RIETI Discussion Paper Series 20-E-027

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

Yoshiyuki ARATA
RIETI

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 20-E-027
March 2020

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

Yoshiyuki Arata (RIETI)*

Abstract

What drives macroeconomic fluctuations? The recent macroeconomic literature (e.g. Gabaix (2011); Acemoglu et al. (2012)) shows that microeconomic shocks are an important source for macroeconomic fluctuations because of the granularity of an economy. I provide probabilistic results that quantify the role of the granularity in the variance of aggregate output and macroeconomic tail probability. First, in the linear model of Hulten's theorem, I show that the limiting behavior of the variance of aggregate output as the number of firms goes to infinity is equivalent to that of the contribution of the largest firm when the distribution of firm size has a Pareto-like tail. Empirical data in Japan also indicates the significance of the granularity in terms of the variance of aggregate output. Second, I obtain a formula that links the macroeconomic tail probability and the tail probability of microeconomic shocks. Utilizing the formula, I find that the empirical granularity is too low to lead to the large deviation of aggregate output in Japan. Third, I consider non-linear models and obtain the upper bounds for the variance of aggregate output and the macroeconomic tail probability. This shows that without the granularity, macroeconomic fluctuations decay very rapidly as the number of firms increases, which is the diversification argument in the non-linear models. In summary, my results imply that the granularity is a necessary condition both in linear and non-linear models, and that both granularity and non-linear terms are needed to explain the micro-originated large deviation of aggregate output.

JEL classifications:
Keywords: Granular hypothesis; Macroeconomic fluctuations; Hulten's theorem

*Research Institute of Economy, Trade and Industry (RIETI), 1-3-1 Kasumigaseki Chiyoda-ku, Tokyo, 100-8901Japan. E-mail: arata-yoshiyuki@rieti.go.jp. This study is conducted as a part of "Economic Growth and Fluctuations In Population Decline" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This research was supported by MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena, Macroeconomic Simulations). This research uses computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID: hp160259, hp170242, hp180177, hp190148). This study is supported by JSPS KAKENHI Grand Numbers 17H02041, 18K12788.

1 Introduction

What drives macroeconomic fluctuations? In the past, the diversification argument that idiosyncratic microeconomic shocks cancel out each other and only macroeconomic shocks matter was dominant (e.g. Lucas (1977)). However, the recent macroeconomic literature gives a completely different view. Gabaix (2011) shows that when firm size is highly heterogeneous and has a Pareto-like tail, macroeconomic fluctuations induced by microeconomic shocks do not die out. He calls this view the granular hypothesis. Acemoglu et al. (2012) consider an input-output production network explicitly and show that the heterogeneity of network structure leads to the granularity of firm size and, in turn, to the micro-originated macroeconomic fluctuations. Building on these model, many subsequent papers add new feature and study a link between microeconomic shocks and macroeconomic fluctuations.

But to what extent does the granularity explain macroeconomic fluctuations? While the literature provides many useful formulas that links microeconomic shocks and macroeconomic variables, it requires further consideration to assess the role of the granularity. The first problem encountered is about the measure of macroeconomic fluctuations. Most of the literature use the variance of aggregate output as the measure of macroeconomic fluctuations, but the macroeconomic tail probability (i.e. the probability of the large deviation of aggregate output) is another important measure (e.g. Acemoglu et al. (2017); Atalay et al. (2018)). The role of the granularity may depend on the choice of the measures of macroeconomic fluctuations. Second, since the distribution of microeconomic shocks affects macroeconomic fluctuations, it is necessary to consider the dependence on the distribution to assess the role of the granularity. Third, while the linear model of Hulten's theorem, which states that aggregate output is the weighted sum of microeconomic shocks, is widely used, non-linear higher-order terms may alter the role of the granularity (e.g. Baqaee and Farhi (2019)). It is important to consider the extension to models including the non-linear higher-order terms and to study how the granularity affects macroeconomic fluctuations. To the best of my knowledge, the consensus about the role of the granularity taking into account these points has not been reached yet.

In this paper, I provide probabilistic results that quantify the role of the granularity and are directly applicable to an empirical analysis. In my analysis, the granularity means that in an economy, there exists a small number of firms having substantial impacts on aggregate output. I study the variance of aggregate output and the macroeconomic tail probability as the measure of macroeconomic fluctuations, in the linear (i.e. the linear model of Hulten's theorem) and non-linear models. My contributions are the following three:

1. In the linear model, I show that the granularity (especially, the size of the largest firm) determines the limiting behavior of the variance of aggregate output when the distribution of firm size has a Pareto-like tail. Empirical data in Japan also implies the empirical significance of the granularity in terms of the variance of aggregate output.

2. In the linear model, I obtain a formula that links the macroeconomic tail probability and the granularity. By using it, I find that the empirical granularity is too low to lead to the large deviation of aggregate output in Japan.

3. In non-linear models, I obtain the upper bounds for the variance of aggregate output and the macroeconomic tail probability. This implies that without the granularity, the upper bounds, i.e. macroeconomic fluctuations, decay very fast as the number of firms increases (the diversification argument in non-linear models).

My results implies that the granularity is a necessary condition for micro-originated macroeconomic fluctuations in both linear and non-linear models. My results also implies that both granularity and non-linear terms are needed to explain the micro-originated large deviation of aggregate output.

My analysis starts with the celebrated Hulten's theorem, which states that in an competitive equilibrium, the first-order approximation of aggregate output Y_n is given by the weighted sum of microeconomic shocks, i.e.

$$Y_n = \sum_{i=1}^{n} \lambda_i \epsilon_i \tag{1}$$

where n is the number of firms, λ_i is the Domar weight (i.e. i's sales divided by GDP), and ϵ_i is productivity shock to firm i. Building on the linear model, I study the contribution of the largest firm to the variance of aggregate output as the number of firms $n \to \infty$. I show that the contribution decays very slowly as $n \to \infty$ if the distribution of firm sizes has a Pareto-like tail. This result is closely related to the finding by Gabaix (2011), which shows that the variance of aggregate output decays very slowly as $n \to \infty$ if the distribution of firm sizes has a Pareto-like tail. Indeed, I show that if the Pareto exponent of the firm size distribution, α, is $1 < \alpha < 2$, both the variance of aggregate output and the contribution of the largest firm decay at the same rate as $n \to \infty$. This result implies that the slowly decaying of the variance of aggregate output found by Gabaix (2011) is driven by that of the largest firm. In addition, I show that if firm size follows Zipf's law (i.e. $\alpha = 1$), the fraction of the contribution of the largest firm to the variance of aggregate output does not converge to 0 but to 1/2 in expectation as $n \to \infty$. This result implies that even if the number of firms is large, the largest firm accounts for a significant part of the variance of aggregate output, suggesting the importance of the granularity in terms of the variance of aggregate output.

Next, I study the macroeconomic tail probability in the linear model. Since the variance is the measure of variation around the mean, the role of the granularity may be different in the macroeconomic tail probability. In particular, I study the limiting behavior of $P(Y_n < -x)$ as $x \to \infty$ with the number of firms fixed. I consider two cases for the distribution of microeconomic shocks ϵ_i. First, I consider the case where the tail of the distribution of microeconomic shocks is lighter than an exponential function. I show that the macroeconomic tail probability is exponentially bounded from above and the bound is determined by the Domar weight of the largest firm. This result implies that if the granularity is low and the Domar weight of the largest firm is small, the macroeconomic tail probability decays very fast as $x \to \infty$. Second, I consider the case where the tail of the distribution of microeconomic shocks is heavier than an exponential function (called subexponential distributions). I show that the macroeconomic tail probabilistic is asymptotically equal to the sum of the probability of the large deviation of microeconomic shocks as $x \to \infty$, i.e. $P(Y_n < -x) \approx \sum_{i=1}^{n} P(\lambda_i \epsilon_i < -x)$. This implies that the large deviation of aggregate output is driven by a single shock to one of firms. If λ_i is highly heterogeneous, it is equivalent to saying that the large deviation of aggregate output is driven by microeconomic shocks to a small number of large firms. This result can be seen as the granular hypothesis for the macroeconomic tail probability.

I apply these results to Japanese firm data and test the empirical significance of the granularity in terms of the variance of aggregate output and macroeconomic tail probability. I calculate the standard deviation of the GDP growth rate induced by large firms. I use empirical Domar weights and the estimated standard deviation of the growth rate of firm-level productivity. I find that the standard deviation of the GDP growth rate induced by the top 100 largest firms is 0.6%. I calculate the fraction of the contribution of the largest firm to the micro-originated variance of aggregate output and find that it is 0.193. Although it is below the expected value of 1/2, it still shows the importance of the granularity. Therefore, microeconomic shocks a non-negligible factor for the standard deviation of the GDP growth rate. However, I find that the empirical granularity is too low to lead to the micro-originated large deviation of the GDP growth rate. Plugging the estimates of the distribution of firm-level productivity shocks and empirical Domar weights into the formulas obtained in Section 4, I find that the macroeconomic tail probability induced by microeconomic shocks is negligible. This implies that in the linear model, microeconomic shocks do not contribute to the large deviation of the GDP growth rate.

Finally, I study the role of the granularity in non-linear models. The non-linear models mean that the mapping f from n microeconomic shocks to aggregate output, i.e. $f : \mathbb{R}^n \mapsto \mathbb{R}$, has not only linear but non-linear terms. In the linear model, $f = \sum_i \lambda_i \epsilon_i$ and the impact of each microeconomic shock on aggregate output is summarized by the Domar weight. To consider its counterpart in the non-linear models, I assume that the impact of firm i on aggregate output is bounded by a constant c_i. Then, I obtain the upper bounds for the variance of aggregate output and the macroeconomic tail probability, which depend on $c_1, ..., c_n$. I show that without the granularity (e.g., $c_i = n^{-1}$ for all i), both upper bounds decay very fast as $n \to \infty$. This implies without the granularity, microeconomic shocks cancel out each other, and the variance of aggregate output and macroeconomic tail probability decay very fast as in the linear model. Hence, without the granularity, the diversification argument holds in non-linear models. Therefore, as in the discussion of Gabaix (2011) in the linear model, the granularity is a necessary condition of micro-originated macroeconomic fluctuations in the non-linear models.

Outline of the paper

The outline of the paper is as follows. In Section 2, I give Hulten's theorem and a multi-sector model, which forms the basis of my analysis. From Section 3 to 5, I study the role of the granularity in the linear model. In Section 3, I study the contribution of the largest firm to the variance of aggregate output. In Section 4, I study the macroeconomic tail probability. In Section 5, I test the predictions obtained in Section 3 and 4 by using Japanese data. In Section 6, I study the variance of aggregate output and the macroeconomic tail probability in the non-linear models. In Section 7, I conclude. In Appendix, I summarize proofs of my propositions.

Related literature

My paper belongs to the literature that studies the microeconomic origins of macroeconomic fluctuations (for a survey, see Carvalho (2014); Carvalho and Tahbaz-Salehi (2019)). As already mentioned, Gabaix (2011) combines Hulten's theorem and the Pareto-like tail of the empirical firm size distribution and shows that microeconomic shocks do not die out even when the number of firms is large. Acemoglu et al. (2012) consider an input-output production network and show that in the Cobb-Douglas production function and log preferences, the linear model in Hulten theorem is not only an approximation but an exact formula. In particular, an analytical expression of the Domar weights via the input-output matrix is given. Their results show that the heterogeneity of the local network structure translates microeconomic shocks into macroeconomic fluctuations. The two papers are the theoretical foundation of this literature.

My results in Section 3 are closely related to the two papers but makes a distinct point. I consider the contribution of the largest firm to the variance of aggregate output, instead of the variance of aggregate output itself, and study its limiting behavior as $n \to \infty$. I show the slowly decaying of the variance of aggregate output found by Gabaix (2011) reflects that of the contribution of the largest firm. I also study the fraction of the contribution of the largest firm to the whole micro-originated variance of aggregate output, which is equal to the ratio $\|\lambda\|_\infty^2 / \|\lambda\|_2^2$ ($\|\lambda\|_\infty$ and $\|\lambda\|_2$ are max-norm and ℓ^2-norm of Domar weights, respectively). This ratio has an important implication for the distribution of aggregate output as $n \to \infty$. Theorem 1(c) of Acemoglu et al. (2012) shows that the convergence in distribution of aggregate output to a Gaussian distribution depends on whether $\|\lambda\|_\infty/\|\lambda\|_2$ converges to 0 as $n \to \infty$ or not. My results show that on average, $\|\lambda\|_\infty/\|\lambda\|_2$ is not degenerate at 0 but strictly positive, and therefore, aggregate output does not converge in distribution to a Gaussian distribution.

Many subsequent papers extend the results of Gabaix (2011) and Acemoglu et al. (2012) in various directions. Some recent papers study an inefficient economy by considering exogenous wedges in the Cobb-Douglas production function (Jones (2011);Jones (2013);Bigio and La'o (2017);Fadinger et al. (2018)) and in a more general (e.g. CES) production function (Caliendo et al. (2018b);Baqaee and Farhi (2018a);Baqaee and Farhi (2018b)). As a further generalization, Grassi (2017) and Baqaee (2018) study an inefficient economy with endogenized wedges. As another type of inefficiency, other papers study financial frictions and their effect on macroeconomic fluctuations (Bigio and La'o (2016); Demir et al. (2017); Altinoglu (2018); Reischer (2019)). Furthermore, other features of different economic subjects are added into the model, e.g. international trade (Caliendo et al. (2018b)), policy intervention (Liu (2018)), inequality and economic growth (Fadinger et al. (2018);Jones (2013)) and firm dynamics (Carvalho and Grassi (2019)).[1]

My results in Section 4 are closely related to Acemoglu et al. (2017), which consider the limiting behavior of the macroeconomic tail probability $P(Y < -x)$ as x and $n \to \infty$. In contrast, I fix n to a constant and study the limiting behavior of $P(Y < -x)$ as $x \to \infty$ only. This results in a simple formula for $P(Y < -x)$ in terms of the tail probability of microeconomic shocks, which is empirically testable. Indeed, by using it, I find that the granularity of empirical firm sizes is too low to lead to the large deviation of aggregate output.

My results in Section 6 are related to Baqaee and Farhi (2019), which show that higher-order terms not considered in the linear model of Hulten's theorem lead to significant macroeconomic fluctuations. My results complement those of Baqaee and Farhi (2019) by showing that even in a model with higher-order terms, microeconomic shocks do not lead to macroeconomic fluctuations without the granularity. Therefore, my results combined with those in the linear model imply that both granularity and non-linear terms are needed to explain the micro-originated large deviation of aggregate output.

Notations

Here, I summarize notations used in this paper. $\sum_i$ denotes $\sum_{i=1}^n$ otherwise stated. F_x denotes the distribution of random variable x. σ_x^2 denote the variance of random variable x. I write $\overline{F}(x) := 1 - F(x)$. For (X_n), I write $X_n \xrightarrow{d} X$ if (X_n) converges in distribution to random variable X as $n \to \infty$. I write $X_n \sim a_n X$ if $X_n/a_n \xrightarrow{d} X$. 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. C denotes a constant, which may change from line to line.

2 The linear model

In this section, I present Hulten theorem and a multi-sector model that is the basis of my analysis. This model is widely used in the literature (e.g. Gabaix (2011), Acemoglu et al. (2012), Acemoglu et al. (2017), Baqaee and Farhi (2019)). I follow the exposition given in Carvalho and Tahbaz-Salehi (2019).

[1] Recent papers that study the empirical significance of the micro-originated macroeconomic fluctuations include Foerster et al. (2011);Atalay (2017);Caliendo et al. (2018a) at the sectoral level and Carvalho and Gabaix (2013);Di Giovanni et al. (2014);Acemoglu et al. (2016);Yeh (2018);Bernard et al. (2018a);Bernard et al. (2018b) at the firm level. Magerman et al. (2017) directly calculate the inverse of Leontief matrix by analyzing Belgium data. Di Giovanni et al. (2018);Di Giovanni et al. (2019);Tintelnot et al. (2019) study micro-originated macroeconomic fluctuations and an input-output production network in the context of international trade.

Consider a static economy consisting of n competitive firms and a representative household. Each firm produces a distinct product by using intermediate goods from other firms and labor.[2] Specifically, firm i produces output y_i by employing the 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 (i.e. deviation from the productivity level at the steady state), x_{ij} and l_i are the amounts of good j and labor used in the production of firm i, respectively. Let ϵ_i be the logarithm of z_i, i.e. $\epsilon_i := \ln(z_i)$ and let s_i be the total revenue of firm i (i.e. sales). In the following, I call ϵ_i the microeconomic shock to firm i. The preferences of the representative household is given by $u(c_1, ..., c_n)$, which is homogeneous of degree 1. The representative household is endowed with one unit of labor and supplies it inelastically. The definition of a competitive equilibrium is given as usual. Let GDP and GDP* be the sum of value added at an equilibrium and its steady-state value, respectively, and let $Y_n = \ln(\text{GDP}/\text{GDP}^*)$.[3] Hulten's theorem is as follows.

Theorem 1 (Hulten (1978)) *In a competitive equilibrium, the first-order macroeconomic impact of the microeconomic shock to firm i is given by*

$$\frac{d \ln GDP}{d \ln(z_i)} = \lambda_i$$

where $\lambda_i := s_i/GDP$ or the Domar weight of firm i.

Hulten's theorem implies that the first-order approximation of Y_n is given by eq.(1), i.e. the weighted sum of microeconomic shock ϵ_i with weight λ_i.

The model in Acemoglu et al. (2012) is a concrete example of Hulten's theorem (the same model is used in Acemoglu et al. (2017)). Assume further that the production function and preferences are given by

$$y_i = z_i l_i^b \prod_j x_{ij}^{a_{ij}}$$

$$u(c_1, ..., c_n) = \sum_i \beta_i \ln(c_i/\beta_i)$$

where b and a_{ij} are the shares of labor and intermediate inputs used in the production of firm i, respectively, and $\beta_i > 0$, $\sum_i \beta_i = 1$. The matrix $A := (a_{ij})$ is an input-output matrix. I assume that the labor share b is common across all firms and strictly positive.

Acemoglu et al. (2012) show that eq.(1) is not only an approximation but the exact formula, i.e., higher-order terms are equal to 0. In addition, Acemoglu et al. (2012) show that the Domar weights have two representation:

$$\lambda_i = \frac{s_i}{b \sum_i s_i} = \sum_j \beta_j \ell_{ji} \tag{2}$$

where ℓ_{ij} is the ijth element of Leontief inverse $L := (I - A)^{-1}$.[4]

I use eq.(1) and (2) in my analysis. It worth noting that the assumption that $\epsilon_1, ..., \epsilon_n$ are independent of each other is fundamentally important. This assumption is the direct consequence of the aim of the paper, i.e. I test whether microeconomic idiosyncratic shocks lead to macroeconomic

[2] While Hulten's theorem and my results in this paper can be extended to an economy with m primary factors, I assume for simplicity that firms use labor only as a primary factor.

[3] To show the dependence on n, I add the subscript n to Y.

[4] The existence of the inverse of $(I - A)$ is guaranteed by the assumption $b > 0$.

fluctuations. Hence, I exclude the comovement of microeconomic shocks induced by, e.g. common aggregate shocks. In the following, I show that the independence assumption is powerful enough to obtain many useful results about macroeconomic fluctuations.

3 The variance of aggregate output in the linear model

In this section, I study how the granularity affects the variance of aggregate output in the linear model, eq.(1). I use the extreme value theory (see Embrechts et al. (1997)) and study the limiting behavior of the contribution of large firms to the variance of aggregate output as $n \to \infty$.

3.1 Results by Gabaix (2011)

To begin with, I review the results given by Gabaix (2011). Recall eq.(1) and let $\sigma^2_{Y_n}$ denote the variance of Y_n. Let $\epsilon_1, ..., \epsilon_n$ be a sequence of iid random variables with the finite variance $\sigma^2_\epsilon < \infty$. Eq.(1) implies that σ^2_Y is given by

$$\sigma^2_{Y_n} = \sigma^2_\epsilon \|\lambda\|^2_2 \tag{3}$$

where $\|\lambda\|_2$ is the ℓ^2-norm of the Domar weights, i.e.

$$\|\lambda\|_2 := \left(\sum_i \lambda_i^2\right)^{1/2} = \left(\frac{\sum_i s_i^2}{b^2(\sum_i s_i)^2}\right)^{1/2}$$

Here, I use eq.(2). Assume that firm sizes $s_1, ..., s_n$ are iid random variables with distribution F_s. Since $\|\lambda\|_2$ is a function of $s_1, ..., s_n$, $\|\lambda\|_2$ is also a random variable and its distribution is determined by F_s. For example, consider the homogeneous case where F_s is degenerate at s^*, i.e. s_i is equal to some constant s^* for all i. Then

$$\|\lambda\|^2_2 = \frac{\sum_i s_*^2}{(b\sum_i s_*)^2} = \frac{1}{b^2 n}$$

It shows that $\sigma^2_{Y_n}$ decays at the rate of n^{-1}.

Gabaix (2011) studies the limiting behavior of $\sigma^2_{Y_n}$ as $n \to \infty$, which is equivalent to that of $\|\lambda\|^2_2$, and shows that if F_s has a Pareto-like tail, $\|\lambda\|^2_2$ decays much slower than n^{-1}.

Theorem 2 (Gabaix (2011)) *Let $s_1, ..., s_n$ be a sequence of iid random variables with distribution F_s. Suppose that F_s has a Pareto-like tail with $\alpha \geq 1$, i.e.*

$$\overline{F}_s(x) \sim Kx^{-\alpha} \quad \text{as } x \to \infty, \ \alpha > 1 \tag{4}$$

where K is a constant. Then $\sigma^2_{Y_n}$ satisfies the following relation:

$$\sigma^2_{Y_n} \sim \frac{v_\alpha}{(\log n)^2}\sigma^2_\epsilon \quad \text{as } n \to \infty \quad \text{for } \alpha = 1$$
$$\sigma^2_{Y_n} \sim \frac{v_\alpha}{n^{2-2/\alpha}}\sigma^2_\epsilon \quad \text{as } n \to \infty \quad \text{for } 1 < \alpha < 2$$
$$\sigma^2_{Y_n} \sim \frac{v_\alpha}{n}\sigma^2_\epsilon \quad \text{as } n \to \infty \quad \text{for } \alpha \geq 2$$

where v_α is a non-degenerate random variable, which is independent of n.

My question is as follows: Why does $\sigma^2_{Y_n}$ decay very slowly? What is the role of the largest firm in the slowly decaying of $\sigma^2_{Y_n}$? To answer it, I study the contribution of the largest firm to $\sigma^2_{Y_n}$ defined by

$$\sigma^2_{\max,n} := \sigma^2_\epsilon \|\lambda\|^2_\infty = \frac{\sigma^2_\epsilon}{b^2} \frac{s^2_{\max,n}}{(\sum_i s_i)^2}$$

where $\|\lambda\|_\infty := \max_i(\lambda_i)$ is the max-norm of the Domar weights and $s_{\max,n} := \max_i(s_i)$ is the sales of the largest firm.

Consider again the homogeneous case (i.e. $s_i = s_*$ for all i). In this case, $s^2_{\max,n} = s^2_*$ and $\|\lambda\|^2_\infty = (bn)^{-2}$, showing that $\sigma^2_{\max,n}$ decays at the rate of n^{-2}. But if F_s is non-degenerate, $s^2_{\max,n}$ increases as n increases.[5] I study how fast $s^2_{\max,n}$ increases as $n \to \infty$, which determines the decaying rate of $\|\lambda\|^2_\infty$, and therefore, that of $\sigma^2_{\max,n}$.

3.2 Exponential tail

For notational convenience, I consider the distribution of sales squared F_{s^2} instead of the distribution of firm size F_s. First, as an example of light tailed distributions, consider an exponential distribution as F_{s^2}:[6] for $x \geq 0$[7]

$$\overline{F}_{s^2}(x) = e^{-x}$$

Since the event $\{s^2_{\max,n} < x\}$ means that all of $s^2_1, ..., s^2_n$ is less than x, the probability of the largest s^2_i is given by

$$P(s^2_{\max,n} < x) = (1 - e^{-x})^n$$

This implies that the density $d_n(x) = n(1-e^{-x})^{n-1}e^{-x}$ and, by taking the derivative of the density and setting it to be 0, the mode $\widetilde{x}_n = \ln n$. Hence, $s^2_{\max,n}$ increases as the rate $\ln n$ as n increases. The increasing rate of $s^2_{\max,n}$ can be confirmed directly: for $x \in \mathbb{R}$,

$$\begin{aligned} P(s^2_{\max,n} - \ln n \leq x) &= \big(P(s^2_i \leq x + \ln n)\big)^n \\ &= (1 - n^{-1}e^{-x})^n \to \exp(-e^{-x}) \ \text{ as } n \to \infty \end{aligned}$$

This means that $(s^2_{\max,n} - \ln n)$ converges in distribution to a non-degenerate random variable which is independent of n.[8] Hence, $s^2_{\max,n}$ increases at the rate of $\ln n$ as n increases.

Compare the increasing rate of $s^2_{\max,n}$ to that of the sum $\sum_i s^2_i$. The law of large numbers implies that $\sum_i s^2_i$ increases at the rate of n as n increases. Hence, $s^2_{\max,n}$ increases more slowly

[5] One can show that $s_{\max,n} \overset{\text{a.s.}}{\to} \infty$ if the right-end point of $F_s = \infty$. The definition of $s^2_{\max,n}$ implies that given a fixed $n \in \mathbb{N}$,

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

for $x \in \mathbb{R}$. Hence, for any $x < \infty$, $P(s^2_{\max,n} \leq x) = F^n_{s^2}(x) \to 0$, which implies that $s^2_{\max,n} \overset{P}{\to} \infty$. Since $(s^2_{\max,n})$ is a non-decreasing sequence, I obtain $s^2_{\max,n} \overset{\text{a.s.}}{\to} \infty$.

[6] One can show that if F_s is a Gaussian distribution, the tail of F_{s^2} is upper-bounded by an exponential function. Mill's ratio implies that for the standard Gaussian distribution F with density f,

$$\frac{\overline{F}(x)}{f(x)} \sim \frac{1}{x}$$

Imagine that F_s has a Gaussian tail. Since $s_i > 0$, Mill's ratio implies that

$$P(s^2_i > x) = P(s_i > \sqrt{x}) \sim \frac{1}{\sqrt{2\pi x}} e^{-x/2} \leq e^{-x/2}$$

This shows that the tail of F_{s^2} is bounded by an exponential function.

[7] If needed, assume that s^2_i is properly normalized by $c_1^{-1}(s^2_i - c_2)$ for some constants $c_1, c_2 > 0$.

[8] The distribution $\exp(-e^{-x})$ is called the Gumbel distribution in the extreme value theory.

than $\sum_i s_i^2$ as n increases. Roughly speaking, in the case of the exponential distribution of F_{s^2}, $s_1^2, ..., s_n^2$ are quite similar to each other and no extremely large one emerges. Therefore, the case of the exponential distribution is similar to the homogeneous case (i.e. $s_i = s_*$ for all i).

Formally, I obtain the following result.

Proposition 1 *Let $s_1^2, ..., s_n^2$ be a sequence of iid random variables with distribution F_{s^2}. Suppose that F_{s^2} has an exponential-like tail with parameter α, i.e.*

$$\overline{F}_{s^2} \sim K e^{-\alpha x} \quad \text{as } x \to \infty$$

where K is a constant. Then

$$\sigma^2_{\max,n} \sim C \frac{\ln n}{n^2} \quad \text{as } n \to \infty$$

where C is a constant, which is independent of n.

Proof. See the Appendix. ■

Similar to the homogeneous case, this shows that the impact of the largest firm decays very fast as n increases.

3.3 Pareto tail

Next, consider as F_{s^2} the distribution having a Pareto-like tail. I show that $\sigma^2_{\max,n}$ exhibits a very slowly decaying compared to the exponential case. Suppose that for $s_i^2 \geq 1, k > 0$[9]

$$\overline{F}_{s^2} = x^{-k}$$

Note a useful relation between exponential and Pareto distributions. Let ξ be a positive random variable drawn from a Pareto distribution with parameter k and let $\eta := \ln \xi$. Then

$$P(\eta > y) = P(\ln \xi > \ln x) = P(\xi > x) = x^{-k} = e^{-k \ln x} = e^{-ky}$$

where $x := e^y$. This shows that the distribution of η is exponential with parameter k. Furthermore, the discussion in the previous subsection shows that if $\eta_1, ..., \eta_n$ are iid random variables with an exponential distribution with parameter k, then $\eta_{\max,n}/(k^{-1} \ln n) \to 1$. This shows that if $\xi_1, ..., \xi_n$ are iid random variables with the Pareto distribution with parameter k, then $\xi_{\max,n} = e^{\eta_{\max,n}}$ increases at rate $n^{1/k}$. Therefore, if $\overline{F}_{s^2}$ is given as above, $s^2_{\max,n}$ increases at the rate of $n^{1/k}$.

This increasing rate of $s^2_{\max,n}$ can be confirmed directly:

$$\begin{aligned} P(n^{-1/k} s^2_{\max,n} \leq x) &= P(s^2_{\max,n} \leq n^{1/k} x) \\ &= \left(1 - \overline{F}_{s^2}(n^{1/k} x)\right)^n \\ &= \left(1 - \frac{1}{n} \frac{1}{x^k}\right)^n \to \exp(-x^{-k}) \quad \text{as } n \to \infty \end{aligned}$$

This shows that $n^{-1/k} s^2_{\max,n}$ converges in distribution to a non-degenerate random variable, which is independent of n.[10] Hence, $s^2_{\max,n}$ increases at the rate of $n^{1/k}$ as n increases.

The increasing rate of $s^2_{\max,n}$ is much higher than in the exponential case. Roughly speaking, if F_{s^2} has the Pareto tail, the heterogeneity of s_i^2 is high and an extremely large s_i^2 emerges as

[9] If needed, assume that s_i^2 is properly normalized by $c_1^{-1} s_i^2$ for some c_1.

[10] The distribution $\exp(-x^{-k})$ is called the Frechet distribution in the extreme value theory.

distribution tail	$\sigma^2_{Y_n}$	$\sigma^2_{\max,n}$
exponential-like	n^{-1}	$n^{-2} \cdot \ln n$
Pareto-like with $\alpha > 2$	n^{-1}	$n^{-2+2/\alpha}$
Pareto-like df with $1 < \alpha < 2$	$n^{-2+2/\alpha}$	$n^{-2+2/\alpha}$

Table 1: Decaying rate of $\sigma^2_{Y_n}$ and $\sigma^2_{\max,n}$

n increases. This implies that the impact of the largest firm does not die out even when n is sufficiently large. I obtain the following result.

Proposition 2 *Let $s_1, ..., s_n$ be a sequence of iid random variables with distribution F_s. Suppose that F_s has a Pareto-like tail with exponent $\alpha > 1$, i.e.*

$$\overline{F}_s(x) \sim Kx^{-\alpha} \quad \text{as } x \to \infty,\ \alpha > 1$$

where K is a constant. Then

$$\sigma^2_{\max,n} \sim C \frac{u_{\alpha/2}}{n^{2-2/\alpha}} \quad \text{as } n \to \infty$$

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

Proof. See the Appendix. ∎

Having established the results of $\sigma^2_{\max,n}$, we are in a position to compare them with those of Gabaix (2011) (Table 1). If F_{s^2} is an exponential distribution (i.e. a light-tailed distribution), the decaying rate of $\sigma^2_{Y_n}$ and $\sigma^2_{\max,n}$ are given by n^{-1} and $n^{-2} \cdot \ln n$, respectively. Compared to $\sigma^2_{Y_n}$, $\sigma^2_{\max,n}$ decays very fast, and therefore, the role of the largest firm becomes negligible when n is sufficiently large. The same statement holds for distributions having a Pareto-like tail with exponent $\alpha > 2$. In this case, the decaying rate of $\sigma^2_{Y_n}$ and $\sigma^2_{\max,n}$ are given by n^{-1} and $n^{-2+2/\alpha}$, respectively, and therefore, the role of the largest firm becomes negligible when n is sufficiently large. However, when the Pareto exponent is less than 2, both $\sigma^2_{Y_n}$ and $\sigma^2_{\max,n}$ decays at the same rate of $n^{-2+2/\alpha}$. Hence, compared to $\sigma^2_{Y_n}$, $\sigma^2_{\max,n}$ does not become negligible even when n is sufficiently large. Note that $\sigma^2_{\max,n}$ is a part of $\sigma^2_{Y_n}$, which means that $\sigma^2_{Y_n}$ does not decay faster than $\sigma^2_{\max,n}$ as $n \to \infty$. Therefore, when F_s has a Pareto-like tail with exponent $\alpha < 2$, $\sigma^2_{\max,n}$, the contribution of the largest firm, determines the slowly decaying of $\sigma^2_{Y_n}$.

3.4 Ratio of $\sigma^2_{\max,n}$ to $\sigma^2_{Y_n}$

Motivated by the results in the previous subsection, I consider the ratio of $\sigma^2_{\max,n}$ to $\sigma^2_{Y_n}$ and its limiting behavior as $n \to \infty$. Let r_n be the ratio. By eq.(2), I obtain

$$r_n := \frac{\sigma^2_{\max,n}}{\sigma^2_{Y_n}} = \frac{\|\lambda\|^2_\infty}{\|\lambda\|^2_2} = \frac{s^2_{\max,n}}{\sum_i s_i^2}$$

Namely, the ratio r_n is equal to that of the max-norm to the ℓ^2-norm of the Domar weights, and to that of the maximum $s^2_{\max,n}$ to the sum $\sum_i s_i^2$. For example, consider again the case where F_{s^2} is exponential. As discussed above, the numerator $s^2_{\max,n}$ increases at the rate of $\ln n$, while the law of large number shows that the denominator increases at the rate of n. This means that r_n converges in probability to 0 as $n \to \infty$, i.e., no impact of the largest firm on $\sigma^2_{Y_n}$. As shown below, this property holds under more general conditions.

Proposition 3 *Let $s_1, ..., s_n$ be a sequence of iid random variables with distribution F_s. Suppose that the variance of F_s is finite. Then*

$$r_n \overset{a.s.}{\to} 0$$

Proof. See the Appendix ■

Note that since the condition of Proposition 3 is equivalent to the existence of the finite mean of F_{s^2}, it includes the cases of the exponential tail and Pareto-like tail with exponent $\alpha > 2$ considered in the previous section. Therefore, under the condition of the finite variance of F_s, $\sigma^2_{\max,n}$ decays faster than $\sigma^2_{Y_n}$, and therefore, the role of the largest firm becomes negligible as $n \to \infty$.

As suggested in the previous section, r_n for F_s having a Pareto tail with exponent $0 < \alpha < 2$ exhibits a different limiting behavior.

Proposition 4 *Let $s_1, ..., s_n$ be a sequence of iid random variables with distribution F_s Suppose that F_s has a Pareto-like tail with exponent $0 < \alpha < 2$. Then r_n converges in distribution to a non-degenerate random variable as $n \to \infty$. In particular, if $\alpha = 1$ (Zipf's law), then*

$$\lim_{n\to\infty} E[r_n] = \frac{1}{2}$$

Proof. See the Appendix. ■

This shows that r_n does not necessarily converges to 0 but the limit is a non-degenerate random variable. Namely, the impact of the largest firm does not die out even when n is sufficiently large. It is consistent with the result that for F_s having a Pareto-like tail with exponent $1 < \alpha < 2$, both $\sigma^2_{\max,n}$ and $\sigma^2_{Y_n}$ decay at the same rate. In particular, if $\alpha = 1$ (Zipf's law), the contribution of the largest firm accounts for the half of $\sigma^2_{Y_n}$ in expectation. These results clearly show the importance of the granularity for macroeconomic fluctuations measured by $\sigma^2_{Y_n}$.

The mechanism behind these results is explained by the limiting behavior of $s^2_{\max,n}$ and $\sum_i s_i^2$ (Figure 1). For the case of the finite variance of F_s including the case of the exponential tail and Pareto-like tail with exponent $\alpha > 2$, the law of large numbers implies that $\sum_i s_i^2$ increases at the rate of n, while the increasing rate of $s^2_{\max,n}$ is lower than n. Hence, the fraction of $s^2_{\max,n}$ in $\sum_i s_i^2$ converges to 0 as $n \to \infty$. On the other hand, for the case of a Pareto tail with exponent $0 < \alpha < 2$, the law of large number for $\sum_i s_i^2$ does not hold. Indeed, the increasing rate of $s^2_{\max,n}$ is higher than n and determines that of $\sum_i s_i^2$. Hence, the fraction of $s^2_{\max,n}$ to $\sum_i s_i^2$ does not converges to 0 even when n is sufficiently large.

My result about r_n has a close relation to Acemoglu et al. (2012) and Acemoglu et al. (2017). Recall the liner model of eq.(1). Theorem 1 in Acemoglu et al. (2012) shows that the convergence in distribution of $Y_n/\|\lambda\|_2$ depends on whether $\|\lambda\|_\infty/\|\lambda\|_2$ converges to 0. In particular, they show that if the iid microeconomic shocks $\epsilon_1, ..., \epsilon_n$ are not Gaussian, $Y_n/\|\lambda\|_2$ does not converge in distribution to Gaussian, i.e. the central limit theorem does not hold. Combined with my result and the Zipf's law, it implies that we cannot apply the central limit theorem to the macroeconomic fluctuations. The ratio r_n is also related to the sectoral dominance δ in Acemoglu et al. (2017) defined by[11]

$$\delta := \frac{\|\lambda\|_\infty}{\|\lambda\|_2/\sqrt{n}}$$

[11] In Acemoglu et al. (2017), shocks at the sectoral level are considered as microeconomic ones.

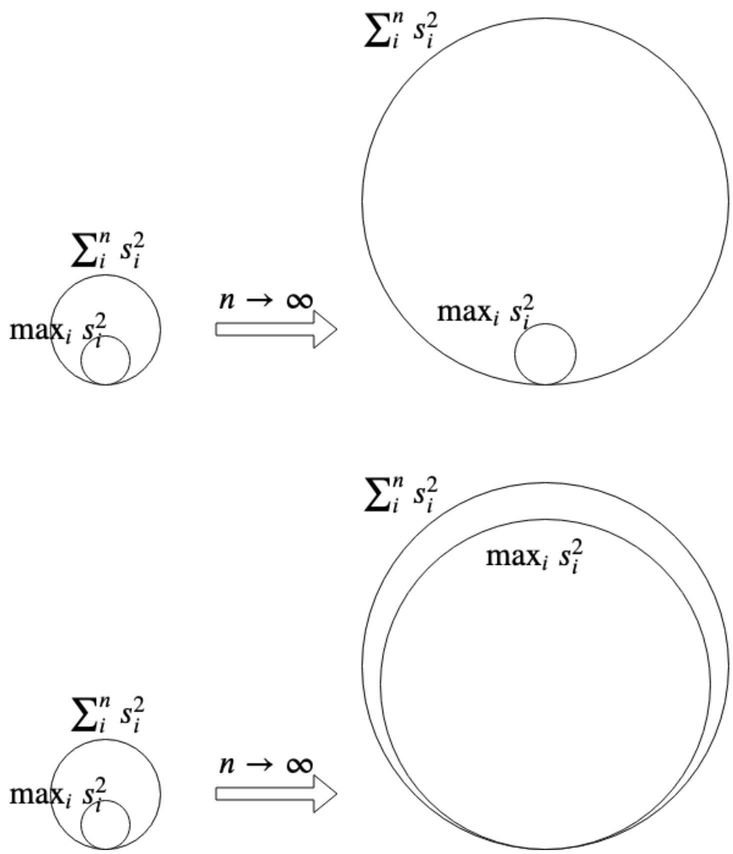

Figure 1: limiting behavior of $s^2_{\max,n}$ and $\sum_i s_i^2$

Acemoglu et al. (2017) show that if the distribution of microeconomic shock ϵ_i has an exponential tail, the existence of macroeconomic tail risk is equivalent to the condition $\lim_{n\to\infty}\delta = \infty$.[12] Since my results shows that $\lim_{n\to\infty} E[\|\lambda\|_\infty/\|\lambda\|_2]$ is strictly positive, $\lim_{n\to\infty} E[\delta]$ diverges, and therefore, the macroeconomic tail risk exists.

3.5 Upper order statistics

In the previous sections, I studied the contribution of the larges firm to $\sigma^2_{Y_n}$. Here, I study the contribution of the mth largest firm to $\sigma^2_{Y_n}$, which is denoted by $\sigma^2_{m,n}$. I show that under the condition of Proposition 2, $\sigma^2_{m,n}$ decays at the same rate of $\sigma^2_{\max,n}$.

As before, let $s_1^2, ..., s_n^2$ be a sequence of iid random variables and let $s^2_{n,n} \leq ... \leq s^2_{1,n}$ denote the ordered sample of $s_1^2, ..., s_n^2$. Hence, $s^2_{1,n} = s^2_{\max,n}$ and $s^2_{m,n}$ is the mth upper order statistics of the sample.

Proposition 5 *Suppose that the conditions in Proposition 2 holds. Then for fixed $m \in \mathbb{N}_+$,*

$$\sigma^2_{m,n} \sim C\frac{v_{\alpha,m}}{n^{2-2/\alpha}} \quad \textit{as } n \to \infty$$

where C is a constant and $v_{\alpha/2,m}$ is a non-degenerate random variable, which depends on $\alpha/2, m$ but not on n.

Proof. See the Appendix. ■

As in Proposition 2, if F_s has a Pareto-like tail with exponential $1 < \alpha < 2$, then $\sigma^2_{m,n}$ decays at the same rate as σ^2_Y as $n \to \infty$. The contribution of the mth (or the sum of the first to mth) largest firms does not become negligible but accounts for a significant part of $\sigma^2_{Y_n}$ even when n is sufficiently large. Note that since m is fixed, m/n becomes arbitrarily small as $n \to \infty$. This result implies that as $n \to \infty$, extremely large firms emerge one after another, and therefore, the small number of large firms have disproportional impacts on aggregate output. Following this result, I calculate the contribution of the top n_{top} firms to $\sigma^2_{Y_n}$, i.e. $\sum_{m=1}^{n_{\text{top}}}\sigma^2_{m,n}$ in Section 5.

4 Macro tail probability

The variance $\sigma^2_{Y_n}$ studied in the previous section is the variation of Y_n around its mean. In this section, I study the tail probability of Y_n and its relation to the granularity. I consider light-tailed and heavy-tailed distributions of microeconomic shocks F_ϵ, separately.

4.1 Setup

First, I review Acemoglu et al. (2017), which is closely related to my analysis. Recall the linear model in eq.(1) again. Let $\epsilon_1, ..., \epsilon_n$ be a sequence of iid random variables with distribution F_ϵ. Acemoglu et al. (2017) consider the measure of the macroeconomic tail risk $R_n(x_n)$ defined by

$$R_n(x_n) := \frac{\ln P\left(\frac{Y_n}{\sigma_{Y_n}} < -x_n\right)}{\ln \Phi(-x_n)}$$

where Φ is the standard Gaussian distribution and (x_n) is a sequence of positive numbers such that $x_n \to \infty$ as $n \to \infty$. The economy exhibits macroeconomic tail risks if $R_n(x_n) = 0$. Acemoglu

[12] For the definition of the macroeconomic tail risk, see the next section.

et al. (2017) show that even if Y_n/σ_{Y_n} converges in distribution to Gaussian, the convergence can be slow at the tail region, and therefore, the economy can exhibit the macroeconomic tail risks.

Note that two limits are considered in $R_n(x_n)$: the number of firms n and tail region $(-\infty, -x_n)$. I employ a simpler strategy, in which n is fixed (Domar weight λ_i is also fixed) and only the limit of x is considered. This has two advantages. First, by directly studying the macroeconomic tail probability $P(Y_n < -x)$, I can obtain a simple formula showing the role of the granularity in $P(Y_n < -x)$. Second, by plugging the empirical Domar weights into the formula, I can estimate the macroeconomic tail probability quantitatively. In the following, I assume that n is large but fixed. I assume further that F_ϵ is symmetric and zero mean. Since the symmetry means $P(Y_n > x) = P(Y_n < -x)$, I consider $P(Y_n > x)$ for notational simplicity.

4.2 Light-tailed distribution

Here, I study two light-tailed distributions as F_ϵ. The first example is a Gaussian distribution, i.e. $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ for all i. The independence of $\epsilon_1, ..., \epsilon_n$ implies $Y_n \sim \mathcal{N}(0, \|\lambda\|_2^2\sigma_\epsilon^2)$. By Mill's ratio, I obtain

$$P(Y_n \geq x) \sim \frac{1}{x}\frac{1}{\|\lambda\|_2\sigma_\epsilon\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\|\lambda\|_2^2\sigma_\epsilon^2}\right)\right) \leq C\exp\left(-\frac{1}{2}\left(\frac{x^2}{\|\lambda\|_2^2\sigma_\epsilon^2}\right)\right) \tag{5}$$

for some constant C. This shows that the tail of $P(Y_n \geq x)$ is upper bounded by the Gaussian decay of $\exp\left(-C\frac{x^2}{\|\lambda\|_2^2}\right)$ for some constant C, which depends on $\|\lambda\|_2^2$. In the homogeneous case (i.e. $s_i = s^*$ for all i), the upper bound becomes $\exp(-Cnx^2)$ for some constant C, showing the very fast decay as $x \to \infty$.

Next, consider as F_ϵ distributions whose tail is upper bounded by an exponential function: There exists some $x^* \geq 0$ such that for $x \geq x^*$,

$$\overline{F}_\epsilon(x) \leq Ke^{-\beta x}$$

where K is a constant. This family of distributions includes ones whose tail is heavier than Gaussian but not than exponential distributions. I have the following proposition.

Proposition 6 *Suppose that the condition above holds. There exists $b, \nu \geq 0$ such that*

$$P(Y_n \geq x) \leq \begin{cases} \exp\left(-\frac{x^2}{2\nu^2\|\lambda\|_2^2}\right) & \text{for } 0 \leq x \leq \frac{\nu^2\|\lambda\|_2^2}{b\|\lambda\|_\infty} \\ \exp\left(-\frac{x}{2b\|\lambda\|_\infty}\right) & \text{for } x > \frac{\nu^2\|\lambda\|_2^2}{b\|\lambda\|_\infty} \end{cases}$$

Proof. See the Append. ■

For the meaning of b and ν, see the Appendix. It shows that the upper bound exhibits a Gaussian decay around the origin and an exponential decay in the tail region. In particular, the exponential decay depends on $\|\lambda\|_\infty$. In the homogeneous case, the upper bound becomes $\exp(-Cnx)$ for some constant C, showing the fast decay as $x \to \infty$. In both cases of the Gaussian distribution and exponential tail bound, the upper bounds of macroeconomic tail probability show the importance of the granularity represented by $\|\lambda\|_\infty$ or $\|\lambda\|_2^2$. Therefore, as x increases, the macroeconomic tail probability decays very fast without the granularity.

4.3 Subexponential distribution

In the remainder of this section, I consider F_ϵ having a heavier tail than an exponential one (see Foss et al. (2011)). Let F_ϵ has a right-unbounded support, i.e. $\overline{F}_\epsilon(x) > 0$ for all x. I give the definition of the heavy-tailedness of distributions:

Definition 1 *The distribution F_ϵ on $\mathbb{R}$ is heavy-tailed if for any $\delta > 0$,*

$$e^{\delta x}\overline{F}_\epsilon(x) \to \infty, \quad \text{as } x \to \infty$$

This means that the tail of F_ϵ decays slower than any exponential functions of type $e^{\delta x}$. In other words, distributions whose tail is upper bounded by an exponential function are not heavy-tailed. An important example of the heavy-tailed distributions is a Pareto distribution.

4.3.1 For distributions on $\mathbb{R}_+$

Before studying the heavy-tailed distribution in general settings, consider ones restricted on the positive half line $\mathbb{R}^+ := (0, \infty)$. Let F_ϵ be a Pareto distribution on $\mathbb{R}^+$, i.e. for $x > 1$

$$\overline{F}_\epsilon(x) = x^{-\beta}, \ \beta > 0$$

F_ϵ has the following property:

Example 1 *Let F_ϵ be given as above. Then*

$$\overline{F_\epsilon * F_\epsilon}(x) \sim 2x^{-\beta} = 2\overline{F}_\epsilon(x)$$

where $*$ denotes the convolution of distributions.

Proof. See the Appendix. ■

This property has an important relation to macroeconomic tail probability. Note that for a general distribution F_ϵ and any $n \in \mathbb{N}^+$,

$$\begin{aligned} P\left(\max_i(\epsilon_i) > x\right) &= \overline{F_\epsilon^n}(x) \\ &= (1 - F_\epsilon^n(x)) \\ &= (1 - F_\epsilon(x))(1 + F_\epsilon(x) + F_\epsilon^2(x) + ... + F_\epsilon^{(n-1)}(x)) \\ &\sim n\overline{F}_\epsilon(x), \ x \to \infty \end{aligned}$$

Hence, the property means that if F_ϵ is a Pareto distribution, the probability of the large deviation of the sum $\epsilon_1 + \epsilon_2$ is asymptotically equal to that of the large deviation of either ϵ_1 or ϵ_2:

$$P(\epsilon_1 + \epsilon_2 > x) \sim P(\max(\epsilon_1, \epsilon_2) > x)$$

Since $\epsilon_1, \epsilon_2 > 0$ implies $\{\omega : \max(\epsilon_1, \epsilon_2) > x\} \subset \{\omega : \epsilon_1 + \epsilon_2 > x\}$, it is equivalent to

$$P(\epsilon_1 + \epsilon_2 > x, \max(\epsilon_1, \epsilon_2) \leq x) = o(\overline{F}_\epsilon(x))$$

Roughly speaking, the contribution by two medium size shocks to $P(\epsilon_1 + \epsilon_2 > x)$ is asymptotically negligible.

This property is related to the heavy-tailedness of the Pareto distribution. Indeed, the following theorem holds:

Theorem 3 ([Theorem 2.12 in Foss et al. (2011)) *Suppose that F_ϵ on $\mathbb{R}^+$ is heavy-tailed. Then*

$$\liminf_{x \to \infty} \frac{\overline{F_\epsilon * F_\epsilon}(x)}{\overline{F}_\epsilon(x)} = 2$$

I assume the regularity of the tail, i.e. the existence of the limit $\lim_{x\to\infty} \frac{\overline{F_\epsilon * F_\epsilon}(x)}{\overline{F}_\epsilon(x)}$. Therefore, if F_ϵ is heavy-tailed and has the limit, $\lim_{x\to\infty} \frac{\overline{F_\epsilon * F_\epsilon}(x)}{\overline{F}_\epsilon(x)} = 2$, which leads to the definition of subexponential distributions.

Definition 2 *A distribution F_ϵ on $\mathbb{R}^+$ is subexponential (denoted by $F_\epsilon \in \mathcal{S}$) if*

$$\overline{F_\epsilon * F_\epsilon}(x) \sim 2\overline{F}_\epsilon(x)$$

4.3.2 For distributions on $\mathbb{R}$

Next, I consider subexponential distributions on $\mathbb{R}$. For a distribution F_ϵ on $\mathbb{R}$, define $F^+(x)$ by

$$F_\epsilon^+(x) := \begin{cases} F_\epsilon(x) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Let $\epsilon^+ := \max(\epsilon, 0)$. If the distribution of ϵ is F_ϵ, that of ϵ^+ is F_ϵ^+.

Definition 3 *A distributions F_ϵ on $\mathbb{R}$ is subexponential (denoted by $F_\epsilon \in \mathcal{S}_\mathbb{R}$) if F^+ on $\mathbb{R}^+$ is subexponential.*

As expected, $F_\epsilon \in \mathcal{S}_\mathbb{R}$ implies[13] that

$$\overline{F_\epsilon * F_\epsilon}(x) \sim 2\overline{F}_\epsilon(x)$$

The defining property of the subexponential distributions extend to the case of arbitrary n convolution. A subexponential distribution F_ϵ on $\mathbb{R}$ satisfies

$$\overline{F_\epsilon^{*n}}(x) \sim n\overline{F}_\epsilon(x)$$

for arbitrary $n \in \mathbb{N}^+$.[14]

As in the case of the Pareto distribution, this property means that the large deviation of the sum is driven by the large deviation of a single random variable, i.e. if $\epsilon_1, ..., \epsilon_n$ are a sequence of iid random variables with distribution $F_\epsilon \in \mathcal{S}_\mathbb{R}$, then

$$P\left(\sum_i \epsilon_i > x\right) \sim P\left(\max_i(\epsilon_i) > x\right), \ x \to \infty$$

In the following, I extend this property to the weighted sum of iid random variables.

4.4 Proposition

I relabel the index i such that $\lambda_n \leq ... \leq \lambda_1$. The Domar weight of the largest firm is λ_1. I obtain the main result of this section.

Proposition 7 *Suppose that $\epsilon_1, ..., \epsilon_n$ are a sequence of iid random variables with $F_\epsilon \in \mathcal{S}_\mathbb{R}$. Then*

$$P\left(Y_n > x\right) = \sum_i P(\lambda_i \epsilon_i > x) + o(P(\lambda_1 \epsilon_1 > x))$$

[13] I need to give a concise proof.

[14] This is the direct consequence of Proposition 7. See the Appendix.

Proof. See the Appendix. ■

This shows that the large deviation of Y_n is driven by the large deviation of a single term $\lambda_i\epsilon_i$. Especially when some λ_i are very large, the tail probability of $\lambda_i\epsilon_i$ with such λ_i directly contributes to the macroeconomic tail probability. This is a formulation of the granular hypothesis for the macroeconomic tail probability. By using this formula, we can explicitly calculate $\Pr(Y_n > x)$ given the estimates of $\overline{F}_\epsilon$. I give two examples below.

Suppose that F_ϵ has a Pareto-like tail, i.e.

$$\overline{F}_\epsilon(x) \sim Kx^{-\beta},\ \beta > 0 \tag{6}$$

For each i,

$$P(\lambda_i\epsilon_i > x) \sim K\left(\frac{x}{\lambda_i}\right)^{-\beta}$$

Proposition 7 implies

$$P(Y_n > x) \sim \left(\sum_i \lambda_i^\beta\right)\overline{F}_\epsilon(x) \tag{7}$$

This implies that the macroeconomic tail probability $P(Y_n > x)$ are linked to the tail probability of microeconomic shocks $\overline{F}_\epsilon(x)$ by the multiplier factor $\left(\sum_i \lambda_i^\beta\right)$. Given the estimates of the parameter β and $\overline{F}_\epsilon(x)$, I can quantitatively estimate the micro-originated macroeconomic fluctuations. This will be done in the next section.

As the second example, suppose that F_ϵ has a Weibull-like tail, i.e.

$$\overline{F}_\epsilon(x) \sim Ke^{-\beta x^\tau},\ \beta > 0, 0 < \tau < 1 \tag{8}$$

One can show that this distribution is subexponential (see, e.g. Example 1.4.3 in Embrechts et al. (1997)). Note that the tail becomes exponential when $\tau = 1$. Hence, the exponential tail can be seen as the limit of the Weibull-like tail as $\tau \to 1$. Since $P(\lambda_i\epsilon_i > x) \sim Ke^{-\beta\left(\frac{x}{\lambda_i}\right)^\tau}$ for each i, I have

$$\frac{P(\lambda_i\epsilon_i > x)}{P(\lambda_1\epsilon_1 > x)} \sim e^{\beta(\lambda_1^{-\tau} - \lambda_i^{-\tau})x^\tau}$$

If the largest Domar weight is strictly large than the others, i.e. $\lambda_1 > \lambda_i$ for $i = 2, ..., n$, this ratio converges to 0 as $x \to \infty$. Hence, Proposition 7 implies that if $\lambda_1 > \lambda_i$ for $i = 2, ..., n$, then

$$P(Y_n > x) \sim P(\lambda_1\epsilon_1 > x) \sim Ke^{-\beta\left(\frac{x}{\lambda_1}\right)^\tau}$$

This implies that if F_ϵ has a Weibull-like tail, only the term of the largest firm contributes to the macroeconomic tail probability as $x \to \infty$. The contribution of other firms on the macroeconomic tail probability becomes negligible as $x \to \infty$. It clearly shows the importance of the granularity in the macroeconomic tail probability.

5 Empirical data in Japan

In this section, I apply the formulas obtained in Section 3 and 4 to Japanese data. I show that the empirical granularity is important for $\sigma_{Y_n}^2$ but not high enough to lead to the micro-originated large deviation of aggregate output.

var.	# of obs.	mean	1st Q.	3rd Q.	max
s_i	10024	69.98	14.20	46.90	11476
L_i	10024	823	161	708	161081
λ_i	10024	1.28×10^{-4}	2.61×10^{-5}	8.60×10^{-5}	0.0211

var.	# of obs.	mean	sd
Δs_i	10024	0.0110	0.140
ΔL_i	10024	0.00913	0.227
$\hat{\epsilon}_i$	10024	0.00965	0.137

Table 2: Summary statistics. The unit of s_i is 1 billion yen.

5.1 Data

I use firm-level data in Japan for 2016-2017 provided by Tokyo Shoko Research (TSR), which is a credit rating agency. The data is comprehensive and includes listed and unlisted firms across all sectors.[15] Since the aim of my analysis is the granularity, I exclude firms whose sales are less than 10 billion yens. I also exclude firms in sectors of agriculture & forestry, fisheries, finance & insurance, medical & health care, or public service sectors. The number of firms analyzed is 10,024. I also calculate the Domar weight λ_i, which is the ratio of sales to GDP.[16] Summary statistics is given in Table 2.

5.2 Summary statistics

Figure 2 shows the complementary cumulative distribution function (CCDF) of firm sales in the log-log scale. The CCDF is close to the straight line in the log-log scale, implying that a Pareto tail is a good approximation for the tail of the distribution of firm sales. I estimate the Pareto tail exponent α in eq.(4) by Hill's method. I use the sales of the top 500 firms in the estimation and find that $\widehat{\alpha} = 1.25$ (s.e. $= 0.050$). Although the estimate $\widehat{\alpha}$ is significantly larger than $\alpha = 1$ of Zipf's law, it is in the range of $1 < \alpha < 2$ and I can apply Propositions 2 and 4.

I construct the productivity growth of firm i, $\widehat{\epsilon}_{it}$, by

$$\widehat{\epsilon}_i := (\ln s_{i,2017} - \ln s_{i,2016}) - b(\ln l_{i,2017} - \ln l_{i,2016})$$

where b is the share of labor cost to sales. I set $b = 0.15$. ¿ I take the value of $b = 0.15$ from Financial Statements Statistics of Corporations by Industry, Annually, compiled by the Ministry of Finance. $\widehat{\epsilon}_i$ is the empirical counterpart of microeconomic shock ϵ_i in my analysis. The summary statistics of $\widehat{\epsilon}_i$ is given in Table 2 and, in particular, the estimate of the standard deviation of $\widehat{\epsilon}_i$ is $\widehat{\sigma}_\epsilon = 0.137$. Figure 3 shows that the density estimation of $\widehat{\epsilon}_i$, showing that its tail is heavier than Gaussian. By using the cross-sectional data of $\widehat{\epsilon}_i$, I estimate the Pareto tail exponent of β in 6 by Hill's method. I find that $\widehat{\beta} = 2.41$(s.e $= 0.066$).

[15] I also use Factset to study the empirical importance of the granularity. See the Appendix.
[16] GDP in Japan is 545 trillion yen in 2017.

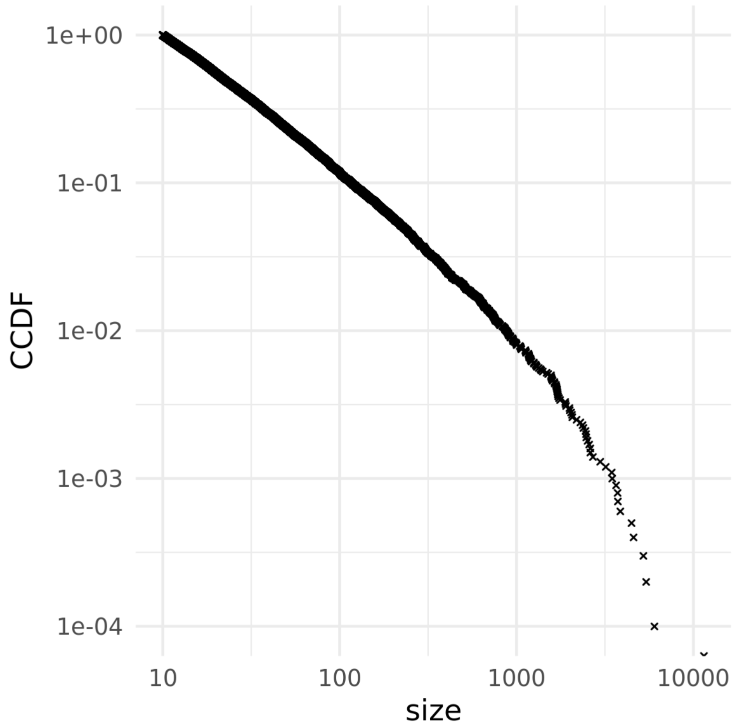

Figure 2: CCDF of firm sales

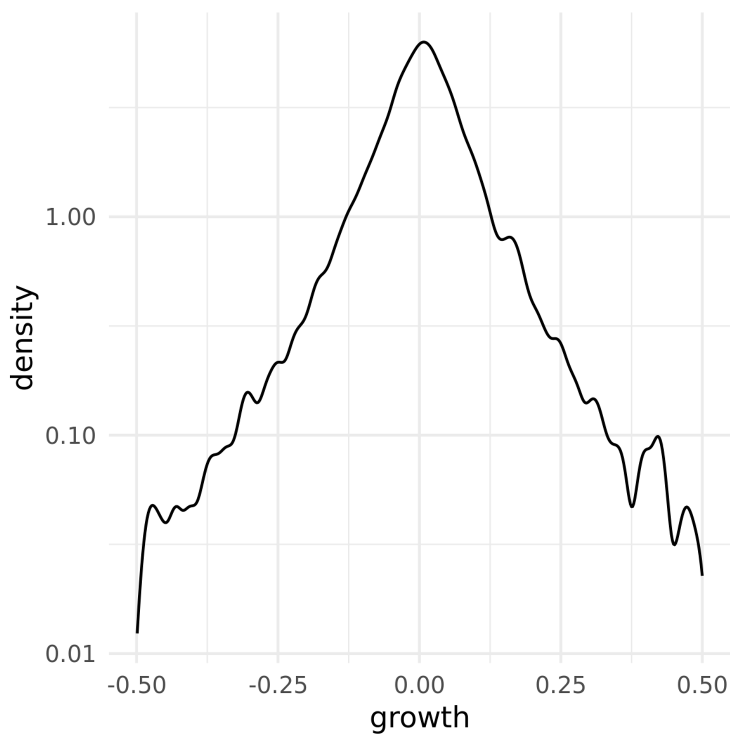

Figure 3: Density estimate of $\widehat{\epsilon}$

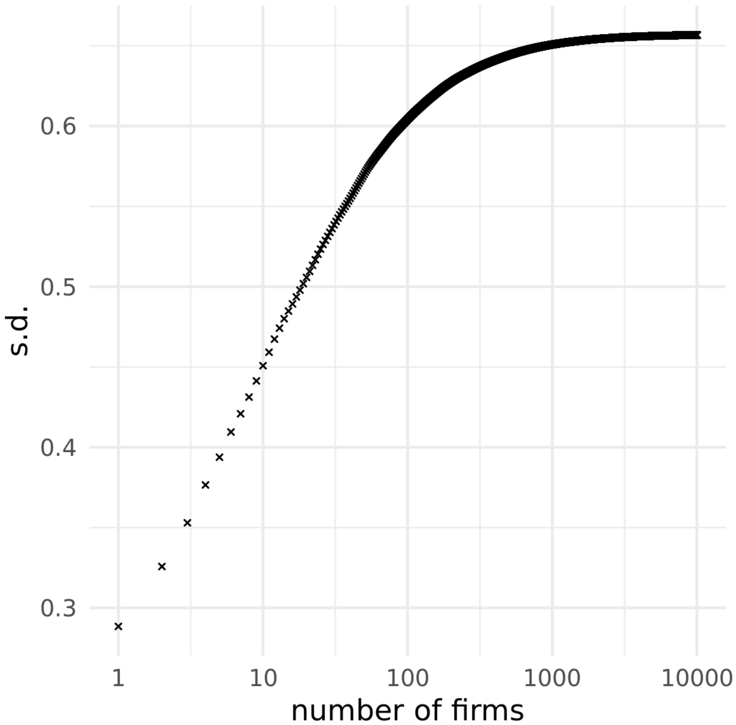

Figure 4: Plot of n_{top} and $\sigma_{n_{\text{top}}}$

5.3 Quantitative analysis

Given the estimates, I quantitatively test the role of the granularity by using the results in Sections 3 and 4. First, I calculate the standard deviation of aggregate output induced by microeconomic shocks to the n_{top} largest firms, i.e.

$$\sigma_{n_{\text{top}}} := \widehat{\sigma}_{\epsilon}\sqrt{\sum_{m=1}^{n_{\text{top}}} \lambda_m^2} \times 100$$

where λ_m is the Domar weight of the mth largest firm.

Figure 4 shows the plot of $\sigma_{n_{\text{top}}}$ as a function of n_{top}. For example, $\sigma_{n_{\text{top}}=1}$ is 0.29, i.e. the largest firm leads to 0.29% standard deviation of aggregate output. Similarly, $\sigma_{n_{\text{top}}=100}$ is 0.60, i.e. the top 100 largest firm leads to 0.60% standard deviation of aggregate output. Microeconomic shocks to the small number of large firms has a sizable impact on the variance of aggregate output. This implies the granularity is an important factor for macroeconomic fluctuations when measured by the variance. Figure 4 shows that, though the $\sigma_{n_{\text{top}}}$ is strictly increasing as n_{top}, $\sigma_{n_{\text{top}}}$ almost converges to around 0.66 (the value of $\sigma_{n_{\text{top}}=10{,}024}$). I calculate the ratio $\sigma^2_{n_{\text{top}}=1}/\sigma^2_{n_{\text{top}}=10{,}024}$ as a proxy of r_n in Proposition 4. I find that the ratio is 0.193. As predicted by Proposition 4, this shows that the largest firm accounts for the significant fraction of $\sigma^2_{Y_n}$.

Next, I calculate the macroeconomic tail probability induced by microeconomic shocks. First, consider the case where F_ϵ is given by eq.(6) and the macroeconomic tail probability is given by 7. With the estimate of β above, I find that $\sum_{i=1}^{10024} \lambda_i^{\beta} = 0.00029$. Since $\overline{F}(x) \leq 1/2$ for $x > 0$, it shows that the macroeconomic tail probability is at most 0.00015, i.e. the large deviation of Y_n occurs once (or at most twice) in 10,000 years. Hence, the macroeconomic tail probability induced by microeconomic shocks is negligible. This shows that the empirical granularity is too low to lead to the micro-originated large deviation of Y_n. I obtain the same implication for the example of the distribution having a Weibull-like tail in eq.(8). Suppose that τ is very close to 1 and $K = 1/2$ (i.e. Laplace distribution). I approximate the macroeconomic tail probability by

$$P(Y_n > x) \sim \frac{1}{2} e^{-\beta\left(\frac{x}{\lambda_1}\right)} = \frac{1}{2}(2\overline{F}_\epsilon(x))^{\frac{1}{\lambda_1}}$$

Table 2 shows that $\lambda_1 = 0.0211$. Hence, $P(Y_n > x)$ is about the 50th power of $2\overline{F}_\epsilon(x)$, which turns out to be very small with a plausible $\overline{F}_\epsilon(x)$. As in the example of the Pareto tail, the empirical granularity is too low to lead to the micro-originated large deviation of Y_n.

6 Non-linear models

In the previous sections, I develop the results in the linear model, eq.(linear). In this section, I consider non-linear models and study the role of the granularity in the variance of aggregate output and the macroeconomic tail probability. I give upper bounds for the variance of aggregate output and macroeconomic tail probability, which decays very fast as $n \to \infty$ without the granularity.

6.1 Example of Baqaee and Farhi (2019)

To begin with, I summarize notations that generalize the linear model. Let $\epsilon := (\epsilon_1, ..., \epsilon_n)$, where $\epsilon_1, ..., \epsilon_n$ are a sequence of iid random variables. Let f be a mapping from $\mathbb{R}^n$ to $\mathbb{R}$ that links microeconomic shocks and aggregate output. i.e.

$$Y_n = f(\epsilon)$$

The mapping f represents a mechanism that translates microeconomic shocks ϵ into macroeconomic fluctuations. For example, the linear model in the previous sections is equivalent to a linear mapping $f(\epsilon) = \sum_i \lambda_i \epsilon_i$. In this section, I consider a general mapping f that contains not only the linear term but non-linear terms. As measures of macroeconomic fluctuations, I consider the variance

$$\mathrm{Var}(f(\epsilon)) := E\left[(f(\epsilon) - E[f(\epsilon)])^2\right]$$

and the macroeconomic tail probability

$$P(f(\epsilon) - E[f(\epsilon)] < -x)$$

In this section, I consider the macroeconomic tail probability not only for the limit of x but for any value of x. I study how granularity is related to $\mathrm{Var}(f(\epsilon))$ and $P(f(\epsilon) - E[f(\epsilon)] < -x)$

Before giving a general result, I study an example of a non-linear mapping $f(\epsilon)$ in Baqaee and Farhi (2019). Assume that the production function is the CES. Assume further that microeconomic shocks ϵ_i is a labor-augmenting productivity shock and the elasticity of substitution in the CES production function is common θ across all firms. Then, Corollary 1 in Baqaee and Farhi (2019) shows that in a competitive equilibrium, $f(\epsilon)$ up to the second-order terms is given by

$$f(\epsilon) = \sum_i \lambda_i^* \epsilon_i + \sum_i \sum_j b_{ij} \epsilon_i \epsilon_j$$

where $b_{ij} := (\theta - 1)\lambda_i^*(1_{\{i=j\}} - \lambda_j^*)$. λ_i^* is the Domar weight for labor-augmenting shocks for firm i, i.e., the labor cost of firm i divided by GDP. In this case, the sum $\sum_i \lambda_i^*$ is equal to 1 and does not depend on ϵ_i.

Let us consider the impact of ϵ_i on $f(\epsilon)$ in this model. To make the analysis simple, assume further that ϵ_i is restricted on the interval $(-1, 1)$. I do not impose any other restrictions on F_ϵ, i.e. F_ϵ is an arbitrary distribution on $(-1, 1)$. The key observation is that the impact of ϵ_i on $f(\epsilon)$ is bounded. To see it, let $g_i(\epsilon_i)$ be the sum of the terms involving ϵ_i, i.e.

$$
\begin{aligned}
g_i(\epsilon) &:= \lambda_i^* \epsilon_i + \sum_j b_{ij} \epsilon_i \epsilon_j + \sum_j b_{ji} \epsilon_j \epsilon_i - b_{ii} \epsilon_i \epsilon_i \\
&= \lambda_i^* \epsilon_i \left(1 + (\theta - 1)(1 + \lambda_i^*) \epsilon_i - 2(\theta - 1) \sum_j \lambda_j^* \epsilon_j \right)
\end{aligned}
$$

Since $\sum_j \lambda_j^* = 1$, $0 \leq \lambda_i^* \leq 1$, I obtain

$$
|g_i(\epsilon)| \leq (1 + 4|\theta - 1|) \lambda_i^*
$$

Imagine that the microeconomic shock to firm i changes from ϵ_i to ϵ_i' while others $\epsilon_1, ...\epsilon_{i-1}, \epsilon_{i+1}, ..., \epsilon_n$ are fixed. It follows that

$$
\begin{aligned}
|f(\epsilon_1, ..., \epsilon_n) - f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)| &= |g_i(\epsilon_1, ..., \epsilon_n) - g_i(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)| \\
&\leq (2 + 8|\theta - 1|) \lambda_i^*
\end{aligned}
$$

This shows that the impact of ϵ_i on $f(\epsilon)$ is bounded and its bound depend linearly on λ_i^*. Note that in the linear model, the bound is given by $2\lambda_i^*$. This shows that in both linear and non-linear model given above, the impact of ϵ_i is bounded by the same form of $C\lambda_i^*$, where C is a constant. In particular, as granularity becomes low, i.e. $\lambda_i^* \to 0$ for all i, the impact of ϵ_i also becomes small.

6.2 General non-linear mappings

Motivated by this example, I consider a non-linear mapping satisfying the following condition:

Definition 4 *A mapping $f : \mathbb{R}^n \to \mathbb{R}$ has the bounded differences condition if there exist non-negative constants $c_1, ..., c_n$ such that for all i,*

$$
\sup_{\substack{\epsilon_1, ..., \epsilon_n \\ \epsilon_i \in \mathbb{R}}} |f(\epsilon_1, ..., \epsilon_n) - f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)| \leq c_i
$$

This means that the maximum change of $f(\epsilon)$ in response to change of ϵ_i is bounded by constant c_i. Constants $c_1, ..., c_n$ measure granularity of an economy. I impose no condition on random variables $\epsilon_1, ..., \epsilon_n$ other than their independence. Given these conditions, I obtain the following result that links granularity $c_1, ..., c_n$ and macroeconomic fluctuations.

Proposition 8 *Let $\epsilon_1, ..., \epsilon_n$ be a sequence of independent random variables. Suppose that a mapping f satisfies the bounded differences condition with constants $c_1, ..., c_n$. Then*

$$
Var(f(\epsilon)) \leq \frac{\sum_i c_i^2}{4}
$$

and for $x \geq 0$

$$
P\left(f(X) - E[f(X)] \geq x\right) \leq \exp\left(-\frac{x^2}{2\sum_i c_i^2}\right)
$$

Proof. See the Appendix. ■

This proposition shows that both variance and macroeconomic tail probability is determined by $\sum_i c_i^2$, and if granularity is low, macroeconomic fluctuations are also small. For example, consider the homogeneous case where $c_i = n^{-1}$, i.e. the maximum impact of ϵ_i is the reciprocal of the number of firms. In this case, $\sum_i c_i^2$ is given by n^{-1}, and therefore, the variance and macroeconomic tail probability decays at the rate of n^{-1} and $\exp(-x^2 n/2)$, respectively, as $n \to \infty$. This implies that without the granularity, microeconomic shocks cancel out each other and do not contribute to macroeconomic fluctuations in the non-linear models. In other words, without the granularity, the diversification argument holds in non-linear models. Therefore, the granularity is a necessary condition for macroeconomic fluctuations in non-linear models as well as the linear model.

6.3 Example of Baqaee and Farhi (2019)—continued

Let us return to the example in Baqaee and Farhi (2019). I can apply Proposition 8 and obtain the following result.

Proposition 9 *Suppose that the conditions above hold. Then*

$$Var(f(\epsilon)) \leq C\|\lambda^*\|_2^2$$

and

$$P\left(f(\epsilon) - E[f(\epsilon)] \geq x\right) \leq \exp\left(-C\frac{x^2}{2\|\lambda^*\|_2^2}\right)$$

where C is a constant, which is independent of λ_i for all i.

Proof. Since $|g_i(\epsilon)| \leq (1 + 4|\theta - 1|)\lambda_i^*$, it is the direct consequence of Proposition 8. ■

Eq.(9) is reminiscent of eq.(3) in the linear model. Even when the second-order term is considered, the variance is controlled by the ℓ^2-norm of the Domar weights. Hence, I can apply the results in Section 3 to $f(\epsilon)$ having the second-order terms. In particular, I conclude that even in the non-linear model, the decaying rate of the variance of aggregate output is related to that of the large firms. Similarly, eq.(9) is reminiscent of eq.(5) in the linear model. The macroeconomic tail probability exhibits the Gaussian decay, which depends on the ℓ^2-norm of the Domar weights. For example, in the homogeneous case, the macroeconomic tail probability is bounded by $\exp\left(-Cnx^2\right)$. Note that eq.(9) holds for any x and n (i.e. λ_i^*), which means that I can also consider the limit of n with x fixed. Therefore, without the granularity, the macroeconomic tail probability decays very fast as n increases.

7 Conclusion

In this paper, I provide probabilistic results that illustrate the role of the granularity in macroeconomic fluctuations. First, in the linear model, I show that the contribution of the largest firm is the key to the slowly decaying of the variance of aggregate output as the number of firms $n \to \infty$. More specifically, I show that the contribution of the largest firm does not become negligible even when n is sufficiently large. The empirical data also shows that the contribution of a small number of large firms accounts for a non-negligible part of the variance of aggregate fluctuations.

However, the role of the granularity in the macroeconomic tail probability is different from in the variance of aggregate output. Especially when the distribution of microeconomic shocks has a heavy tail, I show that the macroeconomic tail probabilistic is asymptotically equal to the sum of the tail probability of $\lambda_i \epsilon_i$, i.e. the Domar weight times microeconomic shock to firm i. This result can be seen the granular hypothesis for the macroeconomic tail probability. By using it, I find that the empirical Domar weights are not so heterogeneous that microeconomic shocks lead to the large deviation of aggregate output. Therefore, in the linear model, the role of the granularity is negligible in the macroeconomic tail probability.

Finally, I extend my analysis to non-linear models. To show the role of the granularity, I consider a non-linear model which exhibits no granularity, i.e. the impact of firm i on aggregate output is bounded by constant c_i. I obtain the upper bounds for the variance of aggregate output and macroeconomic tail probability, which depend on the ℓ^2-norm of the Domar weights. This means that as in the linear model, macroeconomic fluctuations become negligible without the granularity, which can be seen as the diversification argument in non-linear models. Therefore, the granularity is a necessary condition in non-linear as well as linear models.

In short, my results implies that (a) with the granularity but without non-linear terms, microeconomic shocks are important for the variance of aggregate output but not for the macroeconomic tail probability and (b) with non-linear terms but without the granularity, microeconomic shocks have no impact on the variance of aggregate output nor the macroeconomic tail probability. Therefore, both the granularity and non-linear terms are needed to explain the micro-originated large deviation of aggregate output. Indeed, some recent papers make important progress in this direction. A series of paper by Baqaee, Fahri (e.g. Baqaee and Farhi (2019)) consider non-linear higher-order terms of microeconomic shocks in efficient and inefficient economy. My paper complements this line of literature, which is promising for future research.

Appendix

A. Proof of propositions in Section 3

I employ a proof strategy similar to Proposition 2 in Gabaix (2011). Since I study the contribution of the largest firm, I use the extreme value theory, instead of Levy's theorem.

Proof of Proposition 1. Consider the following random variable:

$$n^2(\ln n)^{-1}\|\lambda\|_\infty^2 = \frac{1}{b^2}\frac{(\ln n)^{-1}s_{\max,n}^2}{(n^{-1}\sum_i s_i)^2}$$

First, since the assumption means that $E[s] < \infty$, the strong law of large numbers implies that $(n^{-1}\sum_i s_i)^2$ converges to $E[s]^2$ almost surely as $n \to \infty$. Second, since F_{s^2} has an exponential-like tail, I obtain the following as in the main text:

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

Hence,

$$n^2(\ln n)^{-1}\|\lambda\|_\infty^2 \xrightarrow{P} \frac{1}{b^2\alpha E[s]^2}$$

By setting $C := \frac{\sigma_\epsilon^2}{b^2\alpha E[s]^2}$, the result follows. ■

For the proof of Proposition 2, I employ the same proof strategy as in Proposition 1.

Proof of Proposition 2. For notational simplicity, let $k = \alpha/2 > 1/2$. Consider the following random variable:

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

First, since the assumption means that $E[s] < \infty$, the strong law of large numbers implies that $(n^{-1}\sum_i s_i)^2$ converges to $E[s]^2$ almost surely as $n \to \infty$. Second, since F_{s^2} has an Pareto-like tail with exponent k, I obtain the following as in the main text:

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

where u is a non-degenerate random variable, which is independent of n.

Hence,

$$n^{2-1/k}\|\lambda\|_\infty^2 \overset{d}{\to} \frac{K^{1/k}u}{b^2E[s]^2}$$

By setting $C := \frac{K^{1/k}\sigma_\epsilon^2}{b^2E[s]^2}$, the result follows. ■

Proof of Proposition 3. The assumption $E[s^2] < \infty$ implies that $\sum_{n=1}^\infty P(s_n^2 > \delta n) < \infty$, $\forall\delta > 0$. Hence, Borel-Cantelli lemma implies that $P(s_n^2 > \delta n \text{ i.o.}) = 0$, $\forall\delta > 0$, which is equivalent to $\lim_{n\to\infty} n^{-1}s_n^2 = 0$ a.s. Finally, note the relation between $n^{-1}s_n^2$ and $n^{-1}s_{\max,n}^2$: for some $n_0 \le n$,

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

This implies

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

Since the assumption and the strong law of laws of large numbers implies that $(n^{-1}\sum_i s_i^2) \overset{\text{a.s.}}{\to} E[s^2]$, I obtain

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

■

The convergence of the ratio r_n under general conditions is studied in Chow and Teugels (1979) and Bingham and Teugels (1981). I give a simplified version of their proof.

Proof of Proposition 4. For notational simplicity, let $\xi_1, \xi, ...$ be a sequence of independent random variables with distribution F such that $\overline{F}(x) \sim Kx^{-k}$ with $0 < k < 1$. Let $\xi_{\max,n} := \max_i(\xi_i)$.

The first part of the proposition is equivalent to the convergence of $(U_n, V_n) := \left(a_n^{-1}\sum_i \xi_i, c_n^{-1}\xi_{\max,n}\right)$ to a non-degenerate limit (U, V). To show it, consider the characteristic-distribution function for each n defined by

$$\chi_n(t, v) := E\left[\exp\left(ita_n^{-1}\sum_i \xi_i\right); \xi_{\max,n} < c_n v\right]$$

Since

$$E\left[\exp\left(it\sum_i \xi_i\right); M_n < v\right] = \left(\int_{-\infty}^{v} e^{itu} dF(u)\right)^n$$

$\chi_n(t, v)$ has the following representation:

$$\chi_n(t,v) = \left(1 + \frac{1}{n}\int_{-\infty}^{\infty} n(e^{itw} - 1)dF(a_n w) - \frac{1}{n} g_n(t,v)\right)^n$$

where $g_n(t,v) := \int_v^\infty n \exp(-ita_n^{-1} c_n y) dF(c_n y)$

The stable law implies

$$\int_{-\infty}^{\infty} n(e^{itw} - 1)dF(a_n w) \to \ln \omega_\alpha(t)$$

where $\omega_\alpha(t)$ is the characteristic function of the stable distribution with parameter α. For $g_n(t,v)$, note that the norming constants a_n for the sum and c_n for the maximum increase at the same growth rate. Hence, $a_n^{-1} c_n$ converges to some constant C. In addition, the definition of c_n and the assumption that $\overline{F} \sim Kx^{-k}$ with $0 < k < 1$ imply that

$$n\overline{F}(c_n y) \sim y^{-\alpha}$$

Hence, for $v > 0$,

$$g_n(t,v) \to \int_v^\infty \exp(itCy) d(-y^{-\alpha}) \ \text{ as } n \to \infty$$

Hence, for $t \in \mathbb{R}, v > 0$

$$\chi_n(t,v) \to \chi(t,v)$$

where $\chi(t,v) := E[e^{itU}; V \le v] = \omega_\alpha(t) \cdot \exp\left(\int_v^\infty \exp(itCy) d(-y^{-\alpha})\right)$. This implies that $(U_n, V_n) \xrightarrow{d} (U, Y)$, and therefore, the first part of the result follows.

For the second part of the proposition, note that for a general distribution F, the following equation holds.

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

By taking the derivative of both sides and setting it equal to 0, I obtain

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

Here, I use the assumption that F_ξ is a distribution on $\mathbb{R}^+$.

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

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

Hence, I obtain

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

Therefore, $E[r_n]$ converges to $1/2$.[17] ∎

I use the equivalent relation between the occurrence of extremes and Poisson random variables (see Theorem 4.2.1 in Embrechts et al. (1997)).

Proof of Proposition 5. Let (u_n) be a sequence of real numbers such that $n\overline{F}_{s^2}(u_n) \to \tau \in (0,\infty)$. Let B_n to be the number of exceedances of s_i^2, i.e.

$$
B_n := \sum_i I_{\{s_i^2 > u_n\}}
$$

where $I_{\{\cdot\}}$ is the indicator function. By construction, B_n is a binomial random variable with parameters n and $\overline{F}_{s^2}(u_n)$. Note that the event $s_{m,n}^2 \leq u_n$ is equivalent to the event $\sum_i I_{\{s_i^2 > u_n\}} < m$. Hence, $P(s_{m,n}^2 \leq u_n) = P(B_n < m)$.

Next, consider the limit of $P(B_n < m)$ as $n \to \infty$. Poisson's theorem implies that for $m \in \mathbb{N}_0$,

$$
\lim_{n\to\infty} P(B_n \leq m) = e^{-\tau}\sum_{r=0}^{m}\frac{\tau^r}{r!}
$$

Hence, I obtain

$$
\lim_{n\to\infty} P(s_{m,n}^2 \leq u_n) = \lim_{n\to\infty} P(B_n < m) = e^{-\tau}\sum_{r=0}^{m-1}\frac{\tau^r}{r!}
$$

If F_{s^2} has a Pareto-like tail with exponent k, u_n is given by $u_n(x) = (Kn)^{1/k}x,\ x > 0$. Indeed, $n\overline{F}_{s^2}(u_n(x)) \to x^{-k}$ as $n \to \infty$. Let $\tau(x) := x^{-k}$. Hence, I obtain that for each $m \in \mathbb{N}_0$,

$$
\lim_{n\to\infty} P\left((Kn)^{-1/k}s_{m,n}^2 \leq x\right) = e^{-\tau(x)}\sum_{r=0}^{m-1}\frac{\tau(x)^r}{r!}
$$

This means that $(Kn)^{-1/k}s_{m,n}^2$ converges in distribution to a non-degenerate random variable. Hence, $s_{m,n}^2$ increases at the rate of $n^{-1/k}$ as n increases. Since the increases rate of $s_{m,n}^2$ is equal to that of $s_{\max,n}^2$, I can apply the same reasoning as in the proof of Proposition 2. Therefore, the result follows. ∎

[17] I need to generalize the proof for a general F.

B. Proof of propositions in Section 4

To obtain the upper bound of the tail probability $P(Y_n \geq x)$, I use techniques developed in the theory of concentration inequalities (Boucheron et al. (2012);Wainwright (2019)).

Proof of Proposition 6. First, I give the upper bound for the moment generating function of ϵ. The assumption implies that there exists $c_0 > 0$ such that $E[e^{t\epsilon}] < \infty$ for all $t \in [0, c_0)$. This implies that there exist $\nu \geq 0, b \geq 0$ such that for $t \in [0, \frac{1}{b})$

$$E[e^{t\epsilon}] \leq e^{\nu^2 t^2/2}$$

Indeed, the existence of the moment generating function means that as $t \to 0$

$$E[e^{t\epsilon}] = 1 + \frac{t^2 E[\epsilon^2]}{2} + o(t^2)$$

On the other hand, Taylor's expansion implies that

$$e^{\frac{\nu^2 t^2}{2}} = 1 + \frac{\nu^2 t^2}{2} + o(t^2)$$

Hence, when $\nu^2 > E[\epsilon^2]$, there exists $b \geq 0$ such that $E[e^{t\epsilon}] \leq e^{\frac{\nu^2 t^2}{2}}$ fora all $|t| \leq \frac{1}{b}$. I can apply the same reasoning to random variable $\lambda_i \epsilon_i$.

Next, consider the moment generating function of the weighted sum $\sum_i \lambda_i \epsilon_i$. Given the upper bound for $E[e^{t\lambda_i \epsilon_i}]$ above, I obtain

$$E\left[e^{t\sum_i \lambda_i \epsilon_i}\right] \leq \prod_i E[e^{t\lambda_i \epsilon_i}] \leq \prod_i e^{\nu_i^2 t^2/2} \leq e^{\nu_*^2 t^2/2}$$

where $b_* := \|\lambda\|_\infty b$ and $\nu_* := \|\lambda\|_2 \nu$.

Given these upper bounds, I give the upper bound for $P\left(\sum_i \lambda_i \epsilon_i \geq x\right)$. Markov inequality implies

$$P\left(\sum_i \lambda_i \epsilon_i \geq x\right) \leq e^{-tx} E\left[e^{t\sum_i \lambda_i \epsilon_i}\right] \leq \exp\left(-tx + \frac{\nu_*^2 t^2}{2}\right)$$

I optimize the upper bound with respect to t (Chernoff's bound). Let $g(t,x) := -tx + \frac{\nu_*^2 t^2}{2}$ and let $g_*(x)$ be its infimum, i.e. $g_*(x) := \inf_{t \in [0, b^{-1})} g(t,x)$. If $0 \leq x < \frac{\nu^2}{b_*}$, $g(t,x)$ takes its minimum value at $t_x = \frac{x}{\nu^2}$ and $g_*(x) = -\frac{x^2}{2\nu^2}$. If $x \geq \frac{\nu^2}{b_*}$, $g(t,x)$ takes its minimum value at $t_x = b_*^{-1}$ and

$$g_*(x) = -\frac{x}{b_*} + \frac{1}{2b_*}\frac{\nu_*^2}{b_*} \leq -\frac{x}{2b_*}$$

For the last inequality, I use $x \geq \frac{\nu_*^2}{b}$. Therefore, I obtain

$$P\left(\sum_i \lambda_i \epsilon_i \geq x\right) \leq \begin{cases} \exp\left(-\frac{x^2}{2\nu_*^2}\right) & \text{for } 0 \leq x \leq \frac{\nu_*^2}{b_*} \\ \exp\left(-\frac{x}{2b_*}\right) & \text{for } x > \frac{\nu_*^2}{b_*} \end{cases}$$

■

Proof of Example 1. Let $h(x) := \frac{1}{2}x^{2/3}$. $\overline{F_\epsilon * F_\epsilon}(x)$ can be decomposed into three parts:

$$\overline{F_\epsilon * F_\epsilon}(x) = P(\epsilon_1+\epsilon_2 > x, \epsilon_1 \le h(x))+P(\epsilon_1+\epsilon_2 > x, \epsilon_2 \le h(x))+P(\epsilon_1+\epsilon_2 > x, \epsilon_1 > h(x), \epsilon_2 > h(x))$$

(i) $\overline{F_\epsilon * F_\epsilon}(x) \le 2\overline{F}_\epsilon(x)$

Note that $P(\epsilon_1 + \epsilon_2 > x, \epsilon_1 \le h(x)) \le \overline{F}_\epsilon(x - h(x))$. By the definition of h, I have

$$\overline{F}_\epsilon(x-h(x)) = x^{-\beta}\left(1 - \frac{1}{2}x^{-\frac{1}{3}}\right)^{-\beta} \sim \overline{F}_\epsilon(x)$$

Similarly, $P(\epsilon_1 + \epsilon_2 > x, \epsilon_2 \le h(x)) \le \overline{F}_\epsilon(x - h(x))$ By the independence of ϵ_1 and ϵ_2, $P(\epsilon_1 + \epsilon_2 > x, \epsilon_1 > h(x), \epsilon_2 > h(x)) \le \overline{F}^2(h(x)) = o(\overline{F}_\epsilon(x))$. Hence, I have

$$\limsup_{x\to\infty} \frac{\overline{F_\epsilon * F_\epsilon}(x)}{\overline{F}_\epsilon(x)} \le 2$$

(ii) $\overline{F_\epsilon * F_\epsilon}(x) \ge 2\overline{F}_\epsilon(x)$

Since $\epsilon_1, \epsilon_2 > 0$ a.s, $\{\epsilon_1 + \epsilon_2 > x\} \supset \{\epsilon_1 > x\} \cup \{\epsilon_2 > x\}$. Hence, $\overline{F_\epsilon * F_\epsilon}(x) \ge 2\overline{F}_\epsilon(x)$. The desired result follows. ■

To prove Proposition 7, I introduce the class of long-tailed distributions (denoted by $\mathcal{L}$), which includes subexponential distributions as a subset. I give simplified proof of lemmas below to make it self-contained (for detail discussion about long-tailed distributions, see Ch.2 in Foss et al. (2011)).

Definition 5 *A distribution F is long-tailed ($F \in \mathcal{L}$) if for any fixed $y > 0$*

$$\overline{F}(x+y) \sim \overline{F}(x) \quad \textit{as } x \to \infty$$

Lemma 1 *Subexponential distributions are long-tailed.*

Proof. It suffices to show that $F^+ \in \mathcal{S}$ implies $F^+ \in \mathcal{L}$. First, note that for any distribution F on $\mathbb{R}^+$,

$$\overline{F * F}(x) = \int_0^x \overline{F}(x-y)F(dy) + \overline{F}(x)$$

On the other hand, the definition of the subexponential distribution means that

$$\overline{F * F}(x) \sim 2\overline{F}(x)$$

Hence,

$$\int_0^x \overline{F}(x-y)F(dy) = \overline{F}(x) + o(\overline{F}(x))$$

This implies that

$$\begin{aligned}\int_0^x F(x-y,x]F(dy) &= \int_0^x (\overline{F}(x-y) - \overline{F}(x))F(dy) \\ &= o(\overline{F}(x))\end{aligned}$$

Since $F(x-y,x]$ is non-decreasing function of y, it follows that $F(x-a,x]F(a,x] \le o(\overline{F}(x))$ for $a \le x$. Hence, I obtain $F(x-a,x] = o(\overline{F}(x))$, which is equivalent to $\overline{F}(x-a) \sim \overline{F}(x)$. ■

Definition 6 *Given a strictly positive, non-decreasing function h, an ultimately positive function f is h-insensitive if*

$$\sup_{|y|\le h(x)} |f(x+y) - f(x)| = o(f(x))$$

One can show that a long-tailed distribution F has a function h such that $h(x) \to \infty$ and F is h-insensitive (Lemma 2.19 in Foss et al. (2011)). One can also show that (i) given a sequence of distributions $F_1, ..., F_n$, there exists a function h such that F_i is h-insensitive for all i and (ii) given a long-tailed distributions F and a positive non-decreasing function $\hat{h}$, there exists a function h such that $h(x) \le \hat{h}(x)$ for all x and F is h-insensitive (Proposition 2.20 in Foss et al. (2011)).

Lemma 2 *Let ξ_1, ξ_2 be independent random variables with a common subexponential distribution F on $\mathbb{R}$. Then, there exists a function h such that $h(x) < x/2$, $h(x) \to \infty$ as $x \to \infty$, F is h-insensitive, and*

$$P(\xi_1 + \xi_2 > x, \xi_1 > h(x), \xi_2 > h(x)) = o(\overline{F}(x))$$

Proof. The assumption that $F \in \mathcal{S}_{\mathbb{R}}$ means that $F \in \mathcal{L}$. Hence, there exists a function h such that $h(x) \to \infty$ as $x \to \infty$ and F is h-insensitive. By using h, decompose $P(\xi_1 + \xi_2 > x)$ into three parts:

$$P(\xi_1+\xi_2 > x) = P(\xi_1+\xi_2 > x, \xi_1 \le h(x)) + P(\xi_1+\xi_2 > x, \xi_2 \le h(x)) + P(\xi_1+\xi_2 > x, \xi_1 > h(x), \xi_2 > h(x))$$

Consider the first term of the right-hand side $P(\xi_1 + \xi_2 > x, \xi_1 \le h(x)) = \int_{-\infty}^{h(x)} \overline{F}(x-y)F(dy)$. By using the monotonicity of $\overline{F}(x)$, I obtain its upper bound:

$$\int_{-\infty}^{h(x)} \overline{F}(x-y)F(dy) \le \overline{F}(x - h(x))$$

On the other hand, since $h(x) \to \infty$ as $x \to \infty$, I obtain the following lower bound:

$$\begin{aligned}\int_{-\infty}^{h(x)} \overline{F}(x-y)F(dy) &\ge \int_{-h(x)}^{h(x)} \overline{F}(x-y)F(dy) \\ &\ge F(-h(x), h(x))\overline{F}(x + h(x)) \\ &\sim \overline{F}(x + h(x)) \ \text{ as } x \to \infty\end{aligned}$$

Hence,

$$P(\xi_1 + \xi_2 > x, \xi_1 \le h(x)) \sim \overline{F}(x)$$

Here, I use the property that F is long-tailed. Similarly, I obtain

$$P(\xi_1 + \xi_2 > x, \xi_2 \le h(x)) \sim \overline{F}(x)$$

Recall that the definition of $\mathcal{S}_{\mathbb{R}}$ means that $P(\xi_1 + \xi_2 > x) \sim 2\overline{F}(x)$. Hence, I conclude that

$$P(\xi_1 + \xi_2 > x, \xi_1 > h(x), \xi_2 > h(x)) = o(\overline{F}(x))$$

■

Proof of Proposition 7.

To begin with, assume for simplicity that $\lambda_i < 1$. Let F be the distribution of ϵ_i and let F_i be the distribution of $\lambda_i \epsilon_i$. Then, it is immediate that $F_i \in \mathcal{S}_{\mathbb{R}}$ and $\overline{F}_i(x) = O(\overline{F}(x))$.

I show it by induction. First, consider the case of $n = 2$. Since $F_1, F_2 \in \mathcal{S}_{\mathbb{R}}$, there exists a function h such that $h(x) \leq x/2$, $h(x) \to \infty$, and F_1, F_2 are h-insensitive. By using h, decompose $\overline{F_1 * F_2}(x)$ into three parts:

$$\overline{F_1 * F_2}(x) = P(\xi_1 + \xi_2 > x, \xi_1 \leq h) + P(\xi_1 + \xi_2 > x, \xi_2 \leq h) + P(\xi_1 + \xi_2 > x, \xi_1 > h, \xi_2 > h)$$

where ξ_1, ξ_2 are random variables with distribution F_1 and F_2, respectively. For the first and second terms of the right-hand side, since both F_1 and F_2 are h-insensitive, I obtain

$$\int_{-\infty}^{h(x)} \overline{F}_1(x-y) F_2(dy) = \overline{F}_1(x) + o(\overline{F}_2(x)) = \overline{F}_1(x) + o(\overline{F}(x))$$

$$\int_{-\infty}^{h(x)} \overline{F}_2(x-y) F_1(dy) = \overline{F}_2(x) + o(\overline{F}_1(x)) = \overline{F}_2(x) + o(\overline{F}(x))$$

For the third term of the right-hand side,

$$\begin{aligned}
P(\xi_1 + \xi_2 > x, \xi_1 > h, \xi_2 > h) &= \int_{h(x)}^{\infty} \overline{F}_1(\max(h(x), x-y)) F_2(dy) \\
&\leq \sup_{z > h(x)} \frac{\overline{F}_1(z)}{\overline{F}(z)} \int_{h(x)}^{\infty} \overline{F}(\max(h(x), x-y)) F_2(dy) \\
&= \sup_{z > h(x)} \frac{\overline{F}_1(z)}{\overline{F}(z)} \int_{h(x)}^{\infty} \overline{F}_2(\max(h(x), x-y)) F(dy) \\
&\leq \sup_{z > h(x)} \frac{\overline{F}_1(z)}{\overline{F}(z)} \sup_{z > h(x)} \frac{\overline{F}_2(z)}{\overline{F}(z)} \int_{h(x)}^{\infty} \overline{F}(\max(h(x), x-y)) F(dy)
\end{aligned}$$

Since $\int_{h(x)}^{\infty} \overline{F}(\max(h(x), x-y)) F(dy) \sim o(\overline{F}(x))$, it follows that $P(\xi_1 + \xi_2 > x, \xi_1 > h, \xi_2 > h) \sim o(\overline{F}(x))$.

Next, consider the case of an arbitrary n given that the result holds in the case of $n-1$. Let $G_k := F_1 * ... * F_k$. Since $\overline{G}_{n-1}(x) = \overline{F}_1(x) + ... + \overline{F}_{n-1}(x) + o(\overline{F}(x))$, it follows that G_{n-1} is long-tailed and $\overline{G}_{n-1}(x) = O(\overline{F}(x))$. By applying the result for the case of $n = 2$ to $G_{n-1} * F_n$, the result follows.

■

Note that if $F_i = F$ for all i and $F \in \mathcal{S}_{\mathbb{R}}$, $\overline{F^{*n}}(x) = n\overline{F}(x) + o(\overline{F}(x))$ for arbitrary n. Hence, a subexponential distribution F satisfies $\overline{F^{*n}}(x) \sim n\overline{F}(x)$ for arbitrary n.

C. Proof of proposition in Section 6

The upper bounds for variance and tail probability in proposition 8 are the special cases of Efron-Stein inequality and Azuma-Hoeffding inequality, respectively (see Boucheron et al. (2012);Wainwright (2019)). I give their proof by the martingale method.

Let $E_i[\cdot]$ and $E^{(i)}[\cdot]$ be the conditional expectation conditional on $\epsilon_1, ..., \epsilon_i$ and $\epsilon_1, ..., \epsilon_{i-1}, \epsilon_{i+1}, \epsilon_n$, respectively, i.e. $E_i[f(\epsilon)] := E[f(\epsilon)|\epsilon_1, ..., \epsilon_i]$ and $E^{(i)}[f(\epsilon)] := E[f(\epsilon)|\epsilon_1, ..., \epsilon_{i-1}, \epsilon_{i+1}, ..., \epsilon_n]$. It follows from Fubini theorem that

$$E_i[E^{(i)}[f(\epsilon)]] = E_{i-1}[f(\epsilon)]$$

Define the martingale differences sequence (Δ_k) by

$$\Delta_k := E_i[f(\epsilon)] - E_{i-1}[f(\epsilon)]$$

Note that $f(\epsilon) - E[f(\epsilon)] = \sum_i \Delta_i$. This representation implies that

$$\text{Var}(f(\epsilon)) = E\left[\left(\sum_i \Delta_i\right)^2\right] = \sum_i E[\Delta_i^2] + 2\sum_{j>i} E[\Delta_i \Delta_j]$$

Since $E_i\Delta_j = 0$ for $j > i$, $E_i[\Delta_j \Delta_i] = \Delta_i E_i \Delta_j = 0$ Hence, I obtain the following additive formula:

$$\text{Var}(f(\epsilon)) = E\left[\left(\sum_i \Delta_i\right)^2\right] = \sum_i E[\Delta_i^2]$$

Suppose that f satisfies the bounded differences condition with constants $c_1, ..., c_n$. Let ϵ_i' be an independent copy of ϵ_i. Then,

$$\Delta_i = E_i[f(\epsilon) - E^{(i)} f(\epsilon)] = E_i\left[E^{(i')}\left[f(\epsilon) - f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)\right]\right]$$

Hence, it follows that $|\Delta_i| \leq c_i$.

Proof of Proposition 8. First, I give the proof of the first part of the result. By Jensen's inequality, it follows that

$$\Delta_i^2 \leq E_i\left[(f(\epsilon) - E^{(i)} f(\epsilon))^2\right]$$

Since $\text{Var}(f(\epsilon)) = \sum_i E[\Delta_i^2]$, this implies that

$$\text{Var}(f(\epsilon)) \leq \sum_i E\left[\left(f(\epsilon) - E^{(i)} f(\epsilon)\right)^2\right] = E\left[\sum_i \text{Var}^{(i)}(f(\epsilon))\right]$$

where $\text{Var}^{(i)}(f(\epsilon))$ is the conditional variance defined by $E^{(i)}\left[(f(\epsilon) - E^{(i)} f(\epsilon))^2\right]$.

Consider $\text{Var}^{(i)}(f(\epsilon))$. I use the variance identity: for a real-valued random variable ξ,

$$\text{Var}(\xi) = \inf_{a\in\mathbb{R}} E[(\xi - a)^2]$$

where the infimum is achieved when $a = E[\xi]$. This implies that for every $i = 1, ..., n$

$$\text{Var}^{(i)}(f(\epsilon)) = \inf_{\xi_i} E^{(i)}\left[(f(\epsilon) - \xi_i)^2\right]$$

where the infimum is taken over all $\epsilon^{(i)}$-measurable and square integrable variable ξ_i. Hence, I obtain

$$\text{Var}(Z) \leq \inf_{\xi_i} \sum_i E[(f(\epsilon) - \xi_i)^2]$$

Choose

$$\xi_i = \frac{1}{2}\left(\sup_{\epsilon_i'\in\mathbb{R}} f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n) + \inf_{\epsilon_i'\in\mathbb{R}} f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)\right)$$

Since both $\sup_{\epsilon_i' \in \mathbb{R}} f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)$ and $\inf_{\epsilon_i' \in \mathbb{R}} f(\epsilon_1, ..., \epsilon_{i-1}, \epsilon_i', \epsilon_{i+1}, ..., \epsilon_n)$ are $\epsilon^{(i)}$-measurable, ξ_i is $\epsilon^{(i)}$-measurable. Since $f(\epsilon) - \xi_i \leq c_i/2$, I conclude

$$\text{Var}(f(\epsilon)) \leq \frac{\sum_i c_i^2}{4}$$

Next, I give the proof of the second part of the result. Note that $|\Delta_i| \leq c_i$ implies that $\text{Var}(\Delta_i | \mathcal{F}_{i-1})$ is also bounded by c_i^2. Hoefdding's lemma (e.g. Lemma 2.2 in Boucheron et al. (2012)) implies that for $t \geq 0$,

$$E[e^{t\Delta_i} | \mathcal{F}_{i-1}] \leq e^{\frac{t^2 c_i^2}{2}}$$

This upper bound and conditioning on $\mathcal{F}_{n-1}$ yield

$$\begin{aligned} E e^{t \sum_i \Delta_i} &= E\left[e^{t \sum_i^{n-1} \Delta_i} E[e^{t\Delta_n} | \mathcal{F}_{n-1}]\right] \\ &\leq E\left[e^{t \sum_i^{n-1} \Delta_i}\right] e^{t^2 c_n^2/2} \end{aligned}$$

Iterating this procedure yields $E[e^{t \sum_i \Delta_i}] \leq e^{t^2 \sum_i c_i^2/2}$.

Given the upper bound of $E[e^{t \sum_i \Delta_i}]$, I apply Markov's inequality to $P(f(\epsilon) - Ef(\epsilon) \geq x)$:

$$P(f(\epsilon) - Ef(\epsilon) \geq x) \leq \frac{E e^{t(f(\epsilon) - Ef(\epsilon))}}{e^{tx}} \leq \exp\left(-tx + \frac{t^2 \sum_i c_i^2}{2}\right)$$

I optimize the upper bound with respect to t. It takes its infimum value at $t = \frac{x^2}{\sum_i c_i^2}$ and

$$-tx + \frac{t^2 \sum_i c_i^2}{2} = -\frac{x^2}{2 \sum_i c_i^2}$$

Therefore, I obtain that for all $x \geq 0$,

$$P(f(\epsilon) - E[f(\epsilon)] \geq x) \leq \exp\left(-\frac{x^2}{2 \sum_i c_i^2}\right)$$

■

References

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. (2016). Networks, Shocks, and Systemic Risk. In Bramoullé, Y., Galeotti, A., and Rogers, B., editors, *The Oxford Handbook of the Economics of Networks*, chapter 21.

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

Altinoglu, E. L. (2018). The Origins of Aggregate Fluctuations in a Credit Network Economy. *Working Paper*.

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.

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

Baqaee, D. R. and Farhi, E. (2018a). Macroeconomics with Heterogeneous Agents and Input-Output Networks. *Working Paper*, pages 1–60.

Baqaee, D. R. and Farhi, E. (2018b). Productivity and Misallocation in General Equilibrium. *Working Paper*.

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

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K., and Moxnes, A. (2018a). The Origins of Firm Heterogeneity: A Production Network Approach. *Working Paper*.

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2018b). Production Networks, Geography, and Firm Performance. *Journal of Political Economy*, 127(2):639–688.

Bigio, S. and La'o, J. (2016). Financial Frictions in Production Networks. *Working Paper*.

Bigio, S. and La'o, J. (2017). Distortions in Production Networks. *Working Paper*.

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

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

Caliendo, L., Parro, F., Rossi-Hansberg, E., and Sarte, P.-D. (2018a). The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy. *Review of Economic Studies*, 85(4):2042–2096.

Caliendo, L., Parro, F., and Tsyvinski, A. (2018b). Distortions and the Structure of the World Economy. *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 Grassi, B. (2019). Large Firm Dynamics and the Business Cycle. *American Economic Review*, 109(4):1375–1425.

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

Chow, T. L. and Teugels, J. L. (1979). The Sum and The Maximum of I.I.D Random Variables. In Mandl, P. and Huskova, M., editors, *Proceedings Second Prague Symp. Asymptotic Statistics*, pages 81–82.

Demir, B., Javorcik, B., Michalski, T., and Ors, E. (2017). Financial Constraints and Propagation of Shocks in Production Networks. *Working Paper*.

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. (2018). The Micro Origins of International Business-Cycle Comovement. *American Economic Review*, 108(1):82–108.

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

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modelling of extremal events in insurance and finance*. Springer.

Fadinger, H., Ghiglino, C., and Teteryatnikova, M. (2018). Income Differences, Productivity and Input-Output Networks. *Working Paper*.

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*. Springer, New York.

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

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

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

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. (2018). Industrial Policies in Production Networks. *Working Paper*.

Lucas, R. E. (1977). Understanding Business Cycles. In *Carnegie-Rochester conference series on public policy*, pages 7–29. North-Holland.

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

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

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

Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*.

Yeh, C. (2018). Revisiting the Origins of Business Cycles with the Size-Variance Relationship. *Working Paper*.