The Role of Granularity in the Variance and Tail Probability of Aggregate Output
(Revised)

Yoshiyuki ARATA
RIETI

The role of granularity in the variance and tail probability of aggregate output

Yoshiyuki Arata (RIETI)*

Abstract

According to recent macroeconomic studies, microeconomic shocks can generate aggregate fluctuations (the granular hypothesis). I test this idea by studying the variance and tail probability of aggregate output induced by microeconomic shocks. I show that both of these are determined by the size of the largest firm: The slow decay of the aggregate variance found by Gabaix (2011) is caused by the asymptotic behavior of the size of the largest firm. The tail probability of aggregate output is essentially equivalent to the probability of a large shock hitting the largest firm. I then apply these results to firm-level data in Japan. I find that microeconomic shocks would generate only small fluctuations, and not a large deviation, in aggregate output. Thus, the empirical granularity in Japan (i.e., the size of the largest firm) is not large enough to cause substantial aggregate fluctuations.

Keywords: Granular hypothesis; Aggregate fluctuations; Variance; Tail probability
JEL classifications: E32, E23, D57

The role of granularity in the variance and tail probability of aggregate output

Yoshiyuki Arata[*]

Research Institute of Economy, Trade and Industry[†]

Version: 1.0.2

Date: December 20, 2021

Abstract

According to recent macroeconomic studies, microeconomic shocks can generate aggregate fluctuations (the granular hypothesis). I test this idea by studying the variance and tail probability of aggregate output induced by microeconomic shocks. I show that both of these are determined by the size of the largest firm: The slow decay of the aggregate variance found by Gabaix (2011) is caused by the asymptotic behavior of the size of the largest firm. The tail probability of aggregate output is essentially equivalent to the probability of a large shock hitting the largest firm. I then apply these results to firm-level data in Japan. I find that microeconomic shocks would generate only small fluctuations, and not a large deviation, in aggregate output. Thus, the empirical granularity in Japan (i.e., the size of the largest firm) is not large enough to cause substantial aggregate fluctuations.

Keywords: Granular hypothesis; Aggregate fluctuations; Variance; Tail probability

JEL codes: E32, E23, D57

---

[*]E-mail: y.arata0325@gmail.com.

[†]1-3-1, Kasumigaseki Chiyoda-ku, Tokyo, 100-8901 Japan

(a) GDP   **(b)** GDP growth rate
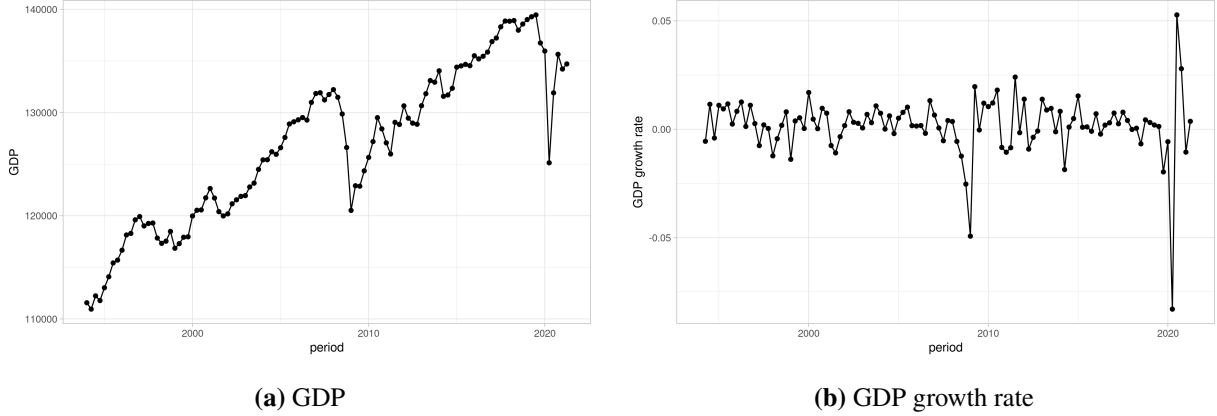
Figure 1: Time series of GDP and GDP growth rates in Japan from 1994Q1 to 2021Q2. The GDP time series are taken from the OECD database. The GDP growth rate g_t is defined as the log difference of the GDP time series, that is, $g_t := \log(\text{GDP}_t) - \log(\text{GDP}_{t-1})$.

1   Introduction

Recent macroeconomic studies (e.g., Gabaix (2011)) argue that microeconomic shocks to firms can drive aggregate fluctuations. This idea is called the granular hypothesis, which states that if firm sizes are highly heterogeneous, microeconomic shocks to giant firms are non-negligible even at the aggregate level. Indeed, high heterogeneity in firm sizes is a stylized fact; for example, in Japan, firms with sales less than 1 billion yen account for 96.8% of all firms,[1] while the sales of Toyota, the largest firm in Japan, are 12.2 trillion yen. Theoretically, previous studies (e.g., Acemoglu et al. (2012); Acemoglu et al. (2017); Baqaee and Farhi (2019)) analyze the coefficients of the expansion of aggregate output $Z := f(\epsilon)$ around its steady state:

$$Z = \sum_i \frac{\partial f(\epsilon)}{\partial \epsilon_i}\big|_{\epsilon=0}\epsilon_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f(\epsilon)}{\partial \epsilon_i \partial \epsilon_j}\big|_{\epsilon=0}\epsilon_i\epsilon_j + \text{higher-order terms} \tag{1}$$

where $\epsilon := (\epsilon_i, ..., \epsilon_n)$ are microeconomic shocks to n firms and f is a mapping from microeconomic shocks to aggregate output. In particular, the coefficients in Eq.(1) are closely related to the firm size and thus highly heterogeneous across firms. For this reason, Eq.(1) and the heterogeneity of firm size are viewed as a foundation for the granular view of macroeconomic fluctuations.

However, previous studies do not fully characterize the *distribution* of aggregate output implied by Eq.(1). Does the distribution of aggregate output converge to a Gaussian distribution as the number of firms increases? What are the variance and tail probability of aggregate output, given Eq.(1) and the empirical heterogeneity of firm sizes? Are the variance and tail probability of aggregate output large enough to explain their empirical counterparts (see **Figure 1** and **Figure 2**)? To answer these questions, we need to study not only the coefficients in Eq.(1) but also the distribution of aggregate output.

---

[1]Source: Economic Census for Business Activity in 2016.
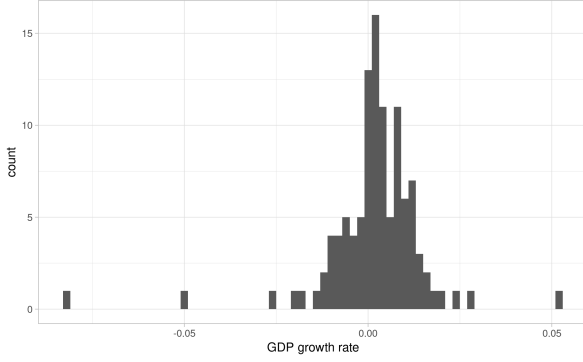
(a) Histogram of the GDP growth rate
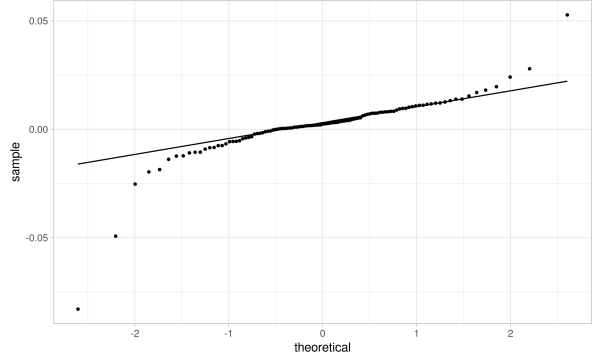
(b) QQ plot of the GDP growth rate

Figure 2: Histogram and QQ plot. The GDP growth rate is the same as in Figure **Figure 1**.

This paper characterizes the distribution of aggregate output in Eq.(1) and tests the empirical validity of the granular hypothesis. Starting with Eq.(1), I characterize the variance and (the upper bound of) tail probability of aggregate output in terms of the heterogeneity of firm size. I show that the size of the largest firm is the key to the distribution of aggregate output. I provide an explicit formula of how the variance and tail probability of aggregate output depend on the size of the largest firm. Then, I apply this result to Japanese firm-level data to quantify micro-originated aggregate fluctuations. I find that the aggregate variance induced by microeconomic shocks is non-negligible, while the tail probability of aggregate output induced by microeconomic shocks is negligible. That is, the empirical heterogeneity of firm size in Japan is not large enough to cause a large deviation in aggregate output.

My theoretical analysis consists of two parts: Sections 3 and 4. Section 3 considers the first-order terms in Eq.(1), where the coefficients are given by Domar weights $w := (w_1, ..., w_n)$.[2] I start with the result in Gabaix (2011), which states that the aggregate variance decays slowly as $n \to \infty$ if the firm size distribution has a Pareto tail.[3] My first question is: why does this slow decay occur? I show that the existence of the largest firm causes this slow decay. More precisely, I show that the fraction of the aggregate variance attributable to the largest firm (i.e., $\|w\|_\infty^2 / \|w\|_2^2$, where $\|w\|_\infty$ and $\|w\|_2$ are ℓ^∞ and ℓ^2-norms of w) does not converge to 0 as $n \to \infty$. That is, the largest firm accounts for a significant fraction of the aggregate variance, even when the number of firms considered is large. In particular, I show that the slow decay rate of the aggregate variance is equivalent to that of the (squared) Domar weight of the largest firm.

The result that $\|w\|_\infty^2 / \|w\|_2^2 \not\to 0$ as $n \to \infty$ has another implication for the central limit theorem (CLT) for aggregate output. I use Theorem 1(c) in Acemoglu et al. (2012): if $\|w\|_\infty^2 / \|w\|_2^2 \not\to 0$ as $n \to \infty$, the CLT does not hold. Combining it with my result, I show that if the firm size distribution has a Pareto

---

[2]Firm i's Domar weight, w_i, is defined by firm i's sales divided by GDP.

[3]The Pareto tail of the firm size distribution is called Zipf's law. See, for example, Axtell (2001) and Gabaix (2009). For the mechanism of Zipf's law and related topics, see, for example, Sornette (2006) and Saichev et al. (2009).

3

tail, the CLT does not hold. The key logic is again the size of the largest firm. If the firm size distribution has a Pareto law, the largest firm is still dominant at the limit of $n \to \infty$. Thus, a shock to the largest firm cannot be cancelled out by shocks to other firms.

Moreover, the size of the largest firm is the key to the tail probability of aggregate output. I show that if $\|w\|_\infty^2/\|w\|_2^2$ is positive, the tail probability of aggregate output deviates from a Gaussian tail and has a fatter tail. In particular, the tail probability of aggregate output is essentially equivalent to the probability of a large shock hitting the largest firm. Intuitively, this result means that a large deviation in aggregate output is driven by a single large shock to the largest firm.

Section 4 considers the second-order terms in Eq.(1). In particular, I use Corollary 1 in Baqaee and Farhi (2019) as an example of the second-order terms. I show that, as in the first-order terms, the two norms of the Domar weights (i.e., $\|w\|_\infty$ and $\|w\|_2$) characterize the distribution of aggregate output. The aggregate variance is proportional to $\|w\|_2^2$, and the tail probability of aggregate output is controlled by $\|w\|_\infty$. The logic is similar to the case of the first-order terms. Owing to its size, a shock to the largest firm dominates shocks to other firms. Thus, the tail probability of aggregate output is essentially equal to the probability of a large shock hitting the largest firm. I show that in both of the first and second-order terms, the size of the largest firm determines the distribution of aggregate output.

Finally, I test the empirical validity of the granular hypothesis by applying the results given in Sections 3 and 4 to Japanese firm-level data. I calculate the empirical values of $\|w\|_\infty$ and $\|w\|_2$ and test whether the variance and tail probability of aggregate output induced by microeconomic shocks are large enough to explain the empirical counterparts, that is, those of the GDP growth rate. My findings are twofold. First, the aggregate variance induced by microeconomic shocks is non-negligible. That is, the contribution of microeconomic shocks to the aggregate variance is economically significant, as suggested by the granular hypothesis. Second, under the assumption that microeconomic shocks follows a Laplace distribution,[4] the tail probability of aggregate output induced by microeconomic shocks is negligible. To summarize, given the empirical heterogeneity of firm size, microeconomic shocks would cause only small fluctuations, and not large deviations, in aggregate output.

Related literature

This paper belongs to the literature on the micro origin of macroeconomic fluctuations (see Carvalho (2014) and Carvalho and Tahbaz-Salehi (2019) for a survey). This literature has two fundamental papers: Gabaix (2011) analyzes the decay rate of the aggregate variance and shows that if the firm size distribution has

---

[4] This assumption is also used in the empirical exercise in Acemoglu et al. (2017). In empirical literature, it is well known that the distribution of the growth rate of firm size is close to a Laplace distribution (e.g., Coad (2009); Dosi et al. (2017); Bottazzi and Secchi (2006); Arata (2019)). Since the firm-level TFP is usually estimated from firms' sales revenue (i.e., TFPR), we can assume that the distribution of TFP growth rates inherits the Laplace shape of the distribution of the growth rate of firm size.

a Pareto tail, the aggregate variance decays slowly as $n \to \infty$. Acemoglu et al. (2012) analyze the granular hypothesis by explicitly considering input-output linkages and provide a condition for the convergence of the distribution of the aggregate output to Gaussian.

I generalizes their results in Section 3. As mentioned above, the ratio $\|w\|_\infty^2 / \|w\|_2^2$ does not converge to 0 but to some positive constant if the firm size distribution has a Pareto tail. Since the aggregate variance is proportional to $\|w\|_2^2$, my analysis reveals that the slow decay of the aggregate variance in Gabaix (2011) is caused by that of $\|w\|_\infty$ (i.e., the Domar weight of the largest firm). The asymptotic behavior of the ratio $\|w\|_\infty^2 / \|w\|_2^2$ is also crucial in Acemoglu et al. (2012). Combining Theorem 1(c) in Acemoglu et al. (2012), my analysis shows that the CLT fails due to the emergence of an extremely large firm, which is expected from the Pareto distribution of firm size.

Subsequent papers extend Gabaix (2011) and Acemoglu et al. (2012) in various directions.[5] My paper is closely related to Acemoglu et al. (2017), which focuses on the tail probability of aggregate output instead of the aggregate variance. As seen in **Figure 2**, the distribution tail of aggregate output deviates from a Gaussian.[6] To explain this empirical fact, Acemoglu et al. (2017) shows that in the tail region, the convergence to Gaussian becomes very slow if the firm size distribution has a Pareto tail. I analyze the same problem but fix n to a constant;[7] that is, I give non-asymptotic results of the tail probability of aggregate output. This is a preferred approach in an empirical analysis because the number of firms in data is fixed and finite. Indeed, using Japanese firm-level data, I find that the tail probability of aggregate output induced by microeconomic shocks is negligible. In contrast to the conclusion in Acemoglu et al. (2017), microeconomic shocks cannot explain the observed large deviation in aggregate output.

As another important extension, Baqaee and Farhi (2019) analyze the second-order terms in Eq.(1) and provide an analytical expression of their coefficients. In Section 4, I use Corollary 1 in Baqaee and Farhi (2019) and complement their result by giving the upper bounds of the variance and tail probability of aggregate output. I show that, as in the case of the first-order terms, the two norms of the Domar weights

---

[5]Some studies extend the model to an inefficient economy with exogenous wedges (Jones (2011); Jones (2013); Baqaee and Farhi (2020b)). In particular, Bigio and La'o (2020), Su (2019), Luo (2020), Altinoglu (2020), and Reischer (2019)) analyze inefficiencies related to financial frictions. As an application, Liu (2019) analyzes the effect of policy intervention in an inefficient economy. Another extension considers the extensive margin, including the firm's entry/exit and rewiring the input-output linkages (Grassi (2017); Baqaee (2018); Acemoglu and Tahbaz-Salehi (2020); Baqaee and Farhi (2020a); Burstein et al. (2020); Acemoglu and Azar (2020); Taschereau-Dumouchel (2020); Oberfield (2018); Elliott et al. (2020); Tintelnot et al. (2019); Huneeus (2018)). For example, Baqaee (2018) and Baqaee and Farhi (2020a) show that an extensive margin amplifies the propagation effect of microeconomic shocks in an input-output network.

[6]The departure from the Gaussian has been documented in empirical literature. To approximate the distribution of the GDP growth rate, Fagiolo et al. (2008) use an exponential power distribution, and Cúrdia et al. (2014) and Clark and Ravazzolo (2015) use student's t distribution. In another literature, Adrian et al. (2019) focus on the evolution of the GDP growth rate distribution, in which a skewed student's t distribution is used.

[7]For the difference from Acemoglu et al. (2017), see also footnote 13.

(i.e., $\|w\|_\infty$ and $\|w\|_2$) characterize the distribution of aggregate output. My analysis shows that the same logic as in the case of the first-order terms holds: the Pareto distribution of firm size leads to the emergence of an extremely large firm, and a shock to such a firm dominates shocks to other firms.

My paper is also related to the empirical literature on the granular hypothesis. Carvalho (2010), Carvalho and Gabaix (2013), and Stella (2015) calculate the granular residual, which is defined as the variation of the sizes of large firms, and finds the high correlation between the granular residual and the GDP growth rate. Di Giovanni and Levchenko (2012), Di Giovanni et al. (2014), Di Giovanni et al. (2019), and Yeh (2019) decompose the variance of aggregate output into common aggregate shocks and individual shocks and show that individual shocks are an important source of the aggregate variance. Foerster et al. (2011), Atalay (2017), and Atalay et al. (2018) filter the time series of aggregate data with a structural factor model and extract underlying microeconomic shocks. Some recent studies (Magerman et al. (2017); Herskovic et al. (2020); Miranda-Pinto (2021)) analyze a firm/sector-level input-output network to measure the relevance of the granular hypothesis. My analysis introduces another method to this literature. Given the empirical values of $\|w\|_\infty$ and $\|w\|_2$ in an economy, my analysis shows how to calculate the variance and tail probability of aggregate output induced by microeconomic shocks. Thus, by comparing them to their empirical counterparts, one can test the empirical relevance of the granular hypothesis.

Outline of this paper

This paper is organized as follows. In Section 2, I review a multi-sector model, which forms the basis of the analysis. In Section 3, I study the distribution of aggregate output when it is given by the first-order terms in Eq.(1). In Section 4, I study the distribution of aggregate output when it is given by the second-order terms in Eq.(1). In Section 5, I apply these results to Japanese firm-level data. In Section 6, I conclude the paper. In the Appendix, I provide the proofs of the propositions.

Notation

$\sum_i$ denotes $\sum_{i=1}^n$. C and c denote constants, which may change from line to line. σ_X^2 denotes the variance of the random variable X. For a sequence (X_n), I write $X_n \overset{\text{a.s.}}{\to} (\overset{P}{\to}, \overset{d}{\to})X$ if (X_n) converges almost surely (in probability, in distribution) to a random variable X as $n \to \infty$. I write $X_n \sim a_n X$ if $X_n/a_n \overset{d}{\to} X$. I write that X_n increases or decreases (or decays) at the rate of a_n if $X_n \sim a_n X$ for some random variable X, which is independent of n.

2 Multi-Sector Model

I review a multi-sector model and its coefficients in Eq.(1). Consider a static economy that consists of n competitive firms. Each firm produces a distinct product using labor and intermediate goods from other

firms. Specifically, firm i produces output y_i by employing constant returns to scale production technology:

$$y_i = z_i f_i(x_{i1}, ..., x_{in}, l_i)$$

where z_i is the productivity shock to firm i, and x_{ij} and l_i are the amounts of good j and labor used in the production of firm i, respectively. Without loss of generality, assume that $z_i = 1$ for all i at the steady state. Let $\epsilon_i := \log(z_i)$ and let s_i be the sales revenue of firm i. Suppose that the economy is inhabited by a representative household endowed with one unit of labor. The representative household supplies its labor inelastically and maximizes their utility $u(c_1, ..., c_n)$, which is homogeneous of degree 1. Finally, assume that the economy is in a competitive equilibrium. Let $Z := \log(\mathrm{GDP}/\overline{\mathrm{GDP}})$, where $\overline{\mathrm{GDP}}$ is the GDP at the steady state. Let w_i denote the Domar weight of firm i, that is, $w_i := s_i/\mathrm{GDP}$, where s_i is i's sales.

Under this setting, Hulten's theorem (Hulten (1978)) states that the coefficient of the first-order terms in Eq.(1) is given by

$$\frac{\partial f(\epsilon)}{\partial \epsilon_i}\Big|_{\epsilon=0} = w_i$$

Thus, the Domar weights are the sufficient statistics for aggregate fluctuations.

The model in Acemoglu et al. (2012) and Acemoglu et al. (2017) gives a concrete example of Hulten's theorem. Assume further a log utility function and Cobb-Douglas production function with a common labor share b across all firms. Then, the coefficient of the first-order terms in Eq.(1) is given by

$$w_i = \frac{s_i}{b \sum_i s_i}, \tag{2}$$

and the higher-order terms are equal to 0. This result implies that aggregate output is the weighted sum of productivity shocks with weight $w_1, ..., w_n$.

Extending this model to CES production technology, Baqaee and Farhi (2019) studies the higher-order terms in Eq.(1). In particular, Corollary 1 in Baqaee and Farhi (2019) gives the coefficient of the second-order terms in a simplified setting. More precisely, if productivity shocks $\epsilon_1, ..., \epsilon_n$ are factor-augmented ones, and the elasticity of substitution is common across all firms (denoted by θ), then

$$\frac{\partial^2 f(\epsilon)}{\partial \epsilon_i \partial \epsilon_j}\Big|_{\epsilon=0} = (\theta - 1)w_i^*(1_{\{i=j\}} - w_j^*) \tag{3}$$

where w_i^* is i's labor cost divided by GDP and can be seen as the Domar weight in this economy.[8] Similar to Hulten's theorem, w_i^* is a sufficient statistic for aggregate fluctuations. In Sections 3 and 4, I study the distribution of Z with the coefficients given by Eq.(2) and Eq.(3), respectively.

3 First-Order Terms

Consider the first-order terms in Eq.(1). Throughout this section, I define Z as follows:

$$Z := \sum_i w_i \epsilon_i, \tag{4}$$

---

[8]Note that the sum of w_i^* is equal to 1, which is independent of $\epsilon_1, ..., \epsilon_n$.

where $w_1, ... w_n$ are the Domar weights and $\epsilon_1, ..., \epsilon_n$ are iid microeconomic shocks with mean 0 and variance σ_ϵ^2. First, I consider the variance of Z in Section 3.1. I show that the asymptotic behavior of the variance of Z is determined by that of the largest w_i. Second, I consider the tail probability of Z in Section 3.2. I give the upper bound of the tail probability of Z with the number of firms fixed.

3.1 Variance

3.1.1 Size of the largest firm

Suppose that Eq.(2) holds and sales $s_1, ..., s_n$ are independently drawn from a common distribution. Let σ_Z^2 denote the variance of Z. Since microeconomic shocks are iid random variables, σ_Z^2 can be represented as follows:

$$\sigma_Z^2 := \sigma_\epsilon^2 \|w\|_2^2 = \frac{\sigma_\epsilon^2}{b^2} \frac{\sum_i s_i^2}{(\sum_i s_i)^2}$$

where $\|w\|_2^2 := \sum_i w_i^2$.

How does σ_Z^2 depend on the heterogeneity of $s_1, ..., s_n$? For example, consider the homogeneous case, where s_i is equal to some constant for all i. In this case,

$$\sigma_Z^2 = \frac{\sigma_\epsilon^2}{b^2} \frac{1}{n}$$

that is, σ_Z^2 decays at the rate of n^{-1}. Because each term in Eq.(4) has the same weight, microeconomic shocks cancel each other out, resulting in the rapid decay of σ_Z^2. In contrast, when $s_1, ..., s_n$ are highly heterogeneous, Gabaix (2011) obtains the following result:

Theorem 3.1 (Proposition 2 in Gabaix (2011)). *Let $\alpha \geq 1$. Suppose that there exists some positive x^* such that*

$$P(s_i > x) = Kx^{-\alpha} \text{ for } x \geq x^*,$$

where K is a positive constant. Then,

$$\sigma_Z^2 \sim \frac{v_\alpha}{n^{2-2/\alpha}} \sigma_\epsilon^2 \quad \text{for } 1 < \alpha < 2$$

$$\sim \frac{c}{n} \sigma_\epsilon^2 \quad\quad \text{for } \alpha > 2$$

where v_α is a non-degenerate random variable, independent of n.[9]

Theorem 3.1 implies that when the heterogeneity is low (i.e., $\alpha \geq 2$), σ_Z^2 decays at the same rate as in the homogeneous case. That is, because of the low heterogeneity, microeconomic shocks cancel each other out as in the homogeneous case. In contrast, for high heterogeneity (i.e., $\alpha < 2$), this averaging effect ceases to work, and σ_Z^2 decays more slowly than n^{-1}.

---

[9]The derivation of the decay rate for $\alpha = 1, 2$ requires careful consideration. However, in an empirical analysis, α is estimated to be in the range of $1 < \alpha < 2$, and the cases of $\alpha = 1, 2$ can be interpreted as the limiting cases. Hence, I remove the cases of $\alpha = 1, 2$ here.

Why does the averaging effect cease to work for $\alpha < 2$? I now show that this is due to the existence of the largest firm. Let us consider the fraction of the variance attributable to shocks to the largest firm.

Definition 3.1. Let $\sigma^2_{\max}$ denote the variance contribution from shocks to the largest firm:

$$\sigma^2_{\max} := \frac{\sigma^2_\epsilon}{b^2} \frac{s^2_{\max}}{(\sum_i s_i)^2} \tag{5}$$

where $s_{\max} := \max_i(s_1, ..., s_n)$.

In the homogeneous case, $\sigma^2_{\max}$ is equal to $\frac{\sigma^2_\epsilon}{b^2} \frac{1}{n^2}$. Since the decay rate of σ^2_Z is equal to n^{-1}, this means that $\sigma^2_{\max}$ decays more rapidly than σ^2_Z. Thus, the contribution of $\sigma^2_{\max}$ to σ^2_Z is negligible for a sufficiently large n.

Note that in the homogeneous case, the decay rate of $\sigma^2_{\max}$ is determined solely by the increasing rate of $(\sum_i s_i)^2$ because $s^2_{\max}$ is constant. When sales are heterogeneous, the decay rate of $\sigma^2_{\max}$ also depends on the increasing rate of $s^2_{\max}$.[10] I give two examples to show that the increasing rate of $s^2_{\max}$ depends on the distribution tail of $s^2_1, ..., s^2_n$. Example 3.1 (Example 3.2) considers the light-tailed (heavy-tailed) distribution for $s^2_1, ..., s^2_n$.

Example 3.1. Suppose that the distribution of s^2_i has an exponential tail:[11] for all i,

$$P(s^2_i > x) = e^{-x} \text{ for } x > 0.$$

Since $s_1, ..., s_n$ are independent of each other,

$$P(s^2_{\max} \le x) = P(s^2_1 \le x, ..., s^2_n \le x) = (1 - e^{-x})^n$$

and its density function $d_n(x)$ is given by $d_n(x) = n(1 - e^{-x})^{n-1}e^{-x}$. By taking the derivative of d_n and setting it to 0, one obtains that the mode of d_n is equal to $\log n$. Thus, the typical value of $s^2_{\max}$ increases at the rate of $\log n$. More formally, for $x \in \mathbb{R}$,

$$P(s^2_{\max} - \log n \le x) = P^n(s^2_i \le x + \log n)$$
$$= (1 - n^{-1}e^{-x})^n \to \exp(-e^{-x}) \text{ as } n \to \infty$$

Thus, $s^2_{\max} - \log n$ converges in distribution to a non-degenerate random variable, independent of n.

---

[10]One can show that $s^2_{\max} \to \infty$ a.s. Indeed, for each $n \in \mathbb{N}$ and $x \in \mathbb{R}$,

$$P(s^2_{\max} \le x) = P(s^2_1 \le x, ..., s^2_n \le x) = P^n(s^2_1 \le x)$$

Thus, if $P(s^2_1 \le x) < 1$ for all $x < \infty$, then $P(s^2_{\max} \le x) \to 0$ for all $x < \infty$. This means that $s^2_{\max} \xrightarrow{P} \infty$. Since $s^2_{\max}$ is non-decreasing as $n \to \infty$, one obtains $s^2_{\max} \xrightarrow{a.s.} \infty$.

[11]This case is closely related to the one in which the distribution of s_i is Gaussian. If s_i follows a standard Gaussian distribution, we obtain by Mills' ratio (see Example 3.3)

$$P(s^2_i > x) = P(s_i > \sqrt{x}) \le e^{-x/2}$$

for a large x. The tail probability of s^2_i is upper bounded by the exponential tail. Thus, the case considered here corresponds to the case in which the heterogeneity of firm size is low.

Compare the increasing rates of $s_{\max}^2$ and $\sum_i s_i^2$ when the distribution of s_i^2 has an exponential tail. Since the law of large numbers implies that $\sum_i s_i^2$ increases at the rate of n, $s_{\max}^2$ increases very slowly compared to $\sum_i s_i^2$. Intuitively, when the distribution of s_i^2 has an exponential tail, the sizes of $s_1^2, ..., s_n^2$ are similar to each other. Since there is no extremely large one among $s_1^2, ..., s_n^2$, the increasing rate of $s_{\max}^2$ is very slow.

Example 3.2. Suppose that the distribution of s_i^2 has a Pareto tail: for $s_i^2 \geq 1$ and $k > 0$,

$$P(s_i^2 > x) = x^{-k}$$

As in Example 3.1, one has

$$P(s_{\max}^2 \leq x) = (1 - x^{-k})^n$$

I use a relation between exponential and Pareto tails. By letting $y := \log x$, one gets

$$P(\log s_i^2 > y) = P(s_i^2 > x) = x^{-k} = e^{-k \log x} = e^{-ky}$$

This shows that when the distribution of s_i^2 has a Pareto tail, the distribution of $\log s_i^2$ has an exponential tail. By using the same argument as in Example 3.1, one can show that $k \log s_{\max}^2$ increases at the rate of $\log n$. Thus, $s_{\max}^2$ increases at the rate of $n^{1/k}$. More formally, for $x \geq 0$,

$$P(n^{-1/k} s_{\max}^2 \leq x) = P(s_{\max}^2 \leq n^{1/k} x)$$
$$= (1 - n^{-1} x^{-k})^n \to \exp(-x^{-k})$$

Thus, $n^{-1/k} s_{\max}^2$ converges in distribution to a non-degenerate random variable, independent of n.

Example 3.2 shows that $s_{\max}^2$ increases at the rate of $n^{1/k}$. This rapid increase is caused by the high heterogeneity of $s_1^2, ..., s_n^2$ described by the Pareto tail. When the distribution of s_i^2 has a Pareto tail, an extremely large s_i^2 emerges one after another as $n \to \infty$. Thus, $s_{\max}^2$ increases rapidly as $n \to \infty$.

Let us return to $\sigma_{\max}^2$. By focusing on the asymptotic behavior of $s_{\max}^2$ discussed in Examples 3.1 and 3.2, I show the decay rate of $\sigma_{\max}^2$. As discussed below, the following results are the key to the understanding of the slow decay of σ_Z^2 found by Gabaix (2011).

Proposition 3.2. *Suppose that there exists x^* such that for $x > x^*$,*

$$P(s_i^2 > x) = K e^{-\alpha x},$$

where K and α are positive constants. Then

$$\sigma_{\max}^2 \sim c \frac{\log n}{n^2}$$

where c is a constant, independent of n.

Proof. See the Appendix. □

Proposition 3.3. *Suppose that there exist x^* such that for $x > x^*$,*

$$P(s_i > x) = K x^{-\alpha}$$

10

where K is a constant and $\alpha > 1$.[12] *Then*

$$\sigma^2_{\max} \sim c \frac{u_{\alpha/2}}{n^{2-2/\alpha}}$$

where c is a constant and $u_{\alpha/2}$ is a non-degenerate random variable, independent of n.

Proof. See the Appendix. □

We are now in a position to compare the decay rates of $\sigma^2_{\max}$ and σ^2_Z. Propositions 3.2, 3.3 and Theorem 3.1 are summarized as follows:

$$\text{Decay rate of } (\sigma^2_Z, \sigma^2_{\max}) = \begin{cases} (n^{-1}, n^{-2}\log n) & \text{for a light tail} \\ (n^{-1}, n^{-2+2/\alpha}) & \text{for a Pareto tail with } \alpha > 2 \\ (n^{-2+2/\alpha}, n^{-2+2/\alpha}) & \text{for a Pareto tail with } 1 < \alpha < 2 \end{cases} \tag{6}$$

First, consider the case of the light-tailed distribution. As discussed in Example 3.1, the sizes of firm sales are similar to each other: the largest firm is not extremely large in size. Thus, the variance contribution of the largest firm is negligible as $n \to \infty$, that is, $\sigma^2_{\max}$ decays more rapidly than σ^2_Z. A similar intuition applies to the case of a Pareto tail with exponent $\alpha > 2$. Although the heterogeneity of firm sales is higher than that of the light-tailed case, it cannot prevent the averaging effect. Since the averaging effect is dominant, $\sigma^2_{\max}$ decays more rapidly than σ^2_Z.

In contrast to these two cases, both σ^2_Z and $\sigma^2_{\max}$ decay at the same rate in the case of a Pareto tail with exponent $1 < \alpha < 2$. This means that $\sigma^2_{\max}$ is of the same order of magnitude as σ^2_Z, that is, the contribution of the largest firm to the aggregate variance is not negligible even for a large n. Intuitively, because of the high heterogeneity of firm sales, $s^2_{\max}$ increases rapidly as $n \to \infty$. Owing to this rapid increase, $\sigma^2_{\max}$ decays at a slower rate than n^{-1}. Since $\sigma^2_{\max}$ is a part of σ^2_Z, σ^2_Z cannot decay more rapidly than $\sigma^2_{\max}$. Thus, σ^2_Z decay at the same rate as $\sigma^2_{\max}$, which is slower than n^{-1}.

3.1.2 Sum and maximum

To further analyze the relation between σ^2_Z and $\sigma^2_{\max}$, I study the ratio of $\sigma^2_{\max}$ to σ^2_Z.

Definition 3.2.

$$r_{\max} := \frac{\sigma^2_{\max}}{\sigma^2_Z} = \frac{\|w\|^2_\infty}{\|w\|^2_2} = \frac{s^2_{\max}}{\sum_i s^2_i}$$

The ratio $r_{\max}$ represents the fraction of the variance of Z attributable to shocks to the largest firm. For example, consider the homogeneous case, in which all of firm sales are equal to some constant. In this case, $r_{\max}$ is equal to $1/n$, which converges to 0 as $n \to \infty$. In other words, the fraction of the largest firm in the variance of Z becomes negligible for a large n. The next proposition shows that this property holds under a more general condition.

---

[12]For later purpose, the condition is represented in terms of the tail probability of s_i instead of s^2_i.

Proposition 3.4. *Suppose that a random variable s_i has a finite second moment: $Es_i^2 < \infty$. Then,*

$$r_{\max} \overset{a.s.}{\to} 0$$

Proof. See the Appendix. □

Distributions with a finite second moment include both light-tailed and Pareto distributions with exponent $\alpha > 2$ in Eq.(6). For these distributions, $r_{\max}$ converges to 0, consistent with the fact that $\sigma_{\max}^2$ decays more rapidly than σ_Z^2.

In contrast, the next result shows that for distributions with a Pareto tail with $\alpha < 2$, $r_{\max}$ does not converge to 0. In other words, $\sigma_{\max}^2$ accounts for a significant part of σ_Z^2 even for a large n.

Proposition 3.5. *Suppose that there exists x^* such that for $x > x^*$,*

$$P(s_i > x) = Kx^{-\alpha},$$

where K is a constant and $0 < \alpha < 2$. Then, $r_{\max}$ converges in distribution to a non-degenerate random variable as $n \to \infty$. In particular, if $\alpha = 1$ (i.e. Zipf's law), then

$$\lim_{n\to\infty} Er_{\max} \geq \frac{1}{2}$$

Proof. See the Appendix. □

This proposition shows that when Zipf's law holds, $\sigma_{\max}^2$ accounts for more than half of σ_Z^2. This result is the key to the understanding of Eq.(6). Since $\sigma_{\max}^2$ accounts for a significant fraction of σ_Z^2 even for a large n, σ_Z^2 decays at the same rate as $\sigma_{\max}^2$. In addition, Proposition 3.3 shows that $\sigma_{\max}^2$ decays at a slower rate than n^{-1} for a Pareto tail with $\alpha < 2$. Thus, σ_Z^2 decays at a slower rate than n^{-1}. This is the mechanism behind the finding by Gabaix (2011). I conclude that when sales are highly heterogeneous, the size of the largest firm is crucial for the variance of Z.

3.2 Tail probability

Let us return to Eq.(4). For simplicity, I assume that the weights are rearranged in decreasing order (e.g., w_1 is the largest weight). I assume that weights are given, that is, the randomness of Z comes only from microeconomic shocks. Motivated by Proposition 3.5, I assume that the weights satisfy the condition $\lim_{n\to\infty} \|w\|_\infty^2 / \|w\|_2^2 > 0$. I study the implication of this condition for the tail probability of Z.

3.2.1 Failure of the CLT

Recall that σ_Z^2 is a decreasing function of n by Theorem 3.1. This implies that Z converges in probability to 0 as $n \to \infty$. For clarity, I formalize this result as follows.

Proposition 3.6. *Suppose that $\sigma_Z^2 \to 0$. Then,*

$$Z \overset{P}{\to} 0$$

12

Proof. Let $B_n := \{Z_n > \delta\}$ for arbitrary positive number δ. Since $\sigma_Z^2 \to 0$, $PB_n \to 0$ as $n \to \infty$. Thus, Z converges in probability to 0. $\qquad\square$

Note that $\sigma_Z^2 \to 0$ holds in all cases considered in Proposition 3.1. That is, even when the weights are highly heterogeneous, Z converges to 0 as $n \to \infty$.

However, this does not mean that Z converges to 0 as predicted by the CLT. Indeed, Theorem 1 (c) in Acemoglu et al. (2012) shows that if $\lim_{n\to\infty} \|w\|_\infty^2/\|w\|_2^2 > 0$, normalized Z does not converge to a Gaussian distribution.

Theorem 3.7 (Theorem 1(c) in Acemoglu et al. (2012)). *Suppose that $\epsilon_1, ..., \epsilon_n$ are not Gaussian random variables and that $\lim_{n\to\infty} \|w\|_\infty^2/\|w\|_2^2 > 0$. Then, $\frac{1}{\|w\|_2 \sigma_\epsilon} Z$ does not converge to a Gaussian distribution.*

This theorem illustrates the challenge of characterizing the distribution of Z under the condition $\lim_{n\to\infty} \|w\|_\infty^2/\|w\|_2^2 > 0$. One cannot approximate the distribution of Z by a Gaussian distribution, and to the best of my knowledge, there is no convergence result for the distribution of Z. In the following, I characterize the distribution of Z with an arbitrary fixed n.[13] In particular, I show that the tail probability of Z is determined by the largest weight (i.e., $\|w\|_\infty$).

3.2.2 Light-tailed distribution

In general, the distribution of Z in Eq.(4) with fixed n depends on the underlying distribution of microeconomic shocks. Here, I consider a light-tailed distribution as the distribution of microeconomic shocks. More precisely, I assume that the moment generating function of ϵ_i exists:

$$Ee^{\lambda \epsilon_i} < \infty \quad \text{for some } \lambda > 0.$$

This condition means that the tail probability $P(\epsilon_i > x)$ vanishes (at least) exponentially fast as $x \to \infty$. Throughout this paper, I assume that the distribution of microeconomic shocks is symmetric, and thus, the condition also implies a rapid decrease in the left tail.

As the first example of the light-tailed distribution, consider a Gaussian distribution:

---

[13] My analysis of the tail probability of Z is related to Acemoglu et al. (2017), who propose the macroeconomic tail probability defined by

$$R_n(\tau_n) := \frac{\log P(Z < -\tau_n \sigma_Z)}{\log \Phi(-\tau_n)}$$

where Φ is the standard Gaussian distribution and $\tau_n \to \infty$ as $n \to \infty$. $R_n(\tau_n)$ measures how rapid the tail probability of Z decays, compared to that of a standard Gaussian random variable. Note that $R_n(\tau_n)$ considers two limits: $n \to \infty$ and $\tau_n \to \infty$. In contrast, I analyze $P(Z \leq -x)$ directly with a fixed n. This approach has two advantages. First, a Gaussian distribution is not needed as a reference. As discussed in Proposition 3.5 and Theorem 3.7, Z does not converge to a Gaussian distribution as $n \to \infty$. Thus, the Gaussian distribution is no longer an appropriate reference in my case. Second, by fixing n, it is not necessary to choose the increasing rate of τ_n as $n \to \infty$. In general, since $R_n(\tau_n)$ depends on the increasing rate, the choice of the increasing rate causes another difficulty. By fixing n, I can focus on the limit of $x \to \infty$ only (i.e., the tail probability).

Example 3.3. Suppose that $\epsilon_1, ..., \epsilon_n$ are iid Gaussian random variables with variance σ_ϵ^2. The invariance property of a Gaussian distribution yields that Z is Gaussian with $\sigma_Z^2 = \|w\|_2^2 \sigma_\epsilon^2$. Mills' ratio (e.g., Proposition 2.1.2 in Vershynin (2020)) shows that if X is a standardized Gaussian random variable,

$$P(X \geq x) \leq \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ for } x \geq 1.$$

Since Z is symmetric, I obtain

$$P(Z \leq -x) \leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\|w\|_2^2 \sigma_\epsilon^2}\right) \text{ for } x \geq \|w\|_2 \sigma_\epsilon.$$

Example 3.3 shows that if microeconomic shocks are Gaussian, Z shows a Gaussian decay in the tail region, which is controlled by $\|w\|_2^2$. Note that this decay is the same as that predicted by the CLT. As shown below, the Gaussian decay can be considered as a benchmark for the tail probability of Z.

Next, let us consider the tail probability of Z when the distribution tail of microeconomic shocks can be fatter than the Gaussian tail. I assume that the distribution tail of microeconomic shocks is exponentially upper bounded so that the moment generating function of ϵ_i exists. This case includes Example 3.3 as a special case.

Proposition 3.8. *Suppose that there exists $x^* \geq 0$ such that for all $x \geq x^*$,*

$$P(\epsilon_i \leq -x) \leq K \exp(-\beta x)$$

where K and β are positive constants. Then, for all $x \geq 0$,

$$P(Z \leq -x) \leq \exp\left(-c \min\left(\frac{x^2}{\|w\|_2^2}, \frac{x}{\|w\|_\infty}\right)\right)$$

where c is a non-negative constant, independent of n.

Proposition 3.8 shows two decay rates: Gaussian decay for a small x (i.e., for $x \leq \|w\|_2^2/\|w\|_\infty$) and exponential decay for a large x (i.e., for $x \geq \|w\|_2^2/\|w\|_\infty$). If $\|w\|_\infty$ is very small compared to $\|w\|_2^2$, the region of Gaussian decay covers almost all $x \geq 0$. In contrast, suppose that the heterogeneity of the weights is high and that $r_{\max}$ does not converge to 0, as in Proposition 3.5. Since $\|w\|_2^2/\|w\|_\infty \leq 1/r_{\max}^2$ for $\|w\|_\infty \leq 1$, Proposition 3.8 implies that the tail probability of Z deviates from Gaussian. Note that the tail probability of Z is controlled by the weight of the largest firm (i.e., $\|w\|_\infty$). In particular, the exponential decay is the same as that of $P(w_1 \epsilon_1 < -x)$, that is, the tail probability of the component of the largest firm (up to a constant c). This suggests that the tail probability of Z is driven mainly by shocks to the largest firm.

This finding is consistent with the interpretation given in Section 3.1: because of the high heterogeneity of firm size, the largest firm is dominant in an economy and determines the distribution properties of aggregate output even when the number of firms is large. The analysis in Section 3.2.3 below shows that this interpretation still holds when we consider the distribution of microeconomic shocks with a tail fatter than exponential.

3.2.3 Subexponential distribution

Section 3.2.2 considers the light-tailed distribution, which has a finite moment generating function. Here, I consider the distribution whose tail is fatter than exponential so that the moment generating function does not exist. More precisely, I assume that for all $\lambda \in \mathbb{R}$,

$$Ee^{\lambda \epsilon} \text{ not exist.}$$

An example of distributions satisfying this condition is a Pareto distribution. As shown below, the Pareto distribution has an important property regarding the tail probability of the sum of independent random variables.

Example 3.4. Let X_1, X_2 be iid random variables on $\mathbb{R}^+$ such that $P(X_i > x) := x^{-\beta}$ for $\beta > 0$. Consider the tail probability of $X_1 + X_2$, which can be decomposed into three parts using a criterion function $h(x)$:

$$\begin{aligned} P(X_1 + X_2 > x) = & P(X_1 + X_2 > x, X_1 \leq h(x)) + P(X_1 + X_2 > x, X_2 \leq h(x)) \\ & + P(X_1 + X_2 > x, X_1 > h(x), X_2 > h(x)) \end{aligned} \tag{7}$$

Here, I take $h(x) := \frac{1}{2}x^{2/3}$. Note that $h(x)$ is small compared to x, especially when x is large. I analyze which term on the right-hand side of Eq.(7) is the main contributor to the tail probability of the sum.

Consider the third term on the right-hand side of Eq.(7), which represents the probability that both X_1 and X_2 are large enough to exceed $h(x)$ and their sum exceeds x. This event is unlikely to occur for a large value of x:

$$P(X_1 + X_2 > x, X_1 > h(x), X_2 > h(x)) \leq P^2(X_1 > h(x)) = o(P(X_1 > x))$$

This probability decays more rapidly than the probability that X_1 exceeds x. Intuitively, when the sum exceeds a large value, it is unlikely that both X_1 and X_2 are large.

Next, consider the first term on the right-hand side of Eq.(7). The definition of h yields

$$P(X_1 + X_2 > x, X_1 \leq h(x)) \leq P(X_2 > x - h(x)) = x^{-\beta}\left(1 - \frac{1}{2}x^{-\frac{1}{3}}\right)^{-\beta} \sim P(X_1 > x)$$

Since X_1 and X_2 are iid random variables, I obtain the same upper bound for the second term.

Thus, by combining these inequalities, I obtain

$$\limsup_{x \to \infty} \frac{P(X_1 + X_2 > x)}{P(X_1 > x)} \leq 2 \tag{8}$$

The opposite direction of Eq.(8) is obvious from the non-negativeness of X_1, X_2. Thus, I obtain

$$P(X_1 + X_2 > x) \sim 2P(X_1 > x) \text{ as } x \to \infty \tag{9}$$

Eq.(9) has a more straightforward interpretation. Note that for any positive integer n,[14]

$$P(\max(X_1, ..., X_n) > x) \sim nP(X > x) \text{ as } x \to \infty,$$

where $X_1, ..., X_n$ are n iid. random variables with the same distribution as X. Thus, Eq.(9) means that the probability of the sum exceeding x is asymptotically equal to the probability of the maximum exceeding x:

$$P(X_1 + X_2 > x) \sim P(\max(X_1, X_2) > x) \text{ as } x \to \infty. \tag{10}$$

In other words, when the sum exceeds a large value x, (at least) one of its components exceeds x. We can ignore the probability that the sum exceeds x while both components are small, which corresponds to the third term on the right-hand side of Eq.(7). This property enables us to approximate the tail probability of the sum using the tail probability of its component.

The property in Eq.(10) is at the heart of the analysis of the tail probability of Z. Indeed, Eq.(10) can be generalized to the case of the weighted sum of iid random variables on $\mathbb{R}$ whose distribution is heavy-tailed.[15] That is, the tail probability of Z is asymptotically equivalent to that of the maximum of the components (i.e., $\max_i(w_i \epsilon_i)$). In the following, I use this property for the calculation of the tail probability of Z. In addition, by specifying the distribution of microeconomic shocks, I clarify the role of the weights in the tail probability of Z.

As examples of heavy-tailed distributions, I consider Pareto and Weibull tails. More precisely, I say that the distribution has a Pareto tail if for $x \geq x^*$,

$$P(\epsilon_i \leq -x) = Kx^{-\beta}, \ \beta > 0 \tag{11}$$

where β and K are some positive constants. Similarly, I say that the distribution has a Weibull tail if for $x \geq x^*$,

$$P(\epsilon_i \leq -x) = Ke^{-\beta x^\tau}, \ \beta > 0, \ 0 < \tau < 1, \tag{12}$$

where β and K are some positive constants. The shape parameter τ controls the heaviness of the tail. In particular, the Weibull tail becomes arbitrarily close to an exponential tail as $\tau \to 1$. Thus, the exponential tail can be considered as the limit of the Weibull tail.

Proposition 3.9. *Suppose that the weight of the largest firm, w_1, is strictly larger than any other weight.*

---

[14]Let $F(x) := P(X \leq x)$. Then,

$$P(\max(X_1, ..., X_n) > x) = (1 - F^n(x))$$
$$= (1 - F(x))(1 + F(x) + F^2(x) + ... + F^{(n-1)}(x)) \sim n(1 - F(x)), \text{ as } x \to \infty.$$

[15]For details, see Appendix 7.1.4.

(i) If the distribution of microeconomic shocks has a Pareto tail, then

$$P(Z \leq -x) \sim \sum_i \left(\frac{w_i}{w_1}\right)^\beta P(w_1\epsilon_1 \leq -x) \ \text{ as } x \to \infty.$$

(ii) If the distribution of microeconomic shocks has a Weibull tail, then

$$P(Z \leq -x) \sim P(w_1\epsilon_1 \leq -x) \ \text{ as } x \to \infty.$$

Proof. See the Appendix. □

The first result in Proposition 3.9 shows that the contribution of each component $w_i\epsilon_i$ to the tail probability of Z is determined by the multiplier $(w_i/w_1)^\beta$. As expected, the smaller the weight, the smaller the impact on the tail probability of Z. Furthermore, the tail probability of Z depends on the tail exponent of the distribution tail of microeconomic shocks. In particular, as β becomes larger (i.e. the tail becomes lighter), only components with large weights contribute to the tail probability of Z.

The second result in Proposition 3.9 has a stronger implication. When the distribution of microeconomic shocks has a Weibull tail, only the component with the largest weight contributes significantly to the tail probability of Z. Compare this finding with Proposition 3.8. Since the exponential tail can be seen as the limit of the Weibull tail, Propositions 3.9 and 3.8 provides a unified view of how the heterogeneity of firm size are related to the tail probability of Z: when the distribution of microeconomic shocks is close to a Laplace, only the size of the largest firm matters for the tail probability of Z.

4   Second-Order Terms

Let us consider the second-order terms in Eq.(1), that is, Z is given by

$$Z := \sum_{i,j} b_{ij}\epsilon_i\epsilon_j$$

where $b_{ij} \in \mathbb{R}$ is a constant and $\epsilon_1, ..., \epsilon_n$ are iid microeconomic shocks. In particular, I use Corollary 1 in Baqaee and Farhi (2019) (see Eq.(3)):

$$b_{ij} := \frac{(\theta-1)}{2} w_i^*(1_{\{i=j\}} - w_j^*).$$

The following analysis characterizes the variance and tail probability of Z in terms of $\|w^*\|_\infty$ and $\|w^*\|_2$.

Before proceeding to the analysis, let us introduce some notations. Let $B := (b_{ij})$ be an $n \times n$ symmetric matrix. Let $\|B\|_{\mathrm{HS}}$ and $\|B\|$ denote the Hilbert-Schmidt norm and the operator norm of B, respectively:

$$\|B\|_{\mathrm{HS}} := \left(\sum_{ij} |b_{ij}|^2\right)^{1/2}, \ \|B\| := \sup\left\{\frac{\|Bv\|_2}{\|v\|_2} : v \in \mathbb{R}^n, v \neq 0\right\}$$

One can show that they can be represented as $\|B\|_{\mathrm{HS}} = (\sum_i \mu_i^2)^{1/2}$ and $\|B\| := \max(|\mu_1|, ..., |\mu_n|)$, where $(\mu_1, ..., \mu_n)$ are the eigenvalues of B. Let $\mathrm{tr}(B)$ denote the trace of B. The independence of microeconomic

shocks yields that the mean of Z is given by

$$EZ = \sum_{ij} b_{ij} E[\epsilon_i \epsilon_j] = \sigma_\epsilon^2 \text{tr}(B)$$

I start with an example in which microeconomic shocks follow a Gaussian distribution. In contrast to the case considered in Section 3, the distribution of Z is no longer Gaussian, as described below.

Example 4.1. Suppose that microeconomic shocks $\epsilon_1, ..., \epsilon_n$ are iid Gaussian random variables with mean 0 and variance σ_ϵ^2. Let $X_i := \epsilon_i / \sigma_\epsilon$ so that X is a vector of standard Gaussian random variables. By setting $A := -B$, I consider $X^T A X$ instead of $X^T B X$. Indeed, $\text{Var}(X^T B X) = \text{Var}(X^T A X)$ and $P(X^T B X - E[X^T B X] < -x) = P(X^T A X - E[X^T A X] > x)$. Note also that $\|A\| = \|B\|$ and $\|A\|_{\text{HS}} = \|B\|_{\text{HS}}$.

First, I diagonalize matrix A. Since A is a real symmetric matrix, there are an orthogonal matrix P and a diagonal matrix D such that $A = P^T D P$. Thus, by letting $Y := PX$, I have

$$X^T A X = \sum_i \mu_{i,A} Y_i^2$$

where $\mu_{i,A}$ is an eigenvalue of A. Since X follows a standard multivariate Gaussian distribution, $Y_1, ..., Y_n$ are independent standard Gaussian random variables. Thus, the right-hand side of the above equation represents the weighted sum of independent χ^2 random variables with weight $\mu_{i,A}$. Since the mean of $X^T A X$ is given by $\text{tr}(A)$, I have

$$X^T A X - E[X^T A X] = \sum_i \mu_{i,A}(Y_i^2 - 1) \tag{13}$$

This representation by Eq.(13) yields the variance of Z. Indeed,

$$\text{Var}(X^T A X) = \sum_i \mu_{i,A}^2 \text{Var}(Y_i^2) = 2 \sum_i \mu_{i,A}^2 = 2\|A\|_{\text{HS}}^2$$

Since $\epsilon_i = \sigma_\epsilon X_i$,

$$\sigma_Z^2 = 2\sigma_\epsilon^4 \|B\|_{\text{HS}}^2 \tag{14}$$

For the tail probability, I apply Chernoff's method to Eq.(13). Note that the moment generating function of each summand in Eq.(13) is upper bounded as follows:

$$\log E e^{\lambda(Y_i^2 - 1)} = \frac{1}{2}(-\log(1 - 2\lambda) - 2\lambda) \leq \frac{\lambda^2}{1 - 2\lambda}$$

for all $\lambda < 1/2$. Thus, the independence of $Y_1, ..., Y_n$ yields

$$\log E e^{\lambda(X^T A X - E[X^T A X])} = \sum_i \frac{1}{2}(-\log(1 - 2\mu_{i,A}\lambda) - 2\mu_{i,A}\lambda) \leq \frac{\lambda^2 \|A\|_{\text{HS}}^2}{1 - 2\lambda\|A\|}$$

for all $\lambda \in (0, \frac{1}{2\|A\|})$. By applying Lemma 7.4 with $v = 2\|A\|_{\text{HS}}^2$ and $c = 2\|A\|$ and using $\epsilon_i = \sigma_\epsilon X_i$, I have

$$P(Z - EZ < -x) \leq \exp\left(-\frac{x^2}{4(\sigma_\epsilon^4 \|B\|_{\text{HS}}^2 + \sigma_\epsilon^2 \|B\|x)}\right) \quad \text{for all } x > 0. \tag{15}$$

Since $(a + b) \leq 2\max(a, b)$ for $a, b \geq 0$, Eq.(15) can be written as

$$P(Z - EZ < -x) \leq \exp\left(-c\min\left(\frac{x^2}{\sigma_\epsilon^4 \|B\|_{\text{HS}}^2}, \frac{x}{\sigma_\epsilon^2 \|B\|}\right)\right) \tag{16}$$

18

where c is some absolute constant.

Eq.(16) shows that the tail probability of Z deviates from a Gaussian in the tail region, even when Gaussian microeconomic shocks are considered. Intuitively, the quadratic form transforms the light tail of the distribution of microeconomic shocks into a fatter tail of the distribution of Z. Eq.(16) also shows that that the tail probability is characterized by $\|B\|$ and that the deviation from the Gaussian in the tail region matters only when $\|B\|_{\mathrm{HS}}^2/\|B\|$ does not converge to 0 as $n \to \infty$.

Eq.(16) is reminiscent of the upper bound for the tail probability of Z in Proposition 3.8. Indeed, $\|w\|_2$ and $\|w\|_\infty$ in Proposition 3.8 are replaced by $\|B\|_{\mathrm{HS}}$ and $\|B\|$ in Example 4.1. This correspondence becomes further apparent by the next lemma.

Lemma 4.1. *Let $B := \{b_{ij}\}$ and $b_{ij} := \frac{(\theta-1)}{2}w_i^*(1_{\{i=j\}} - w_j^*)$. Then,*
$$\|B\| \leq \frac{|\theta - 1|}{2}\|w^*\|_\infty$$
$$c\|w^*\|_2 \leq \|B\|_{HS} \leq C\|w^*\|_2$$
for some constants c and C.

Proof. See the Appendix. □

Lemma 4.1 implies that if $\|w^*\|_2$ converges to 0 as $n \to \infty$, the convergence rate of $\|B\|_{\mathrm{HS}}$ is the same as that of $\|w^*\|_2$. Since the variance of Z is determined by $\|B\|_{\mathrm{HS}}$, we can apply the same reasoning as in Section 3: If the distribution of $w_1^*, ..., w_n^*$ has a Pareto tail, the convergence rate of the variance of Z becomes slow, and its slow rate is due to the asymptotic behavior of the largest w_i^*.

For the tail probability of Z, by using Lemma 4.1, Eq.(16) can be rewritten as
$$P(Z - EZ \leq -x) \leq \exp\left(-c\min\left(\frac{x^2}{\sigma_\epsilon^4\|w^*\|_2^2}, \frac{x}{\sigma_\epsilon^2\|w^*\|_\infty}\right)\right) \tag{17}$$
Thus, the same interpretation given after Proposition 3.8 can be applied here. That is, $Z - EZ$ exhibits Gaussian decay for a small x and exponential decay for a large x. If $\|w^*\|_\infty$ is very small compared to $\|w^*\|_2^2$, the region of Gaussian decay covers almost all x. On the other hand, since $\|w^*\|_\infty \leq 1$, one has
$$\frac{\|w^*\|_\infty^2}{\|w^*\|_2^2} \leq \frac{\|w^*\|_\infty}{\|w^*\|_2^2}.$$
Thus, if $\|w^*\|_\infty^2/\|w^*\|_2^2$ does not converge to 0 (as in Proposition 3.5), there exists a region in which the probability of Z deviates from the Gaussian decay. In addition, the deviation is controlled by $\|w^*\|_\infty$.

These implications can be extended to a more general condition. First, I give the result about the convergence rate of the variance of Z.

Proposition 4.2. *Suppose that $\epsilon_1, ..., \epsilon_n$ are iid random variables with mean 0 and finite fourth moment. Then,*
$$\sigma_Z^2 = (m_4 - 3m_2^2)\sum_i B_{i,i}^2 + 2m_2^2\|B\|_{HS}^2$$

19

where $m_2 := E(\epsilon_i^2)$ and $m_4 := E(\epsilon_i^4)$. In particular, if $b_{ij} := \frac{(\theta-1)}{2} w_i^* (1_{\{i=j\}} - w_j^*)$, then the convergence rate of σ_Z is equal to that of $\|w^*\|_2$.

Proof. See the Appendix. □

This proposition means that, similar to the case discussed in Section 3, the order of σ_Z is equal to that of $\|w^*\|_2$. Thus, if $\|w^*\|_2$ decays slowly as $n \to \infty$, σ_Z decays slowly as well.

For the tail probability of Z, I consider the Laplace distribution as the underlying distribution of microeconomic shocks.

Proposition 4.3. *Let $B := \{b_{ij}\}$ and $b_{ij} := \frac{(\theta-1)}{2} w_i^* (1_{\{i=j\}} - w_j^*)$. Suppose that $\epsilon_1, ..., \epsilon_n$ are iid random variables with a common Laplace distribution with mean 0. Then,*

$$P(Z - EZ \leq -x) \leq \exp\left(-c \min\left(\frac{x^2}{\|w^*\|_2^2}, \left(\frac{x}{\|w^*\|_\infty}\right)^{\frac{1}{2}}\right)\right)$$

where c is a constant, independent of n.

Proof. See the Appendix. □

As in Section 3.2.2, the tail probability of Z shows Gaussian decay in the central region but deviates from Gaussian in the tail region. In the case of the Laplace distribution, the boundary separating the two regions is given by $(\|w^*\|_2^4/\|w^*\|_\infty)^{1/3}$. Since $\|w^*\|_\infty < 1$, I have

$$\frac{\|w^*\|_2^4}{\|w^*\|_\infty} \leq \frac{\|w^*\|_2^4}{\|w^*\|_\infty^4}$$

Thus, as long as the ratio $\|w^*\|_\infty^2/\|w^*\|_2^2$ converges to some non-zero value (as in Proposition 3.5), there exists a region in which the tail probability of Z deviates from a Gaussian.

My analysis in this section provides a unified view about the role of granularity, which is consistent with that in Section 3. That is, $\|w^*\|_\infty, \|w^*\|_2$, and their ratio are sufficient statistics for the variance and tail probability of aggregate output. When the ratio does not converge to 0 (i.e., the largest firm dominates the economy), the aggregate variance decays slowly, and the tail probability of aggregate output deviates from Gaussian. The size of the largest firm characterizes the distribution properties of aggregate output. Given these theoretical results, the next question is to ask whether the size of the largest firm is sufficiently large to generate substantial variance and tail probability of aggregate output. In the next section, I answer this question using Japanese firm-level data.

5   Empirical Analysis

I apply the results obtained in Sections 3 and 4 to Japanese firm-level data. In Section 5.1, I analyze the time series of the quarterly GDP in Japan. In Section 5.2, I give the summary statistics of Japanese firm-level data. In Section 5.3, given the empirical granularity in Japan, I compare the variance and tail probability of aggregate output induced by microeconomic shocks to their empirical counterparts.

5.1 Aggregate output

I analyze the seasonally adjusted quarterly GDP time series from 1994Q1 to 2021Q2, which is taken from the OECD database. I denote by g_t the log-difference of the GDP time series (referred to as the GDP growth rate in the following), that is, $g_t := \log(\text{GDP}_t) - \log(\text{GDP}_{t-1})$. **Figure 1** shows the times series of the GDP and its growth rate over the sample period. The summary statistics of the GDP growth rate are given in **Table 1**.[16]

n	mean	meadian	s.d.	s.d.(mad)	min	max
110	1.73E-03	2.38E-03	0.01374	0.00803	-0.0830	0.0527

Table 1: Summary statistics of the GDP growth rate. In the table, s.d. (mad) represents the estimate of the standard deviation by the median absolute deviation (i.e., $1.4826 \times \text{med}(|g_t - \text{med}(g_t)|)$). This is a consistent estimator when samples follow a Gaussian.

Next, I analyze the distribution property of the GDP growth rate. **Figure 2** shows the histogram (left panel) and QQ plot (right panel) of the GDP growth rate. If the GDP growth rate is independently drawn from a Gaussian distribution, the QQ plot would lies on the straight line in the right panel. The observed departure from the straight line suggests that the GDP growth rate does not follow a Gaussian distribution, especially in the left-tail region.[17] In other words, the probability of a large negative deviation is higher than predicted by a Gaussian distribution.

Finally, I estimate the left-tail probability of the GDP growth rate, that is, $\widehat{P}(g_t < -x)$. I use two methods widely used in extreme value theory: approximation by the generalized extreme value (GEV) distribution and the method of peaks over threshold (POT). The main idea is the use of the limiting theory of extremes: for the GEV approximation, it tells us that properly normalized extremes converges to a generalized extreme value distribution as $n \to \infty$, and for the POT method, it tells us that extremes over a high threshold converges to a generalized Pareto distribution as $n \to \infty$.[18]

Both estimates of the left-tail probability are given in **Figure 3**, in which the empirical counter cumulative distribution function (CCDF) is also plotted for comparison. This figure shows that both estimates fit well with the empirical CCDF and that their differences are small. Thus, I use the GEV approximation as a main estimate in the following. This estimate suggests that the GDP growth rate has a significant left-tail

---

[16]As a measure of scale, sample standard deviation is greatly influenced by outliers. As seen in **Figure 2**, the empirical GDP growth rates have several large deviations, which can overestimate its standard deviation. To mitigate this concern, I use s.d. (mad) as an alternative measure of scale, which is less dependent on outliers, although the GDP growth rate does not seem to follow Gaussian. The large deviations observed in the samples are used to estimate of the tail probability.

[17]I perform normality tests: Kolmogorov-Smirnov ($D = 0.480$), Anderson-Darling ($A = 5.12$), Cramer-von Mises ($W = 0.867$), Shapiro-Wilk tests ($W = 0.776$), where the test statistics are in parentheses. For all the normality tests, the null hypothesis is rejected at the 1% level.
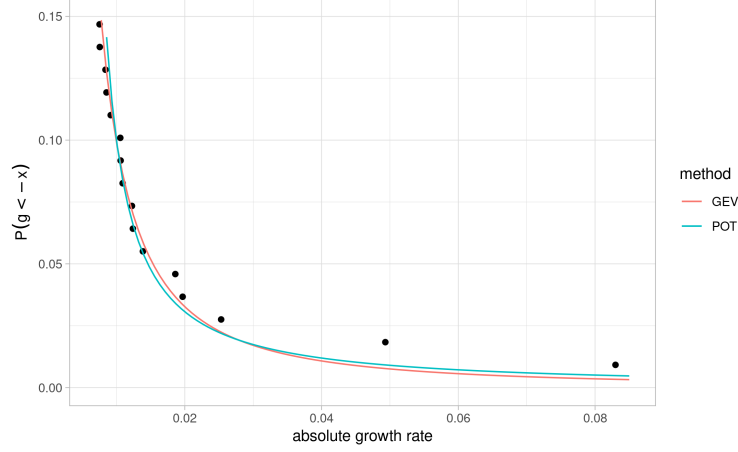
[18]For details, see Appendix 7.2.

Figure 3: Estimation of the left-tail probability. The horizontal axis is the absolute value of the negative GDP growth rate, that is, $|g_t|$. The tail estimates by GEV and POT methods and the empirical CCDF are plotted.

probability; for example, $\widehat{P}(g_t < -0.02)$ is 3.27%. In Section 5.3, I test whether microeconomic shocks can explain this left-tail probability of the GDP growth rate.

5.2 Firm-level data

To measure the empirical granularity, I use firm-level data in Japan in 2017 provided by Tokyo Shoko Research (TSR). This data contains more than one million firms including listed and unlisted firms across all sectors. I exclude firms in sectors of agriculture & forestry, fisheries, finance & insurance, medical & health care, or public service sectors. I also exclude firms whose sales are not available or 0. The total number of firms in my sample is reduced to $1,066,653$.

The summary statistics of annual sales revenue are given in **Table 2**. As is well known in the literature, firms' sales are highly heterogenous and closely follow a Pareto tail. **Figure 4** shows the empirical CCDF of sales in the log-log plot, which is close to the straight line in the tail region. I estimate the slope of this straight line (i.e., the exponent α of the Pareto tail) by Hill's estimate. The estimate is $\widehat{\alpha} = 1.24(0.055)$, which suggests that Zipf's law holds approximately in my sample.

n	mean	meadian	s.d.	s.d.(mad)	min	max
1066653	1143	80	28346	96	0.001	12201443

Table 2: Summary statistics of annual sales revenue. The unit of sales in the table is 1 million yen.

Using firms' sales, I calculate Domar's weights and their norms, which are given in **Table 3**. I set GDP $= 545.897$ trillion yen (GDP in 2017 in Japan) and labor share $= 0.2$.[19] For later purpose, I also give

---

[19]According to Financial Statements Statistics of Corporations by Industry conducted by the Ministry of Finance, the sample average of the labor share in sales fluctuates within the range of $[0.1, 0.2]$ over time. To obtain the upper bound of the micro-originated
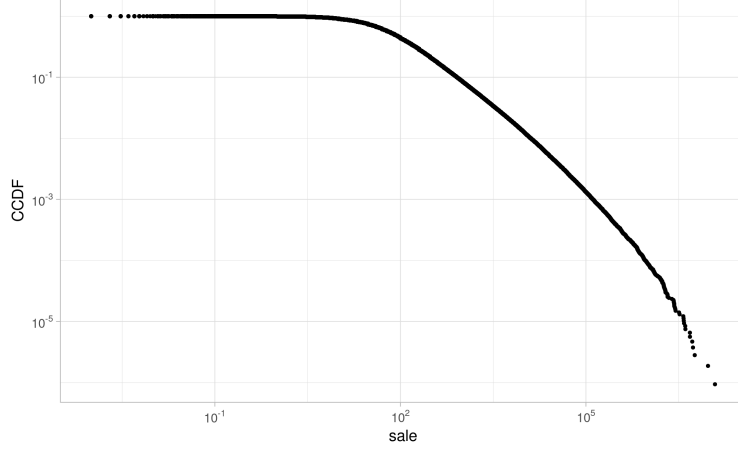
Figure 4: CCDF of firms' sale. The exponent of Pareto's tail is estimated to be $\widehat{\alpha} = 1.24(0.055)$, where the standard error is in parentheses.

$\sqrt{\sum_i B_{i,i}^2}$ and $\|B\|_{\text{HS}}$ in **Table 3**. Recall that these norms are sufficient statistics for the variance and tail probability of aggregate output as discussed in Section 3 and Section 4. Given these norms representing the granularity of the Japanese economy, I test in Section 5.3 below whether microeconomics shocks drive substantial aggregate fluctuations.

$\|w\|_\infty$	$\|w\|_2$	$\|w^*\|_\infty$	$\|w^*\|_2$	$\sqrt{\sum_i B_{i,i}^2}$	$\|B\|_{\text{HS}}$
0.0224	0.0537	0.00447	0.0107	0.00536	0.00536

Table 3: Norms of firms' sales. In the calculation, GDP is set to 546 trillion yen (GDP in 2017 in Japan), and the labor share is set to 0.2.

5.3 Empirical validity of the granular hypothesis

Given the empirical aggregate fluctuations (Section 5.1) and firm-level statistics (Section 5.2) in Japan, we can quantitatively test the granular hypothesis. First, I calculate the empirical counterpart to $r_{\max}$ in Proposition 3.5. I find that $\widehat{r}_{\max} = \|w\|_\infty^2/\|w\|_2^2 = 0.173$; that is, 17.3% of the micro-originated aggregate variance is attributable to the contribution of the largest firm. This implies that the CLT fails because of the high presence of the largest firm, consistent with the granular hypothesis.

Next, I calculate the variance and tail probability of aggregate output induced by microeconomic shocks. Following Gabaix (2011), I assume that the standard deviation of annual productivity shocks is equal to 12%. Since the GDP growth rates given in Section 5.1 are on quarterly basis, I use $\sigma_\epsilon = 6\%$ by assuming that the annual growth rate is the sum of the iid quarterly growth rates. For labor-augmented productivity

---

aggregate fluctuations, I set the labor share in sales equal to 0.2 in my exercise.

shocks (i.e., for the results in Section 4), I assume that $\sigma_\epsilon = 30\%$.[20]

5.3.1 Case of the first-order terms

Consider the aggregate variance when Z is given by Eq.(4). By using the norm in **Table 3**, I obtain

$$\sigma_Z = \sigma_\epsilon \|w\|_2 = 0.32\%.$$

Compared with the empirical counterpart given in **Table 1**, this result suggests that microeconomic shocks are an important source of the aggregate variance. Thus, in terms of the aggregate variance, my empirical analysis supports the granular hypothesis.

Let us consider the tail probability of aggregate output. I consider two distributions for the underlying distribution of microeconomic shocks: Gaussian distribution, for which I apply Mills' ratio, and Laplace distribution, or more precisely, a Weibull distribution whose parameter τ is arbitrarily close to 1.[21] I apply Proposition 3.9 by viewing the Laplace distribution as the limit of τ. The parameter of the Laplace distribution is chosen such that the variance of microeconomic shocks is equal to σ_ϵ. Thus, the tail probability of aggregate output can be approximated as follows:

$$P(Z \leq -x) \leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\|w\|_2^2 \sigma_\epsilon^2}\right) \qquad \text{for Gaussian}$$

$$P(Z \leq -x) \sim P(w_1 \epsilon_1 \leq -x) = \frac{1}{2} \exp\left(-\frac{\sqrt{2}x}{\sigma_\epsilon \|w\|_\infty}\right) \quad \text{for Laplace}$$

By substituting the empirical values of $\|w\|_\infty$ and $\|w\|_2$ in **Table 3** into the above formulae, I plot them along with the estimated tail probability of the GDP growth rate in the left panel of **Figure 5**. The tail probability predicted by microeconomic shocks is negligible compared with the empirical estimate of the tail probability. For example, the GEV estimate of the probability that the GDP growth rate is less than -2.0% is 3.27%, while the tail probability predicted by microeconomic shocks is less than 0.01%. This means that the empirical granularity (especially the size of the largest firm) is too low to cause a large deviation in the GDP growth rate. Thus, in terms of the tail probability of aggregate output, my empirical analysis does not support the granular hypothesis.

---

[20]This value for σ_ϵ is chosen so that the labor-augmented productivity shocks would generate a 6% standard deviation of sales growth rate, that is, $30\% \times 0.2 = 6\%$, where 0.2 is the labor share in my exercise. Although the value of 30% seems to overestimate the variation of microeconomic shocks, the following empirical analysis shows that the contribution of microeconomic shocks to aggregate fluctuations is small. Our empirical analysis can be considered as giving the upper bound for the contribution of microeconomic shocks to aggregate fluctuations.

[21]For the assumption of a Laplace distribution, see footnote 4.

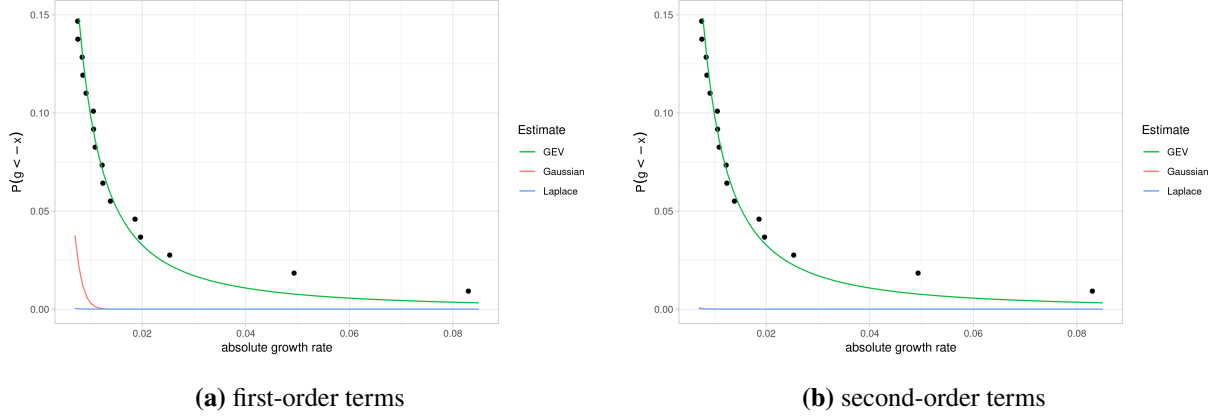**(a)** first-order terms   **(b)** second-order terms

Figure 5: Tail probability. Panel (a) shows the tail probabilities predicted by microeconomic shocks when Z is defined by Eq.(4). Panel (a) also shows the empirical CCDF and GEV estimate of the GDP growth rate, which are the same as in **Figure 3**. Panel (b) shows the tail probabilities predicted by microeconomic shocks when Z is defined as in Section 4. Others are same as in Panel (a).

5.3.2 Case of the second-order terms

Let us consider the aggregate variance when Z is as in Section 4. By using the norms given in **Table 3**, I obtain

$$\sigma_Z = \sqrt{(m_4 - 3m_2^2)\sum_i B_{i,i}^2 + 2m_2^2\|B\|_{\text{HS}}^2} = 0.108\%$$

where I assume $m_4 - 3m_2^2 = 3m_2^2 = 3\sigma_\epsilon^4$. [22] As in Section 5.3.1, microeconomic shocks are a source of aggregate fluctuations in terms of the aggregate variance. However, the magnitude of the aggregate variance caused by the second-order terms is small compared to the result in Section 5.3.1. This result suggests that the first-order approximation (i.e. Hulten's theorem) to aggregate output works well in this setting.

Next, consider the tail probability of Z. I consider Gaussian and Laplace distributions. In the former case, I use Example 4.1. In the latter case, I use an approximation method similar to Example 4.1. Let X_i be a standard Laplace random variable (i.e., $X_i := \epsilon_i/\sigma_\epsilon$) and let $A = -B$. The diagonalization of A yields

$$X^T A X - E[X^T A X] = \sum_i \mu_{i,A}(Y_i^2 - 1)$$

where Y_i is a linear combination of $X_1, ..., X_n$. Note that $Y_1, ..., Y_n$ are linearly independent but not necessarily independent of each other. I assume here that Y_i is a standard independent Laplace random variable.[23] Under this assumption, Proposition 3.9 yields that the tail probability of Z is approximately equal

---

[22]This assumption holds true when microeconomic shocks follow a Laplace distribution. In contrast, when they follow a Gaussian distribution, $m_4 - 3m_2^2$ is equal to 0.

[23]Since the true Y_i is a combination of $X_1, ..., X_n$, the averaging effect dampens the tail probability of true Y_i, compared to the Laplace distribution. That is, this assumption would generate a higher tail probability of Z. Thus, the assumption is considered conservative because even under this assumption, the tail probability of Z predicted by microeconomic shocks turns out to be negligible.

to $P(\mu_{1,A}(Y_1^2 - 1) > x/\sigma_\epsilon^2)$. Note that

$$P(Y_i^2 > y) = P(Y_i > \sqrt{y} \text{ or } Y_i < -\sqrt{y}) = e^{-\sqrt{2}y^{\frac{1}{2}}}.$$

Thus, by substituting $y = \frac{x}{\sigma_\epsilon^2 \|A\|} + 1$ into this equation, I can approximate the tail probability of Z. To summarize, I use the following approximation:

$$P(Z - EZ \le -x) \le \exp\left(-\frac{x^2}{4(\sigma_\epsilon^4 \|B\|_{\text{HS}}^2 + \sigma_\epsilon^2 \|B\|x)}\right) \quad \text{for Gaussian}$$

$$P(Z - EZ \le -x) \sim \exp\left(-\sqrt{\frac{2x}{\sigma_\epsilon^2 \|B\|} + 2}\right) \quad \text{for Laplace}$$

Finally, I use the upper bound of $\|B\| \le \frac{|\theta-1|}{2}\|w^*\|_\infty$.

Substituting the empirical values into these approximations, I plot them in the right panel of **Figure 5**. As in Section 5.3.1, the tail probability induced by microeconomic shocks is negligible compared with the estimated tail probability of the GDP growth rate. Thus, given the empirical granularity in Japan, microeconomic shocks would only generate small fluctuations, and not a large deviation, in aggregate output. In other words, the empirical granularity (especially the size of the largest firm) in Japan is not large enough to explain the observed large deviation in the GDP growth rate.

6 Conclusion

The literature on the granular hypothesis provides new insights into the analysis of aggregate fluctuations. It shows that not only exogenous aggregate shocks but microeconomic shocks can be a source of aggregate fluctuations. Recent studies have proposed many models to analyze how microeconomic shocks are related to aggregate output. Given these developments, it is important to assess the empirical relevance of the granular hypothesis.

This paper focuses on the distribution properties of aggregate fluctuations. My analysis shows that when the firm size distribution has a Pareto tail, shocks to the largest firm dominate those to other firms. For this reason, the CLT fails, and the size of the largest firm is the key variable to the aggregate variance and tail probability of aggregate output. Then, by using firm-level data in Japan, I test whether the empirical granularity (and the size of the largest firm) is large enough to generate the substantial variance and tail probability of aggregate output. I find that it is sufficiently large to increase the aggregate variance but not large enough to contribute to the tail probability of aggregate output. That is, microeconomic shocks would generate only small fluctuations, and not a large deviation, in aggregate output.

Note that my analysis assumes that an economy is efficient and that there is no extensive margin. If these assumptions are relaxed, the micro-originated aggregate fluctuations may be further amplified. For the first assumption, an extension to an inefficient economy is intensively studied in the literature. By introducing wedges (or markup), Bigio and La'o (2020) and Baqaee and Farhi (2020b) analyze how microeconomic

shocks propagate in an input-output network. For the second assumption, some recent papers (Grassi (2017); Baqaee (2018);Baqaee and Farhi (2020a)) tackle this problem by endogenizing network formation. Although their models are much more complicated, if Eq.(1) holds true, my analysis can be applied similarly. The characterization of the distribution properties implied by these models would contribute to a better understanding of the role of granularity in aggregate fluctuations.

7 Appendix

In Section 7.1, I give the proofs of the propositions in the main text. In Section 7.2, I explain the estimation of the tail probability of the GDP growth rate.

7.1 Proof

7.1.1 Proof of Propositions 3.2 and 3.3

I follow the proof in Gabaix (2011). However, unlike it, I use the extreme value theory (e.g., Embrechts et al. (1997); Resnick (1987); De Haan and Ferreira (2006)).

Proof of Proposition 3.2. Consider a random variable given by

$$n^2(\log n)^{-1}\|w\|_\infty^2 = \frac{1}{b^2}\frac{(\log n)^{-1}s_{\max}^2}{(n^{-1}\sum_i s_i)^2}.$$

First, since $Es_i < \infty$ by the assumption, $(n^{-1}\sum_i s_i)^2$ converges to $(Es_i)^2$ a.s. as $n \to \infty$ by the strong law of large numbers. Second, by using the same argument as in Example 3.1, one gets

$$\alpha(s_{\max}^2 - \alpha^{-1}\log(Kn)) \xrightarrow{d} u$$

where u is a non-degenerate random variable. Thus, $(\log n)^{-1}s_{\max}^2 \xrightarrow{P} \alpha^{-1}$. Combining these results, one gets

$$n^2(\log n)^{-1}\|w\|_\infty^2 \xrightarrow{P} \frac{1}{b^2\alpha(Es_i)^2}$$

Thus, the desired result follows. $\qquad\qquad\square$

Proof of Proposition 3.3. For simplicity, let $k := \alpha/2 > 1/2$. Consider a random variable given by

$$n^{2-1/k}\|w\|_\infty^2 = \frac{1}{b^2}\frac{n^{-1/k}s_{\max}^2}{(n^{-1}\sum_i s_i)^2}.$$

The same argument as in the proof of Proposition 3.2 yields that $(n^{-1}\sum_i s_i)^2$ converges to $(Es_i)^2$ a.s. and

$$(Kn)^{-1/k}s_{\max}^2 \xrightarrow{d} u_k,$$

where u_k is a non-degenerate random variable, independent of n. By combining these results together, one obtains

$$n^{2-1/k}\|w\|_\infty^2 \xrightarrow{d} \frac{K^{1/k}u_k}{b^2(Es_i)^2}$$

Thus, the desired result follows. $\qquad\square$

7.1.2 Proof of Propositions 3.4 and 3.5

The relation between the sum and maximum of random variables is studied in the literature on regularly varying functions (see, e.g., Section 8.2.4 in Embrechts et al. (1997) and Section 8.15 in Bingham et al. (1987)). My proof below follows this literature.

Proof of Proposition 3.4. Note that $Es_i^2 < \infty$ is equivalent to $\sum_{n=1}^{\infty} P(s_n^2 > \delta n) < \infty$ for all $\delta > 0$. Thus, the Borel-Cantelli lemma implies $P(s_n^2 > \delta n \text{ i.o.}) = 0$ for all $\delta > 0$, which is equivalent to $\lim_{n \to \infty} n^{-1} s_n^2 = 0$ a.s. Next, I use an equivalent relation between $n^{-1} s_n^2$ and $n^{-1} s_{\max}^2$: for some $n_0 \leq n$,

$$ n^{-1} s_n^2 \leq n^{-1} s_{\max}^2 \leq \max\left(\frac{s_1^2}{n}, ..., \frac{s_{n_0}^2}{n}, \frac{s_{n_0+1}^2}{(n_0+1)}, \frac{s_{n_0+2}^2}{(n_0+2)}, ..., \frac{s_n^2}{n} \right) $$

This inequality implies

$$ \lim_{n \to \infty} n^{-1} s_n^2 = 0 \text{ a.s.} \iff \lim_{n \to \infty} n^{-1} s_{\max}^2 = 0 \text{ a.s.} $$

Finally, the strong law of laws of large numbers yields $(n^{-1} \sum_i s_i^2) \overset{\text{a.s.}}{\to} Es_i^2$. Combining all results together, I obtain

$$ r_{\max} = \frac{(n^{-1} s_{\max}^2)}{(n^{-1} \sum_i s_i^2)} \overset{\text{a.s.}}{\to} 0 $$

$\qquad\square$

The relation between the sum and maximum for random variables with infinite mean is fully characterized by Bingham and Teugels (1981). They provide an equivalent condition for the existence of the non-degenerate limit distribution of $r_{\max}$.

Theorem 7.1 (Theorem 8.15.3 in Bingham et al. (1987)). *Let $\xi_1, ..., \xi_n$ be iid random variables drawn from a distribution F. Let S_n and M_n denote the sum and maximum of $\xi_1, ..., \xi_n$, respectively. Then, the following conditions are equivalent to each other:*

(i) M_n / S_n has a non-degenerate limit distribution.

(ii) F is attracted to a stable law of exponent $k \in (0, 1)$.

(iii) $E(S_n / M_n - 1)$ has a positive finite limit.

Let $\xi_i = s_i^2$. Since a distribution with a Pareto tail with exponent $\alpha \in (0, 2)$ is attracted to the stable law of $k \in (0, 1)$, one can immediately show Proposition 3.5. In the following, I give a simplified proof of (ii) $\Rightarrow$ (iii) for Zipf's law, that is, the second part of Proposition 3.5. This proof shows how one can get the lower bound of the mean of $r_{\max}$.

Proof of the second part of Proposition 3.5. Note that for a general distribution F, the following equation

holds:

$$E\left[\exp\left(it\left(\frac{\sum_i \xi_i}{\xi_{\max}} - 1\right)\right)\right] = n\int_{-\infty}^{\infty} dF(v)\left[\int_{-\infty}^{v}\exp\left(it\frac{x}{v}\right)dF(x)\right]^{n-1}$$

By taking the derivative of both sides and setting t equal to 0, one gets

$$E\left[\frac{\sum_i \xi_i}{\xi_{\max}} - 1\right] = n(n-1)\int_0^{\infty} v^{-1}\int_0^v x\,dF(x)F^{n-2}(v)dF(v)$$

Here, I use the assumption that ξ_i is a non-negative random variable.

For simplicity, consider a simple case where $\overline{F}(x) = x^{-1/2}$ on $[1, \infty)$. In this case, $\int_1^v x\,dF(x) = v^{1/2} - 1$ and

$$(v^{-1/2} - v^{-1})F^{n-2}(v) = (\overline{F}(v) - \overline{F}^2(v))F^{n-2}(v)$$
$$= \overline{F}(v)F^{n-1}(v)$$
$$= F^{n-1}(v) - F^n(v)$$

Thus,

$$E\left[\frac{\sum_i \xi_i}{\xi_{\max}} - 1\right] = n(n-1)\int_1^{\infty} v^{-1}\int_1^v x\,dF(x)F^{n-2}(v)dF(v)$$
$$= n(n-1)\int_1^{\infty}(F^{n-1}(v) - F^n(v))dF(v)$$
$$= n(n-1)\left(\frac{1}{n}[F^n(v)]_1^{\infty} - \frac{1}{n+1}[F^{n+1}(v)]_1^{\infty}\right)$$
$$= n(n-1)\frac{1}{n(n+1)} \to 1$$

Therefore, $E\frac{1}{r_{\max}}$ converges to 2.

Finally, I use Jensen's inequality. Since $f(x) := 1/x$ is a convex function, Jensen's inequality yields $\frac{1}{Er_{\max}} \le E\frac{1}{r_{\max}}$. Thus, I obtain

$$\lim_{n\to\infty} Er_{\max} \ge \lim_{n\to\infty}\frac{1}{E\frac{1}{r_{\max}}} = \frac{1}{2}$$

$\square$

7.1.3 Proof of Proposition 3.8

I follow the proof strategy used in the theory of concentration inequalities. For more details, see Chapter 2 in Wainwright (2019) and Chapter 2 in Vershynin (2020). I need a lemma that characterizes the moment generating function.

Lemma 7.2 (Theorem 2.1.3 in Wainwright (2019))**.** *Let X be a zero-mean random variable. Suppose that*

$$P(|X| > x) \le K\exp(-\beta x) \text{ for all } x \ge 0,$$

where K and β are positive constants. Then, there exist non-negative constants (C, c) such that

$$Ee^{\lambda X} \le e^{C\lambda^2} \text{ for all } |\lambda| < \frac{1}{c}.$$

Proof of Proposition 3.8. Markov's inequality and the independence of $\epsilon_1, ..., \epsilon_n$ yield

$$P(Z \leq -x) \leq e^{-\lambda x} \prod_i E \exp(\lambda w_i \epsilon_i).$$

The lemma implies that there exists non-negative constants (C, c) such that for all $|\lambda| \leq \frac{1}{c\|w\|_\infty}$ and for all i,

$$E \exp(\lambda w_i \epsilon_i) \leq \exp(C\lambda^2 w_i^2).$$

Hence, I have the upper bound of $P(Z \leq -x)$:

$$P(Z \leq -x) \leq \exp(-\lambda x + C\lambda^2 \|w\|_2^2)$$

Next, I minimize the upper bound by choosing an optimal λ subject to the constraint $|\lambda| \leq \frac{1}{c\|w\|_\infty}$. The optimal λ is given by

$$\lambda = \min\left(\frac{x}{2C\|w\|_2^2}, \frac{1}{c\|w\|_\infty}\right),$$

and

$$P(Z \leq -x) \leq \exp\left(-\min\left(\frac{x^2}{4C\|w\|_2^2}, \frac{x}{2c\|w\|_\infty}\right)\right).$$

Here, I used the fact that if $x \geq \frac{2C\|w\|_2^2}{c\|w\|_\infty}$,

$$-\frac{x}{c\|w\|_\infty} + \frac{C\|w\|_2^2}{c^2\|w\|_\infty^2} \leq -\frac{x}{2c\|w\|_\infty}.$$

Therefore, the desired result follows. $\qquad\square$

7.1.4   Proof of Proposition 3.9

The topics of heavy-tailed distributions are one of the fields of probability theory (see, e.g., Embrechts et al. (1997); Foss et al. (2011)). Following Foss et al. (2011), I introduce the classes of distributions $\mathcal{S}$ and $\mathcal{S}_\mathbb{R}$ on $\mathbb{R}^+$ and $\mathbb{R}$, respectively.[24]

Definition 7.1. A distribution F on $\mathbb{R}^+$ belongs to class $\mathcal{S}$ (denoted by $F \in \mathcal{S}$) if the convolution $F * F$ satisfies

$$\overline{F * F}(x) \sim 2\overline{F}(x) \text{ as } x \to \infty.$$

A distribution F on $\mathbb{R}$ belongs to class $\mathcal{S}_\mathbb{R}$ (denoted by $F \in \mathcal{S}_\mathbb{R}$) if $F^+ \in \mathcal{S}$, where F^+ is the distribution of $X^+ := \max(X, 0)$.

The distribution on $\mathbb{R}$ with a Pareto tail belongs to $\mathcal{S}_\mathbb{R}$ because, as shown in Example 3.4, the Pareto distribution on $\mathbb{R}^+$ satisfies the relation $\overline{F * F}(x) \sim 2\overline{F}(x)$. One can also show that the distribution on $\mathbb{R}$ with a Weibull tail belongs to $\mathcal{S}_\mathbb{R}$ (see, e.g., Example 1.4.3 in Embrechts et al. (1997)).

---

[24]The classes of distributions $\mathcal{S}$ and $\mathcal{S}_\mathbb{R}$ are usually called "subexponential," whose notations are first introduced by Athreya and Ney (2004) and widely used in the related literature (e.g., Embrechts et al. (1997); Foss et al. (2011)). However, in the literature on concentration inequality (e.g., Boucheron et al. (2012)), the term "subexponential" means distributions satisfying the property in Lemma 7.2. Since the two usages of "subexponential" are incompatible and confusing, I decided not to use it throughout this paper.

The main tool for the proof of Proposition 3.9 is the following theorem (given as Corollary 3.19 in Foss et al. (2011); see also Embrechts and Goldie (1982)).

Theorem 7.3 (Corollary 3.19 in Foss et al. (2011)). *Suppose that $F \in \mathcal{S}_{\mathbb{R}}$ (the reference distribution). If distributions $G_1, \ldots, G_n$ satisfy $\overline{G}_i(x)/\overline{F}(x) \to c_i$ as $x \to \infty$ for some constant $c_i \geq 0$ and $i = 1, \ldots, n$, then*

$$\frac{\overline{G_1 * \ldots * G_n}(x)}{\overline{F}(x)} \to c_1 + \ldots + c_n \ \text{as } x \to \infty.$$

Proof of Proposition 3.9. Let the distribution of the first component $w_1 \epsilon_1$ be the reference distribution F (i.e., $F(x) := P(w_1 \epsilon_1 \leq x)$), and let G_i be the distribution of $w_i \epsilon_i$ (i.e., $G_i(x) := P(w_i \epsilon_i \leq x)$). Note that the distribution of Z is given by the convolution of G_i, that is, $G_1 * \ldots * G_n$.

For the case of a Pareto tail,

$$\frac{\overline{G}_i(x)}{\overline{F}(x)} = \frac{P(w_i \epsilon_i > x)}{P(w_1 \epsilon_1 > x)} = \left(\frac{w_i}{w_1}\right)^{\beta}$$

Thus,

$$\lim_{x \to \infty} \frac{\overline{G}_i(x)}{\overline{F}(x)} = \left(\frac{w_i}{w_1}\right)^{\beta}$$

Thus, Theorem 7.3 and the symmetry of the distribution of Z yield

$$P(Z \leq -x) \sim \sum_i \left(\frac{w_i}{w_1}\right)^{\beta} P(w_1 \epsilon_1 \leq -x)$$

For the case of a Weibull-tail,

$$\frac{\overline{G}_i(x)}{\overline{F}(x)} = \frac{P(w_i \epsilon_i > x)}{P(w_1 \epsilon_1 > x)} = e^{\beta(w_1^{-\tau} - w_i^{-\tau})x^{\tau}}$$

Since the largest weight w_1 is strictly larger than any other weights,

$$\lim_{x \to \infty} \frac{\overline{G}_i(x)}{\overline{F}_i(x)} = \begin{cases} 1 & \text{for } i = 1 \\ 0 & \text{for } i = 2, 3, \ldots, n \end{cases}$$

Thus, Theorem 7.3 and the symmetry of the distribution of Z yield

$$P(Z \leq -x) \sim P(w_1 \epsilon_1 \leq -x)$$

$\square$

7.1.5 Proof for Section 4

I first define a sub-Gamma random variable and then give a lemma used in Example 4.1. The proof of this lemma is essentially the same as the latter half of the proof of Proposition 5 in Acemoglu et al. (2017).

Definition 7.2. A real-valued centered random variable X is called sub-Gamma with variance factor v and scale c if for all λ such that $0 < \lambda < 1/c$,

$$Ee^{\lambda X} \leq \frac{\lambda^2 v}{2(1 - c\lambda)}.$$

31

Lemma 7.4. *Let X be a sub-Gamma random variable with variance factor v and scale c. Then,*

$$P(X > x) \le \exp\left(-\frac{x^2}{2(v + cx)}\right) \quad \text{for } x \ge 0$$

Proof. I use Chernoff's method. First, by Markov's inequality,

$$P(X > x) \le \exp\left(-\lambda x + \frac{v\lambda^2}{2(1 - c\lambda)}\right)$$

I optimize the upper bound with respect to $\lambda \in (0, 1/c)$:

$$\sup_{\lambda \in (0,1/c)} \left(x\lambda - \frac{\lambda^2 v}{2(1 - c\lambda)}\right) = \frac{v}{c^2} h_1\left(\frac{cx}{v}\right)$$

where $h_1(u) := 1 + u - \sqrt{1 + 2u}$. Finally, by using the following inequality

$$h_1(u) \ge \frac{u^2}{2(1 + u)} \quad \text{for } u > 0,$$

the desired result follows. $\qquad\square$

Proof of Lemma 4.1. For the first result, let Λ be a diagonal matrix whose diagonal entries are given by $w_1^*, ..., w_n^*$. Note that matrix B can be written as $B = \frac{\theta - 1}{2} M$, where $M := \Lambda - w^* w^{*T}$. Consider the quadratic form of M:

$$\langle Mx, x \rangle = \langle \Lambda x, x \rangle - \langle ww^T x, x \rangle$$
$$= \langle \Lambda x, x \rangle - \|w^T x\|^2 \le \langle \Lambda x, x \rangle$$

Since M is a symmetric real matrix, the largest eigenvalue of M is given by $\lambda_{\max} = \max_{\|x\|=1}\langle Mx, x \rangle$. Thus, I have $\lambda_{\max} \le \max(w_1^*, ..., w_n^*) = \|w^*\|_\infty$.

What remains is to show that all eigenvalues of M are non-negative. Assume that $1 - w_i^* \ge 0$ for all i. Since M is a symmetric matrix with real non-negative diagonal entries, all eigenvalues are real. Furthermore, Gershgorin's circle theorem (see, e.g., Garren (1968)) implies that for each eigenvalue λ, there exists an index i such that

$$\lambda \in \left[b_{ii} - \sum_{j \ne i} |b_{ij}|, b_{ii} + \sum_{i \ne j} |b_{ij}|\right]$$

Since matrix M is diagonally dominant, one gets $\lambda \ge 0$. Therefore, the desire result follows.

For the second result, note that

$$\|B\|_{\text{HS}}^2 := \sum_i \sum_j b_{ij}^2$$
$$= \frac{(\theta - 1)^2}{4} \sum_i w_i^{*2} \sum_j (1_{\{i=j\}} - w_j^*)^2$$
$$= \frac{(\theta - 1)^2}{4} \sum_i w_i^{*2}(1 - 2w_i^* + \|w^*\|_2^2)$$
$$= \frac{(\theta - 1)^2}{4} \left(\sum_i w_i^{*2}(1 - 2w_i^*) + \|w^*\|_2^4\right)$$

Assume that $(1 - 2w_i^*) > 0$ for all i. By the above equation, I obtain

$$c\|w^*\|_2^2 < \|B\|_{\text{HS}}^2 < C\|w^*\|_2^2$$

$\square$

Proof of Proposition 4.2. Let us consider the square of $\epsilon'B\epsilon$:

$$(\epsilon'B\epsilon)^2 = \sum_{1\leq i,j,k,\ell\leq n} B_{i,j}B_{k,\ell}\epsilon_i\epsilon_j\epsilon_k\epsilon_\ell$$

The independence of $\epsilon_1, ..., \epsilon_n$ yields that

$$E(\epsilon_i\epsilon_j\epsilon_k\epsilon_\ell) = \begin{cases} m_4 \text{ if } i = j = k = \ell \\ m_2^2 \text{ if } i = j \neq k = \ell \text{ or } i = k \neq j = \ell \text{ or } i = \ell \neq k = j \\ 0 \text{ otherwise} \end{cases}$$

Thus, by taking its expectation, I have

$$E[(\epsilon'B\epsilon)^2] = \sum_i B_{i,i}^2 m_4 + \sum_{1\leq i\neq k\leq n} B_{i,i}B_{k,k}m_2^2 + \sum_{1\leq i\neq j\leq n} B_{i,j}B_{j,i}m_2^2 + \sum_{1\leq i\neq k\leq n} B_{i,k}B_{k,i}m_2^2$$

$$= m_4 \sum_i B_{i,i}^2 + m_2^2 \left[\sum_{1\leq i\neq k\leq n} B_{i,i}B_{k,k} + 2 \sum_{1\leq i\neq j\leq n} B_{i,j}^2 \right]$$

The summations inside the bracket are written as follows:

$$\sum_{1\leq i\neq k\leq n} B_{i,i}B_{k,k} = \sum_i B_{i,i} \sum_k B_{k,k} - \sum_i B_{i,i}^2 = \text{tr}(B)^2 - \sum_i B_{i,i}^2$$

$$\sum_{1\leq i\neq j\leq n} B_{i,j}^2 = \sum_i \sum_j B_{i,j}^2 - \sum_i B_{i,i}^2$$

$$= \sum_i \sum_j B_{i,j}B_{j,i} - \sum_i B_{i,i}^2$$

$$= \sum_i (B^2)_{i,i} - \sum_i B_{i,i}^2 = \text{tr}(B^2) - \sum_i B_{i,i}^2$$

By plugging these equations into the bracket, I obtain

$$E[(\epsilon'B\epsilon)^2] = m_4 \sum_i B_{i,i}^2 + m_2^2 \left[\text{tr}(B)^2 - \sum_i B_{i,i}^2 + 2\text{tr}(B^2) - 2\sum_i B_{i,i}^2 \right]$$

$$= (m_4 - 3m_2^2) \sum_i B_{i,i}^2 + m_2^2[\text{tr}(B)^2 + 2\text{tr}(B^2)]$$

Since $E(\epsilon'B\epsilon) = m_2\text{tr}(B)$, I obtain the variance of Z

$$\sigma_Z^2 = (m_4 - 3m_2^2) \sum_i B_{i,i}^2 + m_2^2[\text{tr}(B)^2 + 2\text{tr}(B^2)] - m_2^2\text{tr}(B)^2$$

$$= (m_4 - 3m_2^2) \sum_i B_{i,i}^2 + 2m_2^2\text{tr}(B^2)$$

Finally, note that $\text{tr}(B^2) = \|B\|_{\text{HS}}^2$. Indeed, since B is a symmetric matrix,

$$\|B\|_{\text{HS}}^2 := \sum_i \sum_j a_{ij}^2 = \sum_i \sum_j a_{ij}a_{ji} = \text{tr}(B^2)$$

Thus, the desired result follows.

Suppose that $b_{ij} := \frac{(\theta-1)}{2} w_i^*(1_{\{i=j\}} - w_j^*)$. Note that Lemma 4.1 implies that $\|B\|_{\mathrm{HS}}^2$ is the same order as that of $\|w^*\|_2^2$. For $\sum B_{i,i}^2$, note that

$$\sum B_{i,i}^2 = \frac{(\theta-1)^2}{4} \sum_i w_i^{*2}(1 - w_i^*)^2$$

Since $w_i^* \leq 1$, $\sum B_{i,i}^2$ is the same order as that of $\|w^*\|_2^2$. Thus, it follows that σ_Z^2 is of the order of $\|w^*\|_2^2$. $\qquad\square$

Proposition 4.3 is an immediate consequence of Lemma 4.1 and Proposition 1.1 given by Götze et al. (2019).[25]

Theorem 7.5 (Proposition 1.1 in Götze et al. (2019))**.** *Suppose that $\epsilon_1, ..., \epsilon_n$ are iid random variables with mean 0 and that B is an $n \times n$ symmetric matrix. Suppose further that $\epsilon_1, ..., \epsilon_n$ satisfy $\|\epsilon_i\|_{\psi_\alpha} \leq C$ for $\alpha \in (0,1] \cup \{2\}$ for $i = 1, ..., n$, where $\|\cdot\|_{\psi_\alpha}$ is the Orlicz norm defined by $\|\epsilon_i\|_{\psi_\alpha} := \inf\{t > 0 : E\exp(\epsilon_i^\alpha/t^\alpha) \leq 2\}$. Then,*

$$P(Z - EZ \leq -x) \leq \exp\left(-c\min\left(\frac{x^2}{C^4\|B\|_{\mathrm{HS}}^2}, \left(\frac{x}{C^2\|B\|}\right)^{\frac{\alpha}{2}}\right)\right) \quad \text{for } x \geq 0,$$

where c is an absolute constant.

Proof of Proposition 4.3. If microeconomic shocks follow a Laplace distribution, the Orlicz norm with $\alpha = 1$ exists by definition, that is, there exists a constant C such that $\|\epsilon_i\|_{\psi_1} \leq C$. Thus, by combining Lemma 4.1, the desired result follows. $\qquad\square$

7.2 Estimation of the tail probability of the GDP growth rate

I explain how the left tail probability of the GDP growth rate is estimated in Section 5.1. Before explaining the estimation method, I provide another visual inspection of the left tail of the distribution of the GDP growth rate. Let X be the absolute value of the negative GDP growth rate in my sample. I define by $e(u)$ the mean excess function over the threshold u:

$$e(u) := E[X - u \mid X > u] \text{ for } u > 0.$$

The dependence of $e(u)$ on u is determined by the underlying distribution of X. For example, if X follows an exponential distribution with parameter λ, then $e(u) = \lambda^{-1}$, that is, $e(u)$ is a constant. In contrast, if X follows a generalized Pareto distribution, then $e(u)$ becomes a linear function of u with a positive slope. Intuitively, if $e(u)$ is an increasing function of u, the underlying distribution of X has a fatter tail than an exponential.

---

[25]The result by Götze et al. (2019) is the extension of the so-called Hanson-Wright inequality. For the Hanson-Wright inequality, see Rudelson and Vershynin (2013).
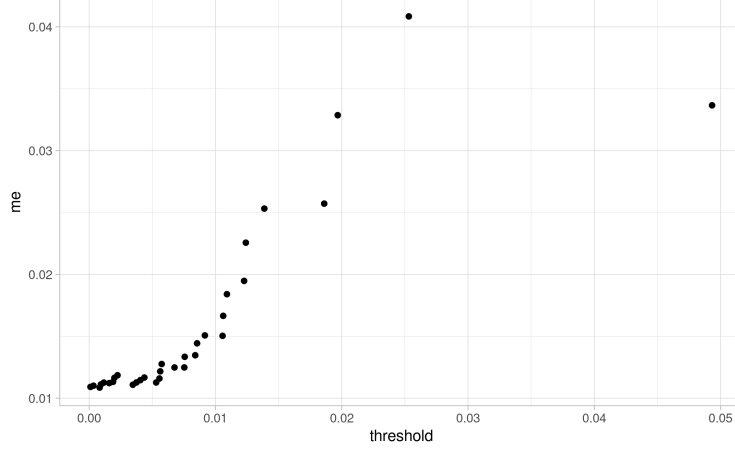
Figure 6: Empirical mean excess function for the negative GDP growth rates.

Let $\widehat{e}(u)$ be its empirical counterpart:

$$\widehat{e}(u) := \frac{1}{\mathrm{card}\Delta(u)} \sum_{t \in \Delta(u)} (X_t - u) \ \text{ for } u > 0,$$

where $\Delta(u)$ is the set of t such that $X_t > u$ and $\mathrm{card}\Delta(u)$ is the cardinality of $\Delta(u)$. **Figure 6** depicts the plot of $\widehat{e}(u)$ over u, showing that $\widehat{e}(u)$ does not converge to a constant value but increases as u increases. This graphical inspection suggests that the distribution of the GDP growth rate has a fatter tail than an exponential.

Keeping this observation in mind, I estimate this fatter tail by two methods based on the extreme value theory. The first method uses the limit theorem, in which $nP(X > c_n x + d_n)$ with some normalized constants c_n and d_n converges to the logarithm of the generalized extreme value (GEV) distribution with parameter ξ as $n \to \infty$.[26] Thus, assuming that this approximation holds well, I estimate its parameter. More precisely, let $X_1, ..., X_n$ be the absolute value of the negative GDP growth rate in my sample, and let $X_{j,n}$ be the jth largest sample among the n samples. I approximate the tail probability of X by

$$\widehat{P}(X > x) = \frac{k}{n}\left(\frac{x}{X_{k+1,n}}\right)^{-1/\widehat{\xi}^{(H)}}$$

The parameter $\xi^{(H)}$ measures the heavy-tailedness of the distribution. I estimate $\widehat{\xi}^{(H)}$ based on the k largest samples (Hill's estimator):

$$\widehat{\xi}^{(H)} := \frac{1}{k}\sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k+1,n}.$$

The estimate of the tail probability is shown in **Figure 3**, where $k = 10$, $\widehat{\xi}^{(H)} = 0.624$, and $X_{k+1} = 0.0106$.

The second method uses the limit theorem that the distribution of X conditional on $X > u$, where u is a threshold, converges to the generalized Pareto distribution (GPD). More precisely, I use the property that

---

[26]For the deviation of this limit theorem, see, for example, Chapter 6 in Embrechts et al. (1997).

for $y > 0$,

$$P(X > u + y) = P(X > u)P(X > u + y \mid X > u).$$

The first term on the right-hand side, $P(X > u)$, can be approximated by the empirical distribution function because the samples in this region are abundant. For the second term $P(X > u + y \mid X > u)$, I use the GPD for the approximation. Thus, the estimate of the tail probability over threshold u is given by

$$\widehat{P}(X > u + y) = \frac{N_u}{n}\left(1 + \widehat{\xi}\frac{y}{\widehat{\beta}}\right)^{-1/\widehat{\xi}} \quad \text{for } y > 0,$$

where N_u is the number of samples exceeding u. I estimate the parameters ξ, β by the maximum likelihood method. The estimate of the tail probability is shown in **Figure 3**, where $\widehat{\xi}^{(H)} = 0.880, \widehat{\beta} = 0.00480$, and $u = 0.01$.

Acknowledgment

References

Acemoglu, D. and Azar, P. D. (2020). Endogenous production networks. *Econometrica*, 88(1):33–82.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A. E., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016.

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2017). Microeconomic Origins of Macroeconomic Tail Risks. *American Economic Review*, 107(1):54–108.

Acemoglu, D. and Tahbaz-Salehi, A. (2020). Firms, Failures, and Fluctuations: The Macroeconomics of Supply Chain Disruptions. *Working Paper*.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–89.

Altinoglu, L. (2020). The origins of aggregate fluctuations in a credit network economy. *Journal of Monetary Economics*, (xxxx).

Arata, Y. (2019). Firm growth and laplace distribution: The importance of large jumps. *Journal of Economic Dynamics and Control*, 103:63–82.

Atalay, E. (2017). How Important are Sectoral Shocks? *American Economic Journal: Macroeconomics*, 9(4):254–280.

Atalay, E., Drautzburg, T., and Wang, Z. (2018). Accounting for the Sources of Macroeconomic Tail Risks. *Economics Letters*, 165:65–69.

Athreya, K. B. and Ney, P. E. (2004). *Branching processes*. Courier Corporation.

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

Baqaee, D. R. (2018). Cascading Failures in Production Networks. *Econometrica*, 86(5):1819–1838.

Baqaee, D. R. and Farhi, E. (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica*, 87(4):1155–1203.

Baqaee, D. R. and Farhi, E. (2020a). Entry vs. Rents. *Working Paper*.

Baqaee, D. R. and Farhi, E. (2020b). Productivity and Misallocation in General Equilibrium. *Quarterly Journal of Economics*, 135(1):105–163.

Bigio, S. and La'o, J. (2020). Distortions in Production Networks. *The Quarterly Journal of Economics*.

Bingham, N., Goldie, C., and Teugels, J. (1987). *Regular Variation*. Cambridge University Press.

Bingham, N. and Teugels, J. (1981). Conditions Implying Domains of Attarction. In *Proceedings 6th Conference Probabability Theory, Brasov*, pages 23–34.

Bottazzi, G. and Secchi, A. (2006). Explaining the distribution of firm growth rates. *The RAND Journal of Economics*, 37(2):235–256.

Boucheron, S., Lugosi, G., and Massart, P. (2012). *Concentration inequalities A nonasymptotic theory of independence*.

Burstein, A., Carvalho, V. M., and Grassi, B. (2020). Bottom-up Markup Fluctuations. *Working Paper*, (December).

Carvalho, V. M. (2010). Aggregate Fluctuations and the Network Structure of Intersectoral Trade. *Working Paper*.

Carvalho, V. M. (2014). From Micro to Macro via Production. *Journal of Economic Perspectives*, 28(4):23–47.

Carvalho, V. M. and Gabaix, X. (2013). The Great Diversification and its Undoing. *American Economic Review*, 103(5):1697–1727.

Carvalho, V. M. and Tahbaz-Salehi, A. (2019). Production Networks: A Primer. *Annual Review of Economics*, 11:635–663.

Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575.

Coad, A. (2009). *The growth of firms: A survey of theories and empirical evidence*. Edward Elgar Publishing.

Cúrdia, V., Del Negro, M., and Greenwald, D. L. (2014). Rare shocks, great recessions. *Journal of Applied Econometrics*, 29(7):1031–1052.

De Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 21. Springer.

Di Giovanni, J. and Levchenko, A. A. (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy*, 120(6):1083–1132.

Di Giovanni, J., Levchenko, A. A., and Mejean, I. (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica*, 82(4):1303–1340.

Di Giovanni, J., Levchenko, A. A., and Mejean, I. (2019). Foreign Shocks as Granular Fluctuations. *Working Paper*.

Dosi, G., Pereira, M. C., and Virgillito, M. E. (2017). The footprint of evolutionary processes of learning and selection upon the statistical properties of industrial dynamics. *Industrial and Corporate Change*, 26(2):187–210.

Elliott, M., Golub, B., and Leduc, M. V. (2020). Supply network formation and fragility. *arXiv preprint arXiv:2001.03853*.

Embrechts, P. and Goldie, C. M. (1982). On convolution tails. *Stochastic Processes and their Applications*, 13(3):263–278.

Embrechts, P., Kll Uppelberg, C., and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Technical report.

Fagiolo, G., Napoletano, M., and Roventini, A. (2008). Are output growth-rate distributions fat-tailed? some evidence from oecd countries. *Journal of Applied Econometrics*, 23(5):639–669.

Foerster, A. T., Sarte, P.-D. G., and Watson, M. W. (2011). Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. *Journal of Political Economy*, 119(1):1–37.

Foss, S., Korshunov, D., and Zachary, S. (2011). *An Introduction to Heavy-Tailed and Subexponential Distributions*.

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3):733–772.

Garren, K. R. (1968). *Bounds for the Eigenvalues of a Matrix*.

Götze, F., Sambale, H., and Sinulis, A. (2019). Concentration inequalities for polynomials in α-sub-exponential random variables.

Grassi, B. (2017). IO in I-O: Size, Industrial Organization, and the Input-Output Network Make a Firm Structurally Important. *Working Paper*.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2020). Firm volatility in granular networks. *Journal of Political Economy*, 128(11):4097–4162.

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

Huneeus, F. (2018). Production Network Dynamics and the Propagation of Shocks. *Working Paper*.

Jones, C. I. (2011). Intermediate Goods and Weak Links in the Theory of Economic Developmen. *American Economie Journal: Macroeconomics*, 3(2):1–28.

Jones, C. I. (2013). Misallocation, Economic Growth, and Input-Output Economics. In Acemoglu, D., Arellano, M., and Dekel, E., editors, *Proceedings of Econometric Society World Congress*, pages 419–55, Cambridge, UK. Cambridge Univ. Press.

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

Luo, S. (2020). Propagation of financial shocks in an input-output economy with trade and financial linkages of firms. *Review of Economic Dynamics*, 36(August 1997):246–269.

Magerman, G., Bruyne, K. D., Dhyne, E., and Hove, J. V. (2017). Heterogeneous Firms and the Micro Origins of Aggregate Fluctuations. *Working Paper*.

Miranda-Pinto, J. (2021). Production network structure, service share, and aggregate volatility. *Review of Economic Dynamics*, 39:146–173.

Oberfield, E. (2018). A theory of input–output architecture. *Econometrica*, 86(2):559–589.

Reischer, M. (2019). Finance-thy-Neighbor. Trade Credit Origins of Aggregate Fluctuations. *Working paper*.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, volume 4. Springer Science & Business Media.

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-gaussian concentration.

Saichev, A. I., Malevergne, Y., and Sornette, D. (2009). *Theory of Zipf's law and beyond*, volume 632. Springer Science & Business Media.

Sornette, D. (2006). *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media.

Stella, A. (2015). Firm Dynamics and the Origins of Aggregate Fluctuations. *Journal of Economic Dynamics and Control*, 55:71–88.

Su, H.-L. (2019). Financial Frictions, Capital Misallocation and Structural Change. *Working Paper*.

Taschereau-Dumouchel, M. (2020). Cascades and Fluctuations in an Economy with an Endogenous Production Network. *Working Paper*.

Tintelnot, F., Kikkawa, K., Mogstad, M., and Dhyne, E. (2019). Trade and Domestic Production Networks. *Working Paper*.

Vershynin, R. (2020). High-Dimensional Probability An Introduction with Applications in Data Science. Technical report.

Wainwright, M. J. (2019). *High-Dimensional Statistics*.

Yeh, C. (2019). Revisiting the origins of business cycles with the size-variance relationship. *Working paper*.