For the explicit derivation of the above see Appendix A.5. In the efficient case, input and output shares are identical so it would always be the case that $\overline{s}_{Y,j} - \overline{s}_{L,j} = 0$ and Hulten's theorem would hold exactly. However, in the presence of misallocation input shares and output shares are not necessarily aligned and we have the extra higher-order term $\frac{\gamma}{1-\gamma}\left(\overline{s}^{d}_{Y,j} - \overline{s}^{d}_{L,j}\right)$.[12]

The expression in Equation (23) is not necessarily positive and that depends on the exact values for the implicit taxes or the severity of miallocation. When the difference between input and output shares is sufficiently pronounced, the elasticity $\overline{\delta_j}$ might be negative. Such a negative elasticity has profound implications for understanding the effects of granular shocks on aggregate volatility as follows. Suppose that a firm has a negative elasticity, it implies that we are so far removed from the efficient case that the input share of this firms is sufficiently high compared to its output share or that $\overline{s}^{d}_{Y,j} < \gamma \overline{s}^{d}_{L,j}$ to be exact. Suppose further that this firm experiences a positive shock. Such a shock draws in more resource into this firm at the expense of other, more productive firms, thus reducing TFP as a result. The converse also holds since a negative shock to this firm frees inputs to be elsewhere employed.

These negative elasticities render the previously introduced transformation impractical. This is because we cannot map between the ability of the firm and the size of the squared elasticity which would be high for extreme negative or positive values. Thus breaking the previously established link in this case. Through this logic our ability to understand the effects of granular shocks on aggregate volatility in this case would depend on the covariance between input and output shares. Specifically, recall that earlier we had that

$$\Psi^2_d = \sum_{i=1}^{N} \overline{\delta}_i^{\,2} = \sum_{i=1}^{N} \left(\overline{s}^{d}_{Y,i}\right)^2, \tag{24}$$

which is also the HHI of the economy computed using output shares. How-

[12]For formal discussion of Hulten's theorem and higher-order effects in that context see Baqaee and Farhi (2019). The particular expression derived here and the effects it generates concerning the propagation of shocks described in the next paragraphs are similar to the effects detailed in Baqaee and Farhi (2020) concerning the propagation of shocks in an horizontal economy in relation to the inverse harmonic mark-up. Examining the formulas derived in Appendix A.5 will show similar ratios between weighted averages of the implicit taxes.

ever, if the allocation of labor is done with full information of the shocks we have that

$$\Psi_d{}^2 = \left[1 + 2\frac{\gamma}{1-\gamma} + \left(\frac{\gamma}{1-\gamma}\right)^2\right] \underbrace{\sum_{i=1}^{N} \left(s_{Y\,i}{}^d\right)^2}_{\text{Sales HHI}} + \tag{25}$$

$$\left(\frac{\gamma}{1-\gamma}\right)^2 \underbrace{\sum_{i=1}^{N} \left(s_{Li}{}^d\right)^2}_{\text{Input HHI}} - 2\left[\frac{\gamma}{1-\gamma} + \left(\frac{\gamma}{1-\gamma}\right)^2\right] \underbrace{\sum_{i=1}^{N} s_{Y\,i}{}^d s_{Li}{}^d}_{\text{Olley-Pakes term}} .$$

Thus, the total volatilty in this economy is a function of input and output concentration and the association shares between the two which is related to the covariance term in the Olley-Pakes decompostion (Olley and Pakes, 1996), but stated in terms of shares instead of firm-level productivity and using a non-centered measure. The total effect of misallocation on volatility in this case will depend crucially on this third term. In the next section I will demonstrate quantitatively that it is still likely the case that the prediction of 2 continues to hold true even in this analytically less predictable case.

4 Quantification and Validation

After introducing the amplification term Ψ and illustrating how it is theoretically influenced by misallocation. In this section I will introduce an approximation strategy for computing Ψ and performing counterfactual analyses concerning it. The major technical challenge in this section is that Ψ is a function of the exact set of N firms in an economy or a sector. That is, it is possible for the same primitive ability distribution $F(a_i)$ to yield different values of Ψ given different sets of realisations of firms. This problem is enhanced given the extremely skewed nature of the empirical firm-size distribution. The firm-size distribution in the US exhibits a tail parameter that is close to unity, e.g. Axtell (2001) finds a tail parameter of 1.059. Therefore, if one were to simulate multiple economies of a large scale using the exact same parametric distribution of for the individual firms' ability a_i the resulting set of economies will have a non-negligable dispersion in the realised values of the amplification term Ψ even if they contain a large number of firms. Thus, the approximation introduced here relates to *the expected value* of Ψ for a given number of firms N and a parametric ability distribution with CDF F.

An additional challenge is thinking about counterfactual scenarios involving the *same* N firms after applying a change to the degree of production efficiency in the model. I.e., one needs to compute not only the values of the elasticities $\overline{\delta_i}$, but also the corresponding values of $\overline{\eta_i}$ for the same realizations of a_i. For tractability, I will start from the efficient production case.

4.1 The expected amplifiction term - approximation strategy

The amplification term Ψ is a function of a randomly drawn sample of N values of a_i drawn from a CDF F. To approximate it, first, observe that by substituting Equation (8) into Equation (9) we obtain

$$\Psi = \sqrt{\sum_{i=1}^{N} \overline{\eta_i}^2} = \sqrt{\sum_{i=1}^{N} \left(\frac{a_i^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}} \right)^2} = G(\boldsymbol{a}). \tag{26}$$

Given that $G(\boldsymbol{a})$ is a function of a random sample, we can always rearrange the vector $\mathbf{a}$ such that it contains ordered observations given the ranks or the order statistics. Order statistics are defined as the largest, second largest, third largest, and so forth realizations of the sample.[13] Now let us define the quantile function as the inverse of the CDF such that $Q(q_i) = a_i$. Thus, drawing a_i from a distribution F directly, is identical to drawing q_i from the uniform distribution such that $q_i \in [0,1]$ and computing a_i using the quantile function.[14] We can thus restate $G(\boldsymbol{a})$ as follows

$$G(\boldsymbol{a}) = \hat{G}(\boldsymbol{q}) = \frac{\sqrt{\sum_{r=1}^{N} Q(q_r)^{\frac{2}{1-\gamma}}}}{\sum_{r=1}^{N} Q(q_r)^{\frac{1}{1-\gamma}}}. \tag{27}$$

Let us now define the vector $\bar{\mathbf{q}}$ as the vector of the expected order statistics of the uniform distribution. This vector is useful since the r^{th} order statistic is distributed according to a Beta distribution $q_r \sim Beta(N+1-r, r)$ and thus its expectations $\bar{q}_r = E[q_r] = \frac{N+1-r}{N+1}$.[15] Therefore, we can derive the following

[13] The above uses the reverse of the conventional ordering whereby order statistics are presented from the smallest, second smallest, and so forth. This has no bearing on any result and is done purely for presentational convenience.

[14] In fact, this is a common method to simulate random draws from a distribution.

[15] For a formal treatment of this with the conventional ordering see Gentle (2009).

first-order Taylor series expansion around $\bar{\mathbf{q}}$

$$\hat{G}(\boldsymbol{q}) \approx \hat{G}(\bar{\mathbf{q}}) + \sum_{r=1}^{N} \frac{\partial \hat{G}}{\partial q_r}(q_r - \bar{q}_r). \tag{28}$$

Thus, taking expectations of the above, letting $\bar{\Psi} = E\left[\hat{G}(\boldsymbol{q})\right]$ we obtain

$$\bar{\Psi} \approx \hat{G}(\bar{\mathbf{q}}) = \frac{\sqrt{\sum_{r=1}^{N} Q(\bar{q}_r)^{\frac{2}{1-\gamma}}}}{\sum_{r=1}^{N} Q(\bar{q}_r)^{\frac{1}{1-\gamma}}}. \tag{29}$$

This approximation will be used later on to qunatify the severity of granular shocks in the model. Note that higher order approximations are conceptually possible since the covariances of order statistics are also straightforward to obtain, see Gentle (2009) for exact formulas. However, these are computationally infeasible because for an economy populated with $N = 10^6$ firms, this covariance matrix has 10^{12} entries. Still, if one wishes to use the approximation strategy suggested here for the sectoral level, the required approximation is given below for completeness

$$\bar{\Psi} \approx \hat{G}(\bar{\mathbf{q}}) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2 \hat{G}(\overline{\mathbf{q}})}{\partial q_i \partial q_j} COV(q_i, q_j) \ . \tag{30}$$

4.2 Introducing Pareto distributions

The Pareto distribution is of a particular interest for works on the firm-size distribution since it is empirically supported and analytically tractable. Many works on the firm-size distribution concern themselves only with the Pareto class. The literature on granular shocks that abstracts from network effect is almost entirely done using Pareto distributions, e.g., Gabaix (2011); Carvalho and Grassi (2019).

Formally, let the ability distribution in the economy take Pareto form with a tail parameter $\zeta_a > 0$ and scale parameter that is normalized to unity[16], i.e. the CDF for a_i is given by $\mathrm{F}(a_i) = \mathrm{Prob}(a < a_i) = 1 - {a_i}^{-\zeta_a}$. The immediate implication of this assumption is as follows

[16]This assumption is without loss of generality because granular amplification is affected by the distribution of relative sizes.

Lemma 3. *The firm-size distribution is a Pareto distribution with tail parameter* $\zeta = \zeta_a (1-\gamma)$.

Proof. Given the firm's labour policy function, output and profits, all definitions of firm size coincide with a term that is proportional to $a_i^{\frac{1}{1-\gamma}}$. Given that a_i is drawn from a Pareto distribution, we can obtain that

$$\text{Prob}\left(a^{\frac{1}{1-\gamma}} > x\right) = \text{Prob}\left(a > x^{1-\gamma}\right) = x^{-\zeta_a(1-\gamma)} = x^{-\zeta}.$$

□

Several clarifications are in order. First, the observable tail parameter in Proposition 3 is $\zeta = \zeta_a (1-\gamma)$ which is the empirically observed tail of the firm-size distribution. E.g., for the U.S. the estimate in Axtell (2001) is that $\zeta = 1.059$. The empirical counterpart of $1-\gamma$ is the profit share in the economy or sector. These two can be combined to infer ζ_a. To keep the notations clear, in what follows ζ will always denote the empirically observed tail of the firm-size distribution. Second, since $1-\gamma < 1$, the tail parameter of the ability distribution is higher than that of the firm-size distribution. Alternatively, the distribution of ability is more equal than the resulting firm-size distribution. This is because the ability to hire labor or other factors magnifies smaller differences in ability into larger differences in size.

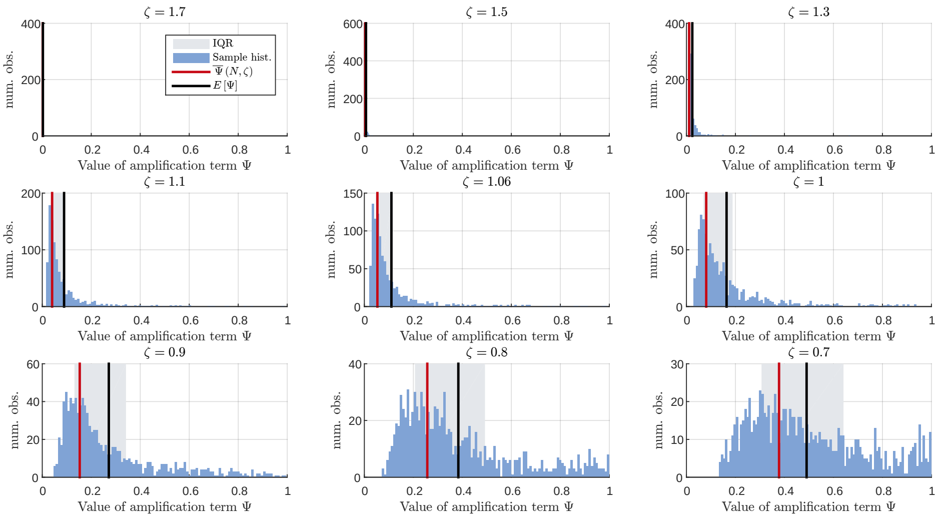
Approximating the amplification term for a Pareto distribution Using the result from Equation (29), the fact that the quantile function of a Pareto distribution is given by $\mathrm{F}^{-1}(a_i) = \mathrm{Q}(q_i) = (1-q_i)^{-\frac{1}{\zeta_a}}$, and the fact that $\mathrm{E}(q_r) = \frac{N+1-r}{N+1}, \forall r \in \{1,\ldots,N\}$ we can obtain the following approximation for the expected amplification term in an economy populated with N firms:

$$\overline{\Psi} = \frac{\sqrt{\sum_{r=1}^{N} r^{-\frac{2}{\zeta}}}}{\sum_{r=1}^{N} r^{-\frac{1}{\zeta}}}. \tag{31}$$

4.3 Assessment of approximation quality

In this section I will assess quantitatively the accuracy of the approximation method introduced above. Following that, I will discuss the amplification po-

Figure 2: The approximation term in randomly drawn economies



Note: This figure reports histograms of the computed amplification term Ψ in 10^3 random samples each containing $N = 4.5 \times 10^6$ firms drawn from a Pareto distribution with a tail parameter ζ. The red line corresponds to the approximate value $\overline{\Psi}(N,\zeta)$, the black one to the sample mean of the amplification term, and the shaded area to the IQR.

tential of different firm-size distributions and will relate them to known results from the granularity literature.

I begin by drawing 10^3 random samples of Pareto distributed firms containing $N = 4.5 \times 10^6$ firms each, using different tail parameters for the firm-size distribution. The results of this exercise are given in Figure 2 which reports for each tail parameter the histogram of resulting values of the amplification term Ψ, the range between the 25^{th} and the 75^{th} percentiles (IQR) and the approximate amplification term $\overline{\Psi}$. The parameter ranges are selected to include a wide range of values of the tail parameter of the firm-size distribution.

Figure 2 shows how the dispersion of the realized amplification term relates to the tail of the firm-size distribution. The fatter the tail (smaller ζ), the more dispersion we observe for the implied amplification term. The first row of Figure 2 corresponds to value that are draw from distributions without a finite second moment. Still, there is little dispersion in the amplification term. The second row reports the amplification term for parameters often used in the literature concerning the US firm-size distribution. In this empirically likely range, we see more dispersion of the amplification term, with some samples obtaining values close to $\Psi = 1$. Recall that values even close to $\Psi = 1$ imply a near perfect pass-through of 'micro' shocks from one firm to the aggregate even in a sample containing millions of firms. When the tail of the firm-size distribution falls below unity, as suggested by some works on the firm-size distribution, we obtain a high degree of dispersion for Ψ and more skewed realizations. Figure 2 demonstrates that the approximation strategy described in this section obtains likely values for the amplification term. Furthermore, it provides a conservative estimate which understates the value of the amplification term compared its sample mean which is highly influenced by extreme realizations.

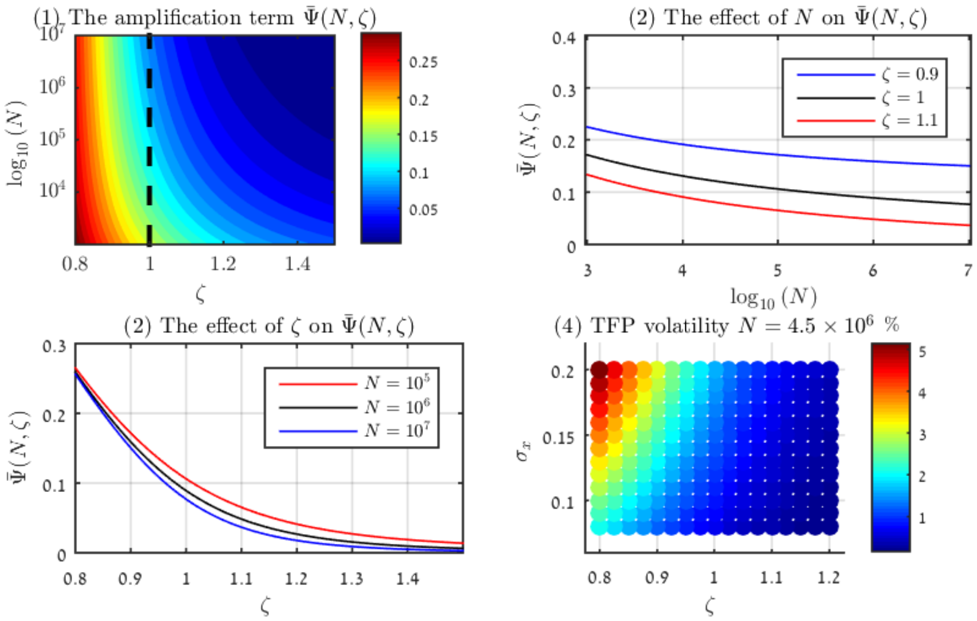
4.4 Validation - How strong is granular amplification?

In this section I will use the approximation method introduced above and estimates from the literature to compute the amplification term under different assumptions. I will illustrate the sensitivity of the amplification term to changing the skewness of the firm-size distribution and the number of firms. Since the closest framework to mine in the literature is Carvalho and Grassi (2019), I will validate my predictions by replicating their headline result. Namely, Carvalho and Grassi (2019) find that firm level idiosyncratic shocks can account for about a quarter of the volatility in TFP (0.25% out of an annual TFP volatility of 1.04%). The results of this section are reported in Figure 3.

Panel (1) of Figure 3, reports the value of the expected amplification term $\bar{\Psi}$ as a function of the number of firms N and the tail of the firm-size distribution ζ which is denoted with some abuse of notation as $\bar{\Psi}(N, \zeta)$. Observe that Panel (1) also suggests that amplification is lower if N or ζ are higher. An illustration of the decay of the amplification term in N is given in Panel (2) for different values of the tail parameter. Observe that this decay is weaker as ζ is smaller. Additionally, the relationship between ζ and the amplification term is illustrated more clearly in Panel (3) holding for different values of N.

I now proceed to examine the overall severity of granular cycles using Equation (9). To apply it, one requires an estimate of the volatility of the firm-level Solow residual. Carvalho and Grassi (2019) provide a survey of the liter-

Figure 3: Computing the Amplification Term and Assessing the Severity of Granular Cycles



Note: Panel (1) reports the values of the amplification term $\Psi(N,\zeta)$ for different values of the tail parameter of the firm-size distribution and N. Panel (2) reports how the amplification term decays as N increases. Panel (3) reports the amplification term as a function of the tail parameter holding the number of firms N constant. Panel (4) reports the volatility of log TFP that arise from firm-level shocks for different combinations of σ_x and the tail.

ature estimating this volatility and report values ranging between 0.08 – 0.2. The analysis in Carvalho and Grassi (2019) employs $\sigma_x = 0.08$ as its benchmark along with a tail estimate of $\zeta = 1.1$ and $N = 4.5 \times 10^6$. Under these parameters the amplification term is given by $\bar{\Psi}\left(4.5 \times 10^6, 1.1\right) = 0.04$. Thus, the volatility of log TFP that can be attributed to granular fluctuations is $\sigma_x \bar{\Psi}\left(4.5 \times 10^6, 1.1\right) =$ 0.32%. This number is within the same order of magnitude to the 0.25% that is reported by Carvalho and Grassi (2019). My number is slightly larger probably due to the numerical implementation in Carvalho and Grassi (2019) which requires truncating the upper support of $F(a_i)$ and discretizating it, whereas my framework has no such requirement. Panel (4) of Figure 3 demonstrates a range of possible estimates for the contribution of granular cycles to total volatility based on different estimates of the tail and σ_x

4.5 The effects of misallocation on granular amplification

To evaluate quantitatively the effects of changes in the degree of misallocation on aggregate volatility. I follow Bento and Restuccia (2017); Poschke (2018a); Buera and Fattal Jaef (2018) in specifying that distortions are positively correlate and take the following single parameter specification

$$(1-\tau_i) = a_i^{-\phi}, \tag{32}$$

where $\phi \in [0,1]$ corresponds to the elasticity of the distortions with respect to ability. Observe that an increase in a_i implies a higher value of τ_i. This form is particularly useful as it illustrates the effect of misallocation on the firm-size distribution. Observe that labor and sales in the model are proportional to $a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}$ that is firm-size is proportional to $a_i^{\frac{1-\gamma\phi}{1-\gamma}}$. Using the logic of Lemma 3, we have that the firm-size distribution has the following observed tail $\zeta = \zeta_a \frac{1-\gamma}{1-\gamma\phi}$ which is larger than in the efficient benchmark if ϕ is larger than zero. A larger tail implies a less skewed firm size distribution.

Viewed through the lens of this parametric example, Theorem 2 becomes a very intuitive result, positively correlated distortions means that the firm-size distribution is less skewed and thus granular volatility is reduced. The converse is also true, reducing ϕ from a positive level to zero or alleviating misallocation implies a more skewed firm-size distribution and more amplification of granular shocks.

Recalling that Ψ_d is given by $\sqrt{\sum_{i=1}^{N} \overline{\delta i}^2}$, we can utilize the quantile function again to approximate the amplification term in the misallocation term as follows

$$\Psi = \sqrt{\sum_{i=1}^{N} \overline{\delta_i}^2} = \sqrt{\sum_{i=1}^{N} \frac{a_j^{\frac{1}{1-\gamma}}\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}}^2} = G_d(\boldsymbol{a}), \tag{33}$$

where applying the quantile function and 32 allows us to obtain

$$G_d(\boldsymbol{a}) = \hat{G}_d(\boldsymbol{q}) = \frac{\sqrt{\sum_{r=1}^{N} Q\left(q_r\right)^{\frac{2(1-\gamma\phi)}{1-\gamma}}}}{\sum_{r=1}^{N} Q\left(q_r\right)^{\frac{1-\gamma\phi}{1-\gamma}}}, \tag{34}$$

Table 1: The effect of misallocation on aggregate volatility

		(a) Efficient case	(b) With misallocation	(c) W/o misallocation
(1)	Baseline $\phi = 0.09$	0.32%	0.32%	0.56%
(2)	$\phi = 0.05$	0.32%	0.32%	0.44%
(3)	$\phi = 0.2$	0.32%	0.32%	1.06%
(4)	$\phi = 0.3$	0.32%	0.32%	1.74%
(5)	Perfect mobility $\phi = 0.09$	0.32%	0.58%	0.64%
(6)	Perfect mobility $\phi = 0.05$	0.32%	0.46%	0.47%
(7)	Perfect mobility $\phi = 0.2$	0.32%	1.04%	1.37%
(8)	Perfect mobility $\phi = 0.3$	0.32%	1.67%	2.34%

which again by using a Taylor series approximation around the order statistics $\overline{\boldsymbol{q}}$ allow us to obtain

$$\overline{\Psi_d} = \hat{G}_d\left(\overline{\boldsymbol{q}}\right) = \frac{\sqrt{\sum_{r=1}^{N} r^{-\frac{2(1-\gamma\phi)}{\zeta a(1-\gamma)}}}}{\sum_{r=1}^{N} r^{-\frac{1-\gamma\phi}{(\zeta a(1-\gamma))}}}. \tag{35}$$

Do misallocation counterfactuals have sizeable stability implications? To quantify the effect of improving production efficiency or a counterfactual change in the severity misallocation in the form of positively correlated distortions, I perform the following exercise. I compute the implied TFP volatility attributable granular shocks in a case with positively correlated distortions and without. To do so, I calibrate the model using $N = 4.5 \times 10^6$, $\zeta = 1.1$, γ and $\sigma_x = 8\%$ following Carvalho and Grassi (2019). Additionally, using the estimates of Hsieh and Klenow (2014) I also calibrate the value of to $\phi = 0.09$. Thus, distortions are assumed to be positively correlated with ability. The results from this exercise are reported in Table 1.

Table 1 reports the aggregate volatility that can be attributed to granular shocks under different assumptions. The first column (a) reports the TFP volatility under the efficient benchmark. This number corresponds to the headline result of Carvalho and Grassi discussed earlier. The second column (b) reports TFP volatility if we assume that the benchmark involves correlated distortions as presented earlier in Equation (32). The last column (c) reports the TFP volatility in a counterfactual scenario where I remove misallocation.

The first row corresponds to the calibrated values introduced in the previous paragraph. Observe that improving production efficiency, going from (b) to (c), increases TFP volatility. I.e., business cycles get worse when we alleviate misallocation as predicted by Theorem 2. The magnitude of the effect is economically meaningful, with a 75% increase in business cycle volatility compared to the baseline.

This result hinges on the size of the elasticity ϕ, which is a difficult object to quantify. The estimate used is a model implied object for the establishment size distribution in the United States. Thus the mapping to a model of firms is not perfect. Hsieh and Klenow (2014) also report an elasticity of 0.50 in India and 0.66 in Mexico. Poschke (2018b) discusses elasticities ranging between zero and 0.3 as consistent with various countries. Thus, compared to the literature, the value of $\phi = 0.09$ is relatively modest. For a discussion of these elasticities and their implications, see Bento and Restuccia (2017). The rows (2) through (4) repeat this exercise for different elasticities ϕ.

Given the discussion in section 3.5, it is also instructive to study the same thought experiment in the case where we assume perfect factor mobility in response to the shock. These are reported in rows (4) to (8). Observe that these cases involve an increase in volatility compared with the corresponding efficient case calibrated to the same parameters. The difference between column (a) and (b) in these rows is between an efficient and inefficient economy exhibiting the same skewness of the firm size distribution as manifested by the input shares.[17] That result is consistent with the analysis reported in Appendix L of Baqaee and Farhi (2019). However, removing misallocation in each of these scenarios involves even more volatility, consistent with the results thus far presented. Some cases even attribute far greater volatility to aggregate TFP in column (b) than the empirically observed one, which is around 1% annually, making them unlikely.

[17]This is because most empirical evidence of the firm size distribution, the definition of size is with respect to employment.

5 Concluding remarks

This paper studies how firm-level shocks may affect macroeconomic stability. Using a tractable heterogeneous firms model, I characterize how misallocation counterfactuals and, indirectly, reforms aimed at improving efficiency increase TFP volatility. The effects generalize to a large class of economies studied in the existing literature, uncovering omitted implications of counterfactual scenarios analyzed in the literature. The findings of this paper may be helpful in guiding future empirical works on the effects of granular shock, misallocation and the interaction of the two.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Atkeson, A. and A. Burstein (2008, December). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Axtell, R. L. (2001). Zipf distribution of us firm sizes. *science 293*(5536), 1818–1820.

Baqaee, D. and E. Farhi (2019, 2019). The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem. *Econometrica 87*(4), 1155–1203.

Baqaee, D. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *Quarterly Journal of Economics*.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013, February). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review 103*(1), 305–34.

Bento, P. and D. Restuccia (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics 9*(3), 267–303.

Buera, F. J. and R. N. Fattal Jaef (2018, June 29). The dynamics of development: Innovation and reallocation. *World Bank Policy Research Working Paper* (8505).

Buera, F. J. and Y. Shin (2013). Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy 121*(2), 221–272.

Carvalho, V. and X. Gabaix (2013, August). The great diversification and its undoing. *American Economic Review 103*(5), 1697–1727.

Carvalho, V. M. and B. Grassi (2019, April). Large firm dynamics and the business cycle. *American Economic Review 109*(4), 1375–1425.

David, J. M. and V. Venkateswaran (2019, July). The sources of capital misallocation. *American Economic Review 109*(7), 2531–67.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2018, January). The micro origins of international business-cycle comovement. *American Economic Review 108*(1), 82–108.

Edmond, C., V. Midrigan, and D. Y. Xu (2015, October). Competition, markups, and the gains from international trade. *American Economic Review 105*(10), 3183–3221.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Gentle, J. (2009). *Computational Statistics.* Statistics and Computing. Springer New York.

Grassi, B. (2018). IO in I-O: Size, Industrial Organization, and the Input-Output NetworkMake a Firm Structurally Important. Technical report.

Hopenhayn, H. A. (1992, September). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica 60*(5), 1127–1150.

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics 6*(1), 735–770.

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

Hsieh, C.-T. and P. Klenow (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics 129*(3), 1035–1084.

Hulten, C. R. (1978, 10). Growth Accounting with Intermediate Inputs. *The Review of Economic Studies 45*(3), 511–518.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy 105*(2), 211–248.

Moll, B. (2014, October). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review 104*(10), 3186–3221.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Poschke, M. (2018a). The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics 10*(3), 1–41.

Poschke, M. (2018b, July). The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics 10*(3), 1–41.

Restuccia, D. and R. Rogerson (2008, October). Policy Distortions and Aggregate Productivity with Heterogeneous Plants. *Review of Economic Dynamics 11*(4), 707–720.

Appendix A Additional derivations

This appendix provides proof that are not included in the main text and additional step-by-step derivation of several results.

A.1 Step by Step Derivation of the Efficient Benchmark

The first order condition for the production problem 2, is given by

$$\gamma a_j {l_j}^{\gamma-1} = w. \tag{A.1}$$

we can rearrange this expression as

$$l_j = \left(a_j e^{x_j}\right)^{\frac{1}{1-\gamma}} \left(\frac{\gamma}{w}\right)^{\frac{1}{1-\gamma}}, \tag{A.2}$$

and use the market clearing condition for labor to obtain

$$\sum_{i=1}^{N} l_i = \left(\frac{\gamma}{w}\right)^{\frac{1}{1-\gamma}} \sum_{i=1}^{N} {a_i}^{\frac{1}{1-\gamma}} = L. \tag{A.3}$$

Substituting the above relationship as $\left(\frac{\gamma}{w}\right)^{\frac{1}{1-\gamma}} = \frac{L}{\sum_{i=1}^{N} \left(a_i e^{x_i}\right)^{\frac{1}{1-\gamma}}}$, into A.2 yields

$$l_j = \frac{{a_j}^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N} {a_i}^{\frac{1}{1-\gamma}}} L, \tag{A.4}$$

and therefore

$$y_j = \frac{{a_j}^{\frac{1}{1-\gamma}} e^{\tilde{x}_j}}{\left(\sum_{i=1}^{N} {a_i}^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}. \tag{A.5}$$

Aggregating this allows us to obtain

$$Y_{\mathbf{s}} = \sum_{i=1}^{N} y_i = \frac{\sum_{i=1}^{N} {a_i}^{\frac{1}{1-\gamma}} e^{\tilde{x}_i}}{\left(\sum_{i=1}^{N} {a_i}^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}. \tag{A.6}$$

A.2 Proof of Proposition 1

Given the idiosyncratic shock structure introduced, TFP in the economy can be restated with some abuse of notation as

$$Z_{\mathbf{s}} = \frac{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}} e^{\tilde{x}_i}}{\left(\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}}\right)^{\gamma}} = Z(\tilde{\mathbf{x}}). \tag{A.7}$$

Using a Taylor series expansion, we can derive the approximation given in Proposition 1 as follows

$$\begin{aligned} &VAR\big[\log(Z(\tilde{\mathbf{x}}))\big] \approx \\ &VAR\left[\log(Z(\tilde{\mathbf{x}}=0)) + \sum_{i=1}^{N} \frac{\partial \log(Z(\tilde{\mathbf{x}}=0))}{\partial \tilde{x}_i} \tilde{x}_i\right] = {\sigma_x}^2 \sum_{i=1}^{N} \overline{\eta_i}^2, \end{aligned} \tag{A.8}$$

where $\overline{\eta_i} = \frac{\partial \log(Z(\tilde{\mathbf{x}}=0))}{\partial \tilde{x}_i}$. This ends the proof.

A.3 Step by Step Derivation of the Economy with Misallocation

The first order condition for the production problem 10, is given by

$$\gamma a_j (1-\tau_i) {l_j}^{\gamma-1} = w. \tag{A.9}$$

we can rearrange this expression and use the market clearing condition for labor to obtain

$$\sum_{i=1}^{N} l_i = \left(\frac{\gamma}{w}\right)^{\frac{1}{1-\gamma}} \sum_{i=1}^{N} (a_i(1-\tau_i))^{\frac{1}{1-\gamma}} = L. \tag{A.10}$$

Thus, in equilibrium we would have that

$$l_j = \frac{\left(a_j\left(1-\tau_j\right)\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N} (a_i(1-\tau_i))^{\frac{1}{1-\gamma}}} L, \tag{A.11}$$

and

$$y_j = \frac{a_j^{\frac{1}{1-\gamma}} \left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_j}}{\left(\sum_{i=1}^{N} (a_i(1-\tau_i))^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}. \tag{A.12}$$

Aggregating this allows us to obtain

$$Y_{\mathbf{s}} = \sum_{i=1}^{N} y_i = \frac{\sum_{i=1}^{N} a_i^{\frac{1}{1-\gamma}} (1-\tau_i)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_i}}{\left(\sum_{i=1}^{N} (a_i(1-\tau_i))^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}. \tag{A.13}$$

A.4 The Model Economy With Size-Dependent Policies

The firm in an environment with SME subsidies solves the problem of maximizing

$$\max_{l_i} \quad R_{SME} a_i l_i^{\gamma-\nu} - w \times l_i, \tag{A.14}$$

with $R_{SME} = (1-t_0)(1+s_0)L^{\nu}$. Using the first order condition we can show that

$$l_i = \left(R_{SME} a_i \frac{\gamma-\nu}{w}\right)^{\frac{1}{1-(\gamma-\nu)}}. \tag{A.15}$$

Using the market clearing condition for labor we can obtain that

$$\sum l_i = \left(\frac{\gamma-\nu}{w} R_{SME}\right)^{\frac{1}{1-(\gamma-\nu)}} \sum (a_i)^{\frac{1}{1-(\gamma-\nu)}} = L. \tag{A.16}$$

Combined with the first order condition for labor we can obtain that output before taxes and subsidies at the firm level is given by

$$y_i = e^{\tilde{x}_i} a_i^{\frac{1}{1-(\gamma-\nu)}} \left[\frac{L}{\sum (a_i)^{\frac{1}{1-(\gamma-\nu)}}}\right]^{\gamma} \tag{A.17}$$

A.5 The Economy with Misallocation when Labor is Allocated Ex-post

The first order condition for the production problem 10 when the value of $\tilde{x}_i$ is known, is given by

$$\gamma a_i e^{\tilde{x}_i} (1-\tau_i) l_i^{\gamma-1} = w. \tag{A.18}$$

we can again rearrange this expression and use the labor market clearing condition to obtain

$$\sum_{i=1}^{N} l_i = \left(\frac{\gamma}{w}\right)^{\frac{1}{1-\gamma}} \sum_{i=1}^{N} (a_i(1-\tau_i))^{\frac{1}{1-\gamma}} = L. \tag{A.19}$$

Thus, in equilibrium we would have that

$$l_j = \frac{\left(a_j e^{\tilde{x}_j}\left(1-\tau_j\right)\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}\left(a_i(1-\tau_i)\right)^{\frac{1}{1-\gamma}}} L, \tag{A.20}$$

and

$$y_j = \frac{\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}} e^{\tilde{x}_j}}{\left(\sum_{i=1}^{N}\left(a_i e^{\tilde{x}_i}(1-\tau_i)\right)^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}. \tag{A.21}$$

Aggregating this by summing across all production units allows us to obtain

$$Y_{\mathbf{s}} = \sum_{i=1}^{N} y_i = \frac{\sum_{i=1}^{N}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}}{\left(\sum_{i=1}^{N}\left(a_i e^{\tilde{x}_i}(1-\tau_i)\right)^{\frac{1}{1-\gamma}}\right)^{\gamma}} L^{\gamma}. \tag{A.22}$$

To derive the value of the elasticity TFP elasticities, recall that TFP is now given by

$$Z_{\mathbf{s}}^{d} = \frac{\sum_{i=1}^{N}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}}{\left[\sum_{i=1}^{N}\left(a_i e^{\tilde{x}_i}(1-\tau_i)\right)^{\frac{1}{1-\gamma}}\right]^{\gamma}}. \tag{A.23}$$

Therefor we can derive the elasticity by

$$\frac{\partial \log Z_{\mathbf{s}}^{d}}{\partial \tilde{x}_j} = \frac{1}{Z_{\mathbf{s}}^{d}} \frac{\frac{1}{1-\gamma}\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}\left[\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{x_i}\right)^{\frac{1}{1-\gamma}}\right]^{\gamma} - \gamma\left[\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}\right]^{\gamma-1}\left(\left(1-\tau_j\right)a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}} \frac{1}{1-\gamma}\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}{\left[\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}\right]^{2\gamma}},$$

which can be condensed into

$$\frac{1}{Z_{\mathbf{s}}^{d}} \frac{\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{1-\gamma} \frac{\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}} - \gamma\left[\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}\right]^{-1}\left(1-\tau_j\right)^{\frac{1}{1-\gamma}}\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}{\left[\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}\right]^{\gamma}} =$$

$$\frac{\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{1-\gamma} \frac{\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}} - \gamma\left(\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}\right)^{-1}\left(1-\tau_j\right)^{\frac{1}{1-\gamma}}\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}} =$$

$$\frac{\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{1-\gamma}\left[\frac{\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}} - \gamma\frac{\left(1-\tau_j\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}\right] =$$

$$\frac{1}{1-\gamma}\left[\frac{\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}} - \gamma\frac{\left(1-\tau_j\right)^{\frac{1}{1-\gamma}}\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}\right] = \delta_{j,\mathbf{s}}.$$

Now, observe that the output share or sales share of firm j is given by

$$s_{Y,j}{}^{d} = \frac{\left(1-\tau_j\right)^{\frac{\gamma}{1-\gamma}}\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}(1-\tau_i)^{\frac{\gamma}{1-\gamma}}\left(a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}, \tag{A.24}$$

and its input share is given by

$$s_{L,j}{}^{d} = \frac{\left(1-\tau_j\right)^{\frac{1}{1-\gamma}}\left(a_j e^{\tilde{x}_j}\right)^{\frac{1}{1-\gamma}}}{\sum_{i=1}^{N}\left((1-\tau_i)a_i e^{\tilde{x}_i}\right)^{\frac{1}{1-\gamma}}}, \tag{A.25}$$

where the dependence of $s_{L,j}{}^{d}$ and $s_{Y,j}{}^{d}$ upon the aggregate state $\mathbf{s}$ is suppressed to economize on notation.

Thus, we finally obtain that

$$\delta_{j,\mathbf{s}} = s_{Y,j}{}^{d} + \frac{\gamma}{1-\gamma}\left(s_{Y,j}{}^{d} - s_{L,j}{}^{d}\right). \tag{A.26}$$