RIETI Discussion Paper Series 21-E-066

The Size of Micro-originated Aggregate Fluctuations: An analysis of firm-level input-output linkages in Japan (Revised)

Yoshiyuki ARATA
RIETI

MIYAKAWA, Daisuke
Hitotsubashi University

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 21-E-066

First Draft: August 2021

Revised: August 2022

The size of micro-originated aggregate fluctuations: an analysis of firm-level input-output linkages in Japan[1]

Yoshiyuki Arata[2]

Research Institute of Economy, Trade and Industry

Daisuke Miyakawa

Hitotsubashi University

Abstract

Recent studies (e.g., Acemoglu et al. (2012)) argue that microeconomic shocks propagate through input-output linkages and result in aggregate fluctuations. This paper challenges this view by quantifying the size of the micro-originated aggregate fluctuations. Our analysis relies on firm-level input-output data for Japan and non-asymptotic probabilistic results, which characterize the variance, tail probability, and distribution shape of aggregate output induced by microeconomic shocks. We find that microeconomic shocks contribute substantially to the aggregate variance but almost nothing to the tail probability of aggregate output. Furthermore, the distribution of aggregate output induced by microeconomic shocks turns out to be very close to a Gaussian, even though the central limit theorem does not hold. Since the distribution of the empirical GDP growth rates has a heavier tail than a Gaussian, microeconomic shocks cannot explain the observed extremes of the GDP growth rates.

Keywords: Micro-origin of aggregate fluctuations; Granular hypothesis; Input-output network

[1]This study is conducted as a part of "Study Group on Corporate Finance and Firm Dynamics" at RIETI. The data used in this paper is made available through the contract between Hitotsubashi University and Tokyo Shoko Research. Arata appreciates the financial support from JSPS (KAKENHI Grant Numbers: 18K12788 and 21K13265). Miyakawa appreciates the financial supports from JSPS (KAKENHI Grant Numbers: 21K01438) and JST-Mirai Program. Conflict of interest: none.

[2]Corresponding author: y.arata0325@gmail.com

1 Introduction

One of the essential features of the modern economy is that firms are connected by transaction relationships, forming a huge input-output network (see **Figure 1**). The last decade has witnessed a surge in research exploring the implications of this network for aggregate fluctuations. An influential paper by Acemoglu et al. (2012) provides a theoretical underpinning for the micro-origin of aggregate fluctuations, showing that microeconomic shocks propagate on the network and drive substantial aggregate fluctuations. They argue that due to the heterogeneity of the network structure, shocks to hub firms in the network do not die out and thus are not negligible even at the aggregate level. This idea is closely related to the granular view proposed by Gabaix (2011), who emphasizes the high heterogeneity of firms in terms of their size. Both ideas have been widely accepted in the recent literature as a new explanation for the source of aggregate fluctuations.[1]

In this paper, we challenge the granular view by focusing on the *non-asymptotic* properties of micro-originated aggregate fluctuations. Most previous studies supporting the granular view are based on asymptotic results as the number of firms (denoted by N) goes to infinity: the decay rate of the aggregate variance (Gabaix (2011)), the decay rate of the tail probability of aggregate output (Acemoglu et al. (2017)), and the condition of the central limit theorem (CLT) (Acemoglu et al. (2012)). However, these asymptotic results do not necessarily imply their corresponding non-asymptotic results with N fixed. For example, the tail probability of aggregate output may be negligible with a fixed N, while it decays very slowly as N goes to infinity. The distribution of aggregate output may be sufficiently close to and economically indistinguishable from a Gaussian with a fixed N, even though the CLT does not hold. We show that the distinction between asymptotic and non-asymptotic properties is crucial to assessing the empirical relevance of the granular view.

Using firm-level input-output data for Japan, we find that the granular view has no explanatory power for the large deviation of the GDP growth rates. More precisely, microeconomic shocks contribute substantially to the aggregate variance but almost nothing to the tail probability of aggregate output. This result is related to the distribution shape of aggregate output. Given the observed heterogeneity of sizes and network structure across firms, the distribution of aggregate output induced by microeconomic shocks turns out to be very close to a Gaussian, even though the CLT does not hold. In light of the empirical fact that the distribution of the GDP growth rates deviates from a Gaussian in the tail region, the granular view cannot explain the observed extremes of the GDP growth rates.

We begin with the probabilistic methods to assess the role of the granularity in aggregate fluctuations non-asymptotically. More concretely, we analyze the distribution properties of the weighted sum of independent and identically distributed (iid) shocks. The iid shocks represent idiosyncratic shocks to firms, and the weights are given by the general equilibrium models proposed by Acemoglu et al. (2012) and Baqaee

[1] We use "the granular view" and "the micro-origin of aggregate fluctuations" interchangeably in the following.

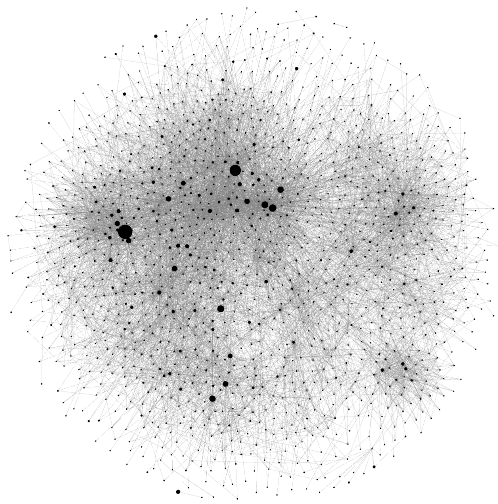


Figure 1: Input-output network in Japan in 2018. For data sources, see Section 3. Only firms with sales of more than 100 billion yen and their input-output linkages are included. The size of each circle represents the size of the firm's sales.

and Farhi (2020b), in which the structure of the input-output network determines the weights, that is, firms lying at the center of the network have a large weight. We focus on the non-asymptotic properties of the variance, the tail probability, and the distribution shape of the weighted sum with the weights fixed. We use a concentration inequality to obtain the upper bound of the tail probability and show that the upper bound is determined solely by the largest weight. To approximate the distribution of the weighted sum, we use the Edgeworth series approximation, which is based on its cumulant generating function. Using this method, we can measure how far the distribution of the weighted sum is from a Gaussian. Thus, when the weights are calculated (e.g., from an empirical input-output network), these methods enable us to analyze the distribution properties of the weighted sum with the given weights rather than their asymptotic properties at the limit of N.

For the empirical input-output network, we use firm-level data compiled by Tokyo Shoko Research (TSR). This dataset contains firm-level information, such as sales, and the identification of the suppliers and customers of firms, from which we construct an input-output network covering more than $300,000$ firms in Japan. As is well known in the literature, the network structure is highly heterogeneous: the number of linkages is heterogeneous across firms, and some large firms play the role of hubs in the network. Following the models by Acemoglu et al. (2012) and Baqaee and Farhi (2020b), we calculate the weights for each firm from this empirical network. By combining these empirical weights and the non-asymptotic methods, we quantify the size of the micro-originated aggregate fluctuations with the given input-output network fixed.

We find that microeconomic shocks are an important source of the aggregate variance because of the high heterogeneity of the weights. The weights calculated from the empirical input-output network follow a distribution with a Pareto tail with an exponent close to 1. That is, there exist firms having a disproportionate impact on aggregate output, and shocks to such firms do not die out even at the aggregate level. For this reason, the CLT does not hold, meaning that the distribution of aggregate output induced by microeconomic shocks does not converge to a Gaussian. Moreover, we find that this high heterogeneity comes mainly from the input-output network; that is, Leontief's inverse matrix of the network is the main factor determining the heterogeneity of the weights. These results are consistent with previous studies supporting the granular view and, in particular, the finding by Magerman et al. (2017) for the Belgian economy, showing that an input-output network is an importance source of the aggregate variance.

However, our finding suggests that, in contrast to the aggregate variance, microeconomic shocks contribute almost nothing to the tail probability of aggregate output. The tail probability of aggregate output induced by microeconomic shocks is negligible compared to its empirical counterparts (i.e., the tail probability of the GDP growth rates). Put differently, the heterogeneity of the weights (i.e., the heterogeneity of the input-output network) is too low to explain the observed extremes of the GDP growth rates.

This difference between the variance and tail probability of aggregate output can be explained by the distribution shape of aggregate output. Our use of the Edgeworth series approximation reveals that the distribution of aggregate output induced by microeconomic shocks is very close to a Gaussian given the empirical weights. This finding means that despite the heterogeneity of the weights, microeconomic shocks to a considerable extent cancel each other out; that is, the empirical granularity found in Japan is not sufficiently large to completely prevent this averaging effect. It is worth mentioning that this result does not contradict the asymptotic result that the distribution of aggregate output does not converge to a Gaussian as N goes to infinity. This asymptotic result makes a statement only about the limiting behavior of the distribution at the limit of N but not about how far the distribution is from a Gaussian with N fixed. Our analysis reveals that given the observed granularity in Japan, the distribution of aggregate output is not a Gaussian at the limit of N but practically close enough to a Gaussian with the given N. This finding, obtained by focusing on the non-asymptotic properties of the distribution of aggregate output, sheds new light on the granular view and is the key contribution of our paper. Since the non-asymptotic results are more important from an empirical perspective, we conclude that microeconomic shocks only cause small fluctuations and not a large deviation of aggregate output.

Related literature

Our paper belongs to the literature on the microeconomic origins of aggregate fluctuations. Gabaix (2011) proposes the granular view and shows that if the firm size distribution has a Pareto tail, the aggregate variance decays more slowly than predicted by the CLT. Acemoglu et al. (2012) analyze the slow decay rate of the aggregate variance when an input-output network is explicitly considered. They also provide the CLT condition, in which the convergence to a Gaussian is determined by the contribution of shocks to the largest firm to the aggregate variance. Related to this result, Arata (2021) shows that this variance contribution of the largest firm accounts for a significant part of the aggregate variance even at the limit of N when the distribution of the weights has a Pareto tail. The study closest to ours is Acemoglu et al. (2017), who analyze the tail probability of aggregate output and argue that microeconomic shocks can generate a large deviation of aggregate output. These previous studies focus on the asymptotic properties of the distribution of aggregate output as N goes to infinity. In contrast, the contribution of our paper is to show that when focused on the non-asymptotic properties, the granular view cannot explain the observed aggregate fluctuations, especially the observed extremes of aggregate output. In other words, the asymptotic results discussed in the previous studies do not necessarily imply the empirical relevance of the granular view.

While the studies above consider an efficient economy, recent theoretical studies also propose models for an inefficient economy. For instance, Jones (2011) and Jones (2013) introduce exogenous wedges and study the relation between input-output linkages and economic growth. Meanwhile, Bigio and La'o (2020), Su (2019), Luo (2020), and Altinoglu (2021) consider inefficiencies induced by financial frictions. Further, Baqaee and Farhi (2020b) provide a general framework for an inefficient economy with exogenous wedges, which nests the model in Acemoglu et al. (2012) as a special case. Our empirical analysis uses their model and quantifies the size of micro-originated aggregate fluctuations with inefficiencies.[2][3]

Our paper is related to the empirical literature on micro-originated aggregate fluctuations. Carvalho (2010), Carvalho and Gabaix (2013), and Stella (2015) use the fact that firms' sales are sufficient statistics for aggregate fluctuations in an efficient economy and test the relevance of microeconomic shocks. Related to them, a line of studies, including Di Giovanni et al. (2014), Di Giovanni et al. (2018), and Di Giovanni et al. (2019), decompose the observed aggregate fluctuations into exogenous aggregate shocks and microeconomic shocks. Another strand of studies (Foerster et al. (2011); Atalay (2017); Atalay et al. (2018)) estimates the underlying microeconomic shocks by filtering the time series of aggregate variables through a structural model.

[2]Some recent studies introduce extensive margin, that is, changes in input-output linkages and entry/exit of firms (Grassi (2017); Burstein et al. (2020);Baqaee (2018);Acemoglu and Tahbaz-Salehi (2020);Taschereau-Dumouchel (2020);Baqaee and Farhi (2020a). Although the extensive margin generates a more complex relation between aggregate output and microeconomic shocks, it is difficult to construct its empirical counterparts because the characterization of the extensive margin depends on unobservable parameters. In our analysis, we focus on the intensive margin only.

[3]Another important extension is to incorporate the higher-order terms of microeconomic shocks into the model. Baqaee and Farhi (2019) consider the CES technology and analyze the second-order terms of microeconomic shocks. Dew-Becker (2022) considers all higher-order terms of microeconomic shocks and characterizes the tail probability of aggregate output. Since our analysis is based on the first-order terms both in efficient and inefficient cases, our finding that the distribution of aggregate output is very close to a Gaussian suggests that these higher-order terms are needed to explain the observed features of the GDP growth rates. See also the discussion in the Conclusion.

It should be noted, however, that empirical works that directly analyze the size of micro-originated aggregate fluctuations based on an empirical firm-level input-output network are very sparse.[4] To the best of our knowledge, the exception is Magerman et al. (2017), who analyze an empirical input-output network constructed from Belgian tax data. While Magerman et al. (2017) argue the importance of microeconomic shocks using the aggregate variance as the measure of aggregate fluctuations, our paper, by focusing not only on the variance but on the tail probability of aggregate output, shows that the role of microeconomic shocks in explaining aggregate fluctuations is limited.

Outline

This paper is organized as follows. In Section 2, we provide probabilistic methods that characterize the distribution of aggregate output and review multi-sector models. In Section 3, we provides an overview of our data used in Section 4. In Section 4, we give our main empirical results. In Section 5, we conclude. In the Appendix, we give mathematical proofs and details of our data.

2 Quantitative Model

We provide a quantitative method to measure the size of the micro-originated aggregate fluctuations. In Section 2.1, we provide probabilistic methods that characterize the distribution of aggregate output induced by microeconomic shocks. In Section 2.2, we review the multi-sector models used in Acemoglu et al. (2012) and Baqaee and Farhi (2020b).

2.1 Probabilistic Methods

2.1.1 Variance and CLT

Consider an economy in which aggregate output Z is represented as the weighted sum of iid microeconomic shocks $\epsilon_1, ..., \epsilon_N$ with weights $w_1, ..., w_N$:

$$Z := \sum_{i=1}^{N} w_i \epsilon_i \tag{1}$$

where N is the number of firms and w_i represents the impact of i's microeconomic shock on aggregate output.[5] Let σ_ϵ^2 be the variance of microeconomic shocks. Thus, the variance of Z (denoted by σ_Z^2) is equal to $\sigma_\epsilon^2 \|w\|_2^2$, where $\|w\|_2$ is the ℓ^2-norm of the weights (i.e., $\|w\|_2 := \sqrt{\sum_i w_i^2}$). The simplest example of Eq.(1) is the homogeneous economy with no input-output linkages, that is, $w_i = 1/N$ for all i. In this case,

[4] For a sector-level analysis, Miranda-Pinto (2021) uses the OECD data and calculates Leontief's inverse matrix of the input-output network. He empirically shows that the diversification of the input-output network (i.e., a more homogenous network) is the key to the decrease in the variance of the GDP growth rates.

[5] The analytical expression of $w_1, ..., w_N$ in multi-sector models is given in Section 2.2.

the CLT holds, and Z converges to a Gaussian distribution with the variance decaying at the rate of $N^{-1/2}$. That is, as the number of firms increases, micro-originated aggregate fluctuations decay rapidly.

In contrast, when the weights are highly heterogeneous, the rapid decay of the aggregate fluctuations does not occur. Gabaix (2011) shows that when the distribution of the weights has a Pareto tail with an exponent $1 \leq \alpha < 2$, σ_Z^2 decays more slowly than $N^{-1/2}$. Furthermore, the heterogeneity of the weights, especially the largest weight, is related to the CLT and the convergence to a Gaussian distribution. Let $\|w\|_\infty$ denote ℓ^∞-norm of the weights (i.e., $\|w\|_\infty := \max_i(w_1, w_2, ..., w_N)$). Let $r_{\max}$ denote the ratio of the two (squared) norms of the weights:

$$r_{\max} := \frac{\|w\|_\infty^2}{\|w\|_2^2}$$

Since the variance contribution of the firm with the largest weight to σ_Z^2 is given by $\sigma_\epsilon^2 \|w\|_\infty$, $r_{\max}$ represents the fraction of σ_Z^2 attributable to this firm. Acemoglu et al. (2012) show that the convergence of Z to a Gaussian is determined by the convergence of $r_{\max}$ to 0.

Theorem 2.1 (Theorem 1(c) in Acemoglu et al. (2012))**.** *Suppose that $\epsilon_1, ..., \epsilon_N$ are not Gaussian random variables and $\lim_{N\to\infty} r_{\max} > 0$. Then, $\frac{1}{\sigma_\epsilon \|w\|_2} Z$ does not converge to a Gaussian distribution.*

Why does the CLT not hold when $\lim_{N\to\infty} r_{\max} > 0$? To understand its mechanism, recall Cramer's decomposition theorem (Cramér (1936); see also Linnik and Ostrovskii (1977)): If a Gaussian random variable ξ admits decomposition as the sum $\xi = \xi_1 + \xi_2$ of two independent random variables, then both ξ_1 and ξ_2 are Gaussian. More generally, if the sum of independent random variables follows a Gaussian distribution, each random variable must be Gaussian. The condition in Theorem 2.1 means that the firm with the largest weight, which accounts for a significant part of Z, does not follow a Gaussian even at the limit of N. This explains why the asymptotic behavior of $r_{\max}$ is crucial for the (non-)convergence of Z to a Gaussian.

The next question is: when does the condition $\lim_{N\to\infty} r_{\max} > 0$ hold? Arata (2021) shows that the asymptotic behavior of $r_{\max}$ is related to the tail of the distribution of $w_1, ..., w_N$.

Proposition 2.2 (Proposition 3.5 in Arata (2021))**.** *(i) If the distribution of $w_1, ..., w_N$ has a finite second moment, then*

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

(ii) If the distribution of $w_1, ..., w_N$ has a Pareto tail with the tail exponent $\alpha \in (0, 2)$, then $\lim_{N\to\infty} r_{max}$ is a non-degenerate random variable and its expectation is strictly positive:

$$\lim_{N\to\infty} E r_{\max} > 0$$

This result explains how the CLT is related to the heterogeneity of the weights. When the heterogeneity of the weights is not high, the variance contribution of the firm with the largest weight becomes negligible as N increases. Since each shock has a negligible impact on aggregate output, the distribution of Z converges to a Gaussian. In contrast, when the weights are highly heterogeneous, this variance contribution does not diminish. Even when N is large, shocks to the firm with the largest weight dominate aggregate output and cannot be canceled by shocks to other firms. The resultant distribution of Z depends on shocks to the firm with the largest weight, and the distribution of Z does not converge to a Gaussian. Once it is confirmed that the weights (e.g., firm sizes) follow a distribution with a Pareto tail with exponent α sufficiently close to 1, we can conclude that the CLT does not hold because of the high heterogeneity of the weights. This logic serves as the foundation for the granular view in the literature.

However, it should be noted that the results above are asymptotic (i.e., the properties of Z as N goes to infinity). In general, we can not deduce the non-asymptotic properties of Z from its asymptotic properties. In particular, the asymptotic result regarding the (non-)convergence of Z to a Gaussian says nothing about how far the distribution of Z with a given N is from a Gaussian. There remains a possibility that given a fixed N, the distribution of Z is very close to a Gaussian, while the CLT does not hold. For example, suppose that $r_{\max}$ does not converge to 0 but is very small or that the distribution of microeconomic shocks is different from but very close to a Gaussian. The resultant distribution of Z would be arbitrarily close to a Gaussian. In this case, although the CLT would not hold, it would not be possible to argue the relevance of the granular view. To assess the empirical relevance of the granular view, we need to analyze the non-asymptotic properties of Z with the weights fixed. In Section 2.1.3, we explain how to analyze the non-asymptotic properties of Z. Before that, we show below that the distinction between the asymptotic and non-asymptotic results is also important for the tail probability of Z.

2.1.2 Tail probability

Let us consider the tail probability of Z in Eq.(1), that is, the probability that microeconomic shocks lead to a large deviation of aggregate output. For the simplest example, consider the homogeneous economy with weight $w_i = 1/N$ for all i. Furthermore, since the tail probability of Z also depends on the distribution of microeconomic shocks, assume that they are drawn from the standard Gaussian distribution. Then, the normalized aggregate output $\sqrt{N}Z$ is also a Gaussian, and thus, we have

$$P(Z \leq -x) = \Phi(-\sqrt{N}x)$$

Thus, using Mills' ratio, we obtain the convergence ratio of the tail probability of Z:

$$\frac{1}{N}\log P(Z \leq -x) \to -\frac{x^2}{2} \tag{2}$$

as $N \to \infty$.[6] That is, under this condition, the tail probability of Z decays at the rate of $e^{-Nx^2/2}$ as $N \to \infty$. Note that while the typical value of Z is of the order of $1/\sqrt{N}$ and the variance decays at the rate of N^{-1},

[6]Mills' ratio for the standard Gaussian distribution means that its tail probability can be approximated by using its density: as $x \to \infty$,

$$\frac{1-\Phi(x)}{\phi(x)} \sim \frac{1}{x}$$

where Φ and ϕ are the standard Gaussian distribution and its density function, respectively.

the tail probability of Z decays at the exponential rate. The difference in the convergence rates shows that microeconomic shocks are less likely to contribute to the tail probability of aggregate output than to the aggregate variance when the heterogeneity is low.

This rapid decay rate depends on the degree of heterogeneity and the distribution of microeconomic shocks. For the distribution of microeconomic shocks, following Acemoglu et al. (2017), we consider a Laplace distribution. The Laplace distribution has a heavier tail than a Gaussian but still has finite moments of all orders.[7]

Assumption 2.1. Microeconomic shocks follow a Laplace distribution with mean 0 and variance $2b^2$, whose probability density function is given by

$$d(x; b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$$

Given this assumption, Acemoglu et al. (2017) show that the convergence rate of the tail probability of Z is determined by that of $\sqrt{r_{\max}}$. That is, when the dominance of the largest weight (i.e., $\|w\|_\infty$) in the aggregate variance (i.e., $\|w\|_2$) does not diminish quickly, the tail probability of Z decays more slowly than Eq.(2). However, this asymptotic result is not suitable for the question of whether the heterogeneity of weights given by empirical data is sufficiently high to generate the non-negligible tail probability of Z. Instead, the following non-asymptotic result gives us the formula to measure the tail probability of Z with the weights fixed.

Proposition 2.3. *Under Assumption 2.1, we have*

$$P(Z \leq -x) \leq \exp\left(-\frac{x^2}{4b^2\|w\|_2^2 + 2b\|w\|_\infty x}\right) \tag{3}$$

Proof. See the Appendix. □

Eq.(3) shows that the two statistics of weights, $\|w\|_2$ and $\|w\|_\infty$, are sufficient statistics for the upper bound of the tail probability of Z. Intuitively, for small x, the tail probability is bounded by a Gaussian decay with $\|w\|_2$, similar to Eq.(2). For large x, because of Assumption 2.1, the tail probability deviates from the Gaussian decay and is controlled by an exponential decay with $\|w\|_\infty$. In particular, if microeconomic shocks generate a large deviation of aggregate output, $\|w\|_\infty$ (i.e., the largest weight) must be sufficiently large. In Section 4.2, substituting the empirical values of $\|w\|_2$ and $\|w\|_\infty$ into Eq.(3), we test whether the heterogeneity of the weights is sufficiently high to generate a non-negligible tail probability of aggregate output.

2.1.3 Distribution approximation

Proposition 2.1 describes the condition for the (non-)convergence of Z to a Gaussian. However, even when Z does not converge to a Gaussian, this does not necessarily mean that the distribution of Z is far from

[7] For the empirical validity of this assumption, see Section 3.2.2.

a Gaussian. In the following, we discuss how to measure the closeness of the distribution of Z to a Gaussian with given weights.

One useful way of comparing two distributions is to use their cumulants. The cumulants of a random variable are moment related quantities that are uniquely determined by the distribution. Formally, the jth order cumulant κ_j of a random variable X is the jth order derivative of the cumulant generating function K_X at 0, where K_X is defined as

$$K_X(t) := \log Ee^{tX}$$

The first- and second-order cumulants are equal to the mean and variance of X, respectively. The fourth-order cumulant κ_4 is related to the excess kurtosis; the excess kurtosis is a measure of the peakedness and tailedness of its probability density and is defined by $\mu_4/\mu_2^2 - 3$, where μ_k is the kth central moment. When X is normalized, κ_4 is equal to the excess kurtosis.

The cumulants have several useful properties. If X_1 and X_2 are independent random variables, the cumulant of $X_1 + X_2$ satisfies additivity:

$$\kappa_j(X_1 + X_2) = \kappa_j(X_1) + \kappa_j(X_2)$$

In addition, $\kappa_j(aX)$ for some scalar a is equal to $a^j \kappa_j(X)$.

Let us consider concrete examples of cumulants. Suppose that X follows a Gaussian distribution with mean 0 and variance σ_X^2. Then, one can show that $\kappa_1 = 0$, $\kappa_2 = \sigma_X^2$, and $\kappa_j = 0$ for $j \geq 3$. In particular, the Gaussian distribution has no excess kurtosis (i.e., $\kappa_4 = 0$). This means that if the cumulant κ_j for $j \geq 3$ is close to 0, the distribution of X is close to a Gaussian. That is, the deviation of cumulants for $j \geq 3$ from 0 can be used to measure how far the distribution of X is from a Gaussian. For another example, suppose that X follows a Laplace distribution with mean 0 and variance $2b^2$. Then, its cumulant κ_j is given by

$$\kappa_j = \begin{cases} 0, & \text{if } j \text{ is odd} \\ 2(j-1)!b^j, & \text{if } j \text{ is even} \end{cases}$$

In particular, if X is normalized, the excess kurtosis (i.e., κ_4) equals 3. As shown in **Figure 2**, the probability density function is sharply peaked at the center and has a heavier tail than a Gaussian.

The properties of cumulants are useful in the analysis of the weighted sum of independent random variables. Consider the normalized version of Eq.(1):

$$Z^* = \sum_i w_i^* \epsilon_i^* \tag{4}$$

where $w_i^* := w_i/\|w\|_2$ and $\epsilon_i^* := \epsilon_i/\sigma_\epsilon$. The properties of the cumulants above yield that

$$\kappa_j(Z^*) = \kappa_j(\epsilon^*) \sum_i w_i^{*j} \tag{5}$$

This equation means that the cumulants of the normalized aggregate output Z^* can be separated into two factors: $\kappa_j(\epsilon^*)$ is the cumulant of the normalized microeconomic shocks, and $\sum_i w_i^{*j}$ summarizes the effect of the heterogeneity of the weights on the jth-order cumulant. For example, when an economy consists of a

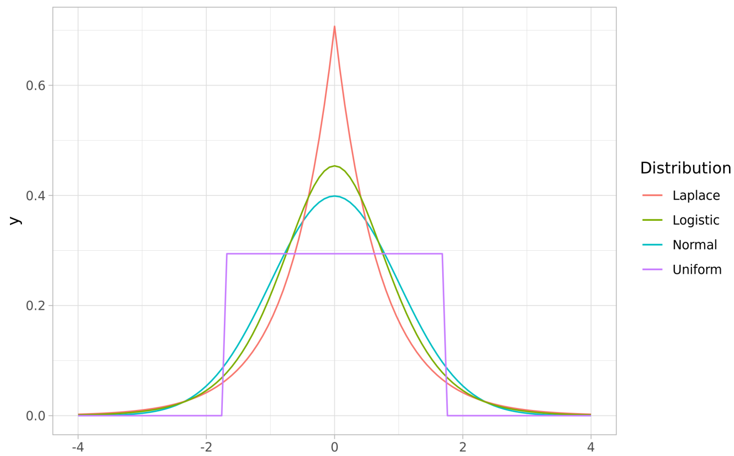


Figure 2: The comparison of the probability density functions with different kurtosis. The densities of normal ($\kappa_4 = 0$), Laplace ($\kappa_4 = 3$), logistic ($\kappa_4 = 1.2$), and uniform ($\kappa_4 = -1.2$) distributions with the variance set to 1 are plotted.

single firm, $\sum_i w_i^{*j} = 1$, and the cumulants of Z^* are equivalent to those of ϵ^*. On the other hand, for the case of the homogeneous economy (i.e., $w_i = 1/N$ for all i), $\sum_i w_i^{*j}$ is equal to $N^{-j/2+1}$, and the jth-order cumulant of Z^* decays at this rate. That is, by calculating the factor $\sum_i w_i^{*j}$, we can measure how much the jth-order cumulant of Z^* approaches 0 (i.e., the cumulant of a Gaussian) through the summation in Eq.(4).

Note that the cumulants are defined as the coefficients of Taylor's expansion of the cumulant generating function around 0. Thus, if a sequence of cumulants is given, we can approximate the original cumulant generating function. Since the cumulant generating function uniquely determines its corresponding distribution, we can reconstruct the distribution of Z^* from a series of its cumulants. This method, called Edgeworth series approximation, is usually stated as an asymptotic expansion of the sample mean of iid random variables:[8]

Proposition 2.4 (See Chapter 2 in Hall (2013))**.** *Let S_n be the sample mean of n iid random variables. The*

[8]More precisely, let f be the probability density function that we aim to approximate, and let ϕ be a reference density function such as a Gaussian. Their characteristic function is given as follows:

$$\widehat{f}(t) = \exp\left[\sum_{j=1}^{\infty} \kappa_j \frac{(it)^j}{j!}\right], \quad \widehat{\phi}(t) = \exp\left[\sum_{j=1}^{\infty} \gamma_j \frac{(it)^j}{j!}\right]$$

where κ_j and γ_j are the jth-order cumulants of f and g, respectively. When the density of the standard Gaussian distribution with the same mean and variance as f is used as the reference density, we obtain the following formal identity:

$$\widehat{f}(t) = \exp\left[\sum_{j=3}^{\infty} \kappa_j \frac{(it)^j}{j!}\right] \widehat{\phi}(t)$$

The Edgeworth series approximation is based on the inversion of this identity. Finally, terms used for approximation are arranged according to the order of n, which yields Eq.(6).

Edgeworth series approximation for the distribution of S_n is given by

$$P(S_n \leq x) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + \ldots + n^{-j/2} p_j(x)\phi(x) + \ldots \quad (6)$$

where Φ and ϕ are the distribution and probability density of the standard Gaussian, and p_j is a polynomial of degree $3j - 1$. The Edgeworth series approximation for the probability density of S_n is given by the first derivative of Eq.(6).

The right-hand side of Eq.(6) is the expansion around the Gaussian distribution (the reference distribution), and its correction terms are grouped according to the order of n. The correction terms are determined by the cumulants of the random variable. For example, $p_1(x)$ and $p_2(x)$ are given as follows:

$$p_1(x) = -\frac{1}{6}\kappa_3(x^2 - 1)$$

$$p_2(x) = -x\left\{\frac{1}{24}\kappa_4(x^2 - 3) + \frac{1}{72}\kappa_3^2(x^4 - 10x^2 + 15)\right\}$$

Formally, we apply Proposition 2.4 with $n = 1$ to Z^* in our case. In addition, since the distribution of Z^* is symmetric in our setting, the cumulants for odd j are equal to 0. Thus, the leading term in the correction terms is determined by the fourth-order cumulant κ_4, that is, the excess kurtosis. Put differently, when the excess kurtosis is close to 0, the distribution is expected to be close to the standard Gaussian. In Section 4.3, given the empirical weights, we calculate the excess kurtosis and approximate the distribution of Z^* by Eq.(6).[9]

2.2 Multi-Sector Model

The results in Section 2.1 show how micro-originated aggregate fluctuations are related to the heterogeneity of the weights in Eq.(1). Here, we review the analytical expression of the weights in the models proposed by Acemoglu et al. (2012) and Baqaee and Farhi (2020b). Consider a static economy with N firms, each of which produces a distinct product by using labor and intermediate goods from other firms. The production technology of firm i is represented by the CES production function with the elasticity of substitution θ_i and Hick's neutral productivity A_i. Firm i sets its price so that it is equal to its markup μ_i times marginal cost. Firms' profits are rebated to the final consumer. The final expenditure (i.e., the nominal GDP) is equal to the sum of labor income and profits rebated from firms.

Let us introduce our notation. Let β_i be the final expenditure share for good i, that is, the share of the spending on good i in the total expenditure. Let $\Omega := \{\Omega_{ij}\}$ denote a revenue-based input-output matrix, where Ω_{ij} is the share of the cost of good j in firm i's sales revenue. Similarly, let $\widetilde{\Omega} := \{\widetilde{\Omega}_{ij}\}$ denote a cost-based input-output matrix, where $\widetilde{\Omega}_{ij}$ is the share of the cost of good j in firm i's total costs. The two matrices satisfy

$$\widetilde{\Omega} = \mathrm{diag}(\mu)\Omega$$

[9] In Section 4.3, we use the first 20 cumulants of Z^* (i.e., κ_j for $j \leq 20$) for the correcting terms in Eq.(6).

where $\text{diag}(\mu)$ is the diagonal matrix with diagonal element $\mu_1, ..., \mu_N$. Let λ be the vector of Domar weights defined by the ratio of firm i's sales to GDP. From the accounting identity, λ satisfies

$$\lambda' = \beta'(I - \Omega)^{-1} \tag{7}$$

Using this relation, we define a counterpart of λ where Ω is replaced with $\widetilde{\Omega}$:

$$\widetilde{\lambda}' := \beta'(I - \widetilde{\Omega})^{-1}$$

Finally, let $\Psi_L := (I - \Omega)^{-1}L$ and $\Lambda_L := \beta'\Psi_L$, where L is the vector of labor shares in sales.

Let us return to Eq.(1). Here, Z and ϵ_i represent the logarithms of GDP and A_i, respectively, and the weights depend on the model considered. First, consider the model of an efficient economy with $\theta_i = 1$ for all i. Acemoglu et al. (2012) show that in a competitive equilibrium, the weights in Eq.(1) are given by Domar weights:

$$w' = \lambda' = \beta'(I - \Omega)^{-1} \tag{8}$$

Eq.(8) shows that the weight of firm i's shock is determined by two components: the network position represented by Leontief's inverse matrix $(I - \Omega)^{-1}$ and preference β. Intuitively, shocks to firms lying at the center of the network or producing goods important for the representative household have a larger impact on aggregate output. Following Acemoglu et al. (2017), we introduce the notion of two types of heterogeneity: network heterogeneity and primitive heterogeneity. Network heterogeneity is defined by the weights in Eq.(8) with $b_i = 1/N$ for all i:

$$\lambda'_{\text{net}} := \frac{1}{N}1'(I - \Omega)^{-1} \tag{9}$$

That is, λ'_{net} represents the case in which the heterogeneity of the weights comes only from the structure of the input-output network. Similarly, primitive heterogeneity is defined by the weights in Eq.(8) when there is no heterogeneity coming from the input-output network:

$$\lambda'_{\text{prm}} = \beta'$$

Thus, by comparing λ_{net} and λ_{prm}, we can quantify which heterogeneity is the main source of the heterogeneity of the weights.

Next, consider an inefficient economy with markup, that is, some firms earn positive profits (i.e., $\mu \neq 1$) and the two input-output matrices do not coincide (i.e., $\Omega \neq \widetilde{\Omega}$). Under this setting, Baqaee and Farhi (2020b) give an analytical expression of the weights for arbitrary θ_i. Specifically, when $\theta_i = 0$ or 1 for all i

(i.e., the Cobb-Douglas and Leontief cases), the expression of the weights is simplified as follows:[10] [11]

$$w_i = \widetilde{\lambda}_i \qquad \text{for } \theta_i = 1 \text{ for all } i$$

$$w_i = \lambda_i \frac{\Psi_{L,i}}{\Lambda_L} \qquad \text{for } \theta_i = 0 \text{ for all } i$$

Note that with the help of these theoretical results, we can calculate the empirical weights at the firm level, given input-output network data and individual firms' information such as sales and profits. The next section provides an overview of our firm-level data and provides basic properties of the input-output network.

3 Overview of Data

We provide an overview of our data. In Section 3.1, we describe the distribution of Japan's GDP growth rates. In Section 3.2, we describe our firm-level data for Japan and analyze their three crucial features: the high heterogeneity of firms' sales, the distribution of microeconomic shocks, and the connectivity of the input-output network.

[10] Although the main reason for the choice of the two cases is the simplicity of the expression, there is another reason for including Leontief's case. Recent empirical studies show that the response to a negative shock from its suppliers is well described by the technology with low substitutability, especially in the short term. For example, Boehm et al. (2019) show that after the earthquake in Japan, firms in the US that imported intermediate goods from the damaged area were unable to substitute their intermediates with other goods and reduced their production.

[11] When only a single factor (i.e., labor) is used, Equation (10) in Baqaee and Farhi (2020b) reduces to

$$\frac{\partial \log Y}{\partial \log A_i} = \widetilde{\lambda}_i - \frac{\partial \log \Lambda_L}{\partial \log A_i}$$

$$\frac{\partial \log \Lambda_L}{\partial \log A_i} = \sum_j \frac{\lambda_j}{\mu_j}(\theta_j - 1)\mathrm{Cov}_{\widetilde{\Omega}(j)}\left(\widetilde{\Psi}_{(i)}, \frac{\Psi_{(L)}}{\Lambda_L}\right)$$

Furthermore, the response of the markup/wedge shocks (i.e., Equation (13) in Baqaee and Farhi (2020b)) is reduced to

$$\frac{\partial \log Y}{\partial \log \mu_i} = -\widetilde{\lambda}_i - \frac{\partial \log \Lambda_L}{\partial \log \mu_i}$$

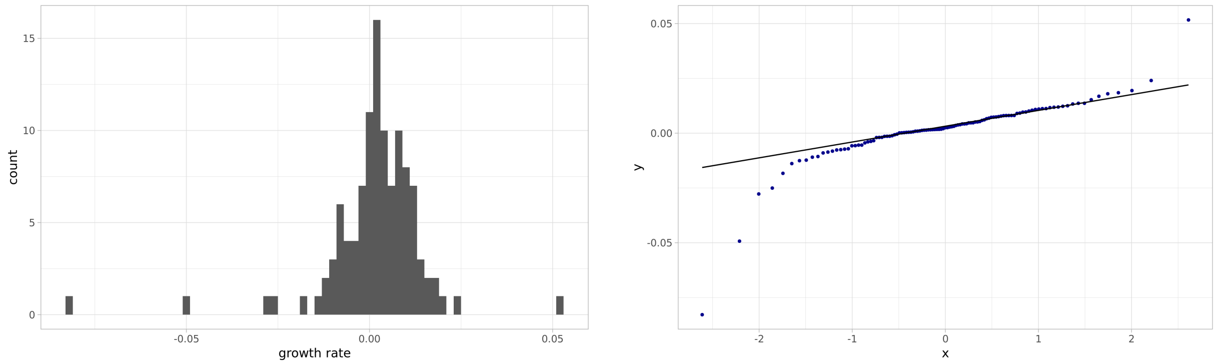
$$\frac{\partial \log \Lambda_L}{\partial \log \mu_i} = -\sum_j \frac{\lambda_j}{\mu_j}(\theta_j - 1)\mathrm{Cov}_{\widetilde{\Omega}(j)}\left(\widetilde{\Psi}_{(i)}, \frac{\Psi_{(L)}}{\Lambda_L}\right) - \lambda_i \frac{\Psi_{iL}}{\Lambda_L}$$

Proposition 4 in Baqaee and Farhi (2020b) states that if $\theta_i = 0$ for all i, then

$$\frac{\partial \log Y}{\partial \log \mu_i} = 0$$

Combining these equations, we obtain

$$\frac{\partial \log Y}{\partial \log A_i} = \lambda_i \frac{\Psi_{iL}}{\Lambda_L}$$



(a) Histogram

(b) QQ-plot

Figure 3: The histogram and QQ-plot of the GDP growth rate. In (b), the straight line represents the hypothesized Gaussian distribution.

3.1 Aggregate Fluctuations

We study Japan's quarterly GDP time series (seasonally-adjusted and constant prices) from 1994Q1 to 2021Q4 taken from the OECD database. Let g_t denote the log difference of the GDP in successive periods, that is, $g_t := \log \mathrm{GDP}_t - \log \mathrm{GDP}_{t-1}$. We call g_t the GDP growth rate, which corresponds to Z in Eq.(1). The summary statistics of the GDP growth rates are given in **Table 1**.

	count	mean	s.d.	s.d.(mad)	min	max
g_t	111	0.0017	0.0136	0.0078	-0.0828	0.0517

Table 1: The summary statistics of the GDP growth rate from 1994Q1 to 2020Q4. In the table, s.d.(mad) represents the estimate of the standard deviation based on the median absolute deviation, that is, $1.4826 \times \text{median}(|g_t - \text{median}(g_t)|)$. This is a consistent estimator when g_t follows a Gaussian distribution.

Our focus is on the distribution of the GDP growth rates, especially in the tail region. **Figure 3** depicts the histogram and QQ-plot of the GDP growth rates. The graphical inspection suggests that the distribution of the GDP growth rates does not follow a Gaussian but has a heavier tail. That is, the probability of large deviations of the GDP growth rate is larger than expected by a Gaussian. We perform the normality tests, confirming that the departure from a Gaussian is statistically significant.[12]

Next, we quantify the tail probability of the GDP growth rates, that is, $P(g < -x)$. The simplest estimate of the tail probability is the empirical counter cumulative distribution function (CCDF) defined by $\frac{1}{N_g}\#\{g|g < -x\}$, where N_g is the number of observations. The empirical CCDF is plotted in **Figure 4**. In addition, we provide another estimate of the tail probability using the extreme value theory. The main idea of the estimation is to use the limit theorem that as the sample size increases, the tail probability of the

[12] We perform four normality tests (the KS, AD, CvM, SW tests). All four tests reject the null hypothesis that the GDP growth rate follows a Gaussian at the 99 percent confidence level.

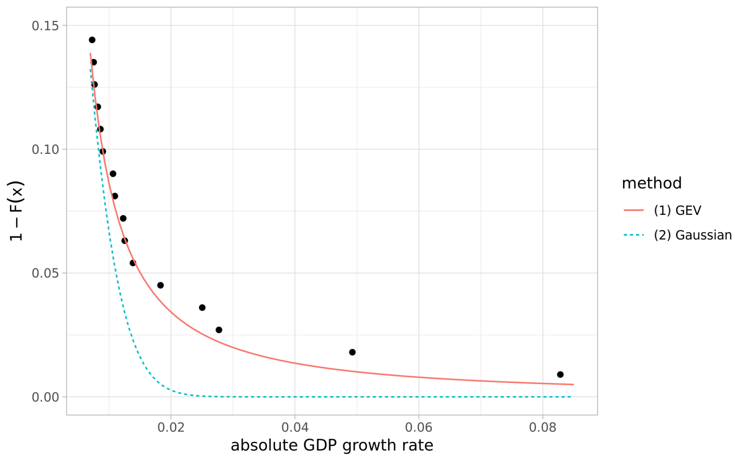


Figure 4: Estimates of the tail probability of the negative GDP growth rate. The horizontal axis represents the absolute value of the negative GDP growth rates, that is, $|g|$ with $g < 0$. The dots represent the empirical CCDF. GEV represents the estimate based on the GEV distribution. Gaussian represents the tail probability of the Gaussian distribution with parameters based on the sample mean and median absolute deviation given in **Table 1**.

properly normalized g converges to the generalized extreme value (GEV) distribution.[13] By estimating the parameter of the GEV distribution, we approximate the tail probability of the GDP growth rate. The result is given in **Figure 4**. For comparison, we also depict the Gaussian tail probability, whose parameters are based on the sample mean and median absolute deviation given in **Table 1**.

Figure 4 shows that both estimates (i.e., the empirical CCDF and GEV estimate) deviate from the Gaussian counterpart. Consistent with the QQ-plot in **Figure 3**, this result suggests that the tail probability of the GDP growth rate is larger than predicted by a Gaussian. That is, extremes of the GDP growth rates are more likely to occur than predicted by the corresponding Gaussian distribution. In Section 4, we study whether the granular view can explain this departure from a Gaussian in the tail region.

3.2 TSR data

We use Japanese firm-level data provided by Tokyo Shoko Research (TSR). This data consists of two types of information: individual firm information such as sales and profits and the identification of customer-supplier relationships. The latter enables us to identify which firms are the main suppliers/customers of a firm and construct an input-output network covering the Japanese economy. The sample period ranges from

[13] More precisely, let $X_{1,n} \geq X_{2,n} \geq \ldots \geq X_{j,n} \geq \ldots \geq X_{n,n}$ be the arrangement of the absolute values of the negative GDP growth rates in their decreasing order. We estimate the tail probability of X as follows:

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

where $\widehat{\xi}^{(H)}$ is Hill's estimate. In this estimation, we set k equal to 11. For detail, see Section 6.4.3 in Embrechts et al. (1997).

2012 to 2018.[14] In our analysis, we exclude firms with sales of less than 100 million yen and firms for which sales data is unavailable.[15] We also exclude firms in the sector of banking, insurance, and government. For customer-supplier linkages, we assume that a linkage between two firms exists in year t if it is reported within the last three years; that is, if a linkage between two firms is reported in 2016 and not updated after that, we assume that it still exists in 2018. By this sample selection, the numbers of firms and customer-supplier linkages in 2018 are reduced to $330,426$ and $1,874,514$, respectively.

3.2.1 Individual firm information

We provide a brief overview of the firm-level information relevant to our analysis. The variables of interest are annual sale revenue, labor share, and markup, whose summary statistics are given in **Table 2**.

year	name	count	mean	median	sd	max
2018	sale	330426	3696	490	51538	12201443
	labor share	330426	0.188	0.162	0.117	1
	markup	330426	1.029	1.009	0.083	3

Table 2: Summary statistics of variables for individual firms. The unit of sales is 1 million yen. Regarding the markup, we replace outliers (observations with a markup of more than 3) with 3. The summary statistics for other years are similar and omitted.

One of the stylized facts in the empirical literature on firm dynamics is Zipf's law: the distribution of sales has a Pareto tail with an exponent α close to 1. That is, for a sufficiently large x,

$$P(\text{Sale} > x) \sim Cx^{-\alpha}$$

where C is a constant. Thus, if Zipf's law holds, the tail of the distribution should be a straight line in the log-log scale. **Figure 5** confirms this property. The CCDF of sales in the log-log scale is close to a straight line, meaning that the tail of the distribution of sales is close to a Pareto tail. The tail exponent α is estimated by Hill's method. We find that the estimates $\widehat{\alpha}$ are close to 1 and stable over our sample periods; for example, the point estimate of α for 2018 is 1.17.[16] Consistent with the previous literature, firms' sales in the Japanese economy exhibit high heterogeneity and can be approximated by Zipf's law.

We mention how we calculate the markup and labor share from our data. The markup is estimated using the accounting approach (see, e.g., De Loecker et al. (2020)); that is, the markup is defined by the ratio of

[14] In our analysis, we mainly use the data for 2018. We use the data for other years to check the stationarity of the distribution properties of the variables of interest.

[15] The Economic Condition Survey in 2020 conducted by the Ministry of Internal Affairs and Communications shows that the total sales in Japan (i.e., the sum of sales of all firms in Japan) is equal to 1,499 trillion yen, and firms with sales larger than 100 million yen account for 98.1% of the total sales (i.e., for 1,471 trillion yen). Thus, it is highly unlikely that this criterion affects our analysis.

[16] In Hill's estimate, the top 1000 largest values of sales are used for the estimation.

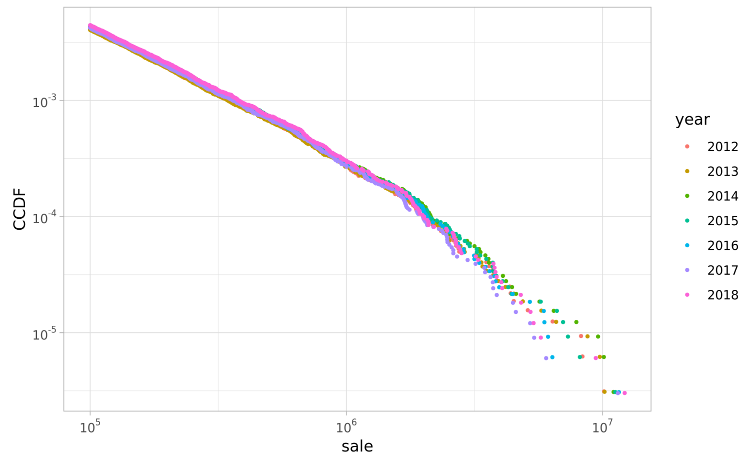


Figure 5: The CCDF of sales for 2012-2018 in the log-log plot. Only firms with sales larger than 100 billion yen are plotted.

sales to the total cost (= sales minus profits).[17] Since we assume that the markup is larger than 1 in Section 2.2, we replace negative profits with 0 so that the resulting markup satisfies this assumption. Turning to the labor share, labor costs are available only for 4% of all firms in our data. To impute the missing values for the others, we develop a machine learning model. The labor share is defined by the ratio of the predicted values of labor costs to sales.

3.2.2 Productivity shocks

To measure the size of micro-originated aggregate fluctuations, knowing the distribution of microeconomic productivity shocks is crucial. Since productivity shocks are not directly observable from the data, we use the growth rates of annual sales revenue and labor productivity as a proxy for productivity shocks.[18] Here, labor productivity is defined by annual sales revenue divided by the number of employees. We consider only large firms when estimating productivity shocks because only shocks to large firms are relevant for

[17] An alternative approach is the production function approach proposed by De Loecker and Warzynski (2012). In this approach, the markup is obtained as the ratio of the output elasticity of a variable input to the input's cost share in sales revenue. Once the output elasticity of a variable input is properly estimated, the markup obtained from the production function approach is equal to the markup of price over marginal cost. However, Bond et al. (2021) point out that if the revenue elasticity instead of output elasticity is employed, the markup obtained from the production approach is valid only when the true markup is 1. That is, the production function approach can recover the true markup only when the price is equal to the marginal cost. This is because the revenue elasticity is equivalent to the output elasticity only when a firm cannot affect the output price by changing the quantity. Since the ratio of the revenue elasticity to the cost share does not contain any information about the true markup except for this extreme case, we employ the accounting approach in our analysis.

[18] Here, the frequency considered is annual, though the quarterly frequency is used for the GDP growth rates in Section 3.1. For this difference, see Footnote 21.

micro-originated aggregate fluctuations, and the labor productivity defined above is not reliable when the number of employees is small. We restrict our samples to firms with sales and employees of more than 10 billion yen and 300 employees. The summary statistics of the growth rates are given in **Table 3**.

year	var	count	mean	s.d.	mad
2018	growth rate of sale	5783	0.039	0.112	0.061
	growth rate of labor productivity	5783	0.021	0.135	0.069

Table 3: Summary statistics of the growth rates of sales and labor productivity. Summary statistics for other years are similar and omitted.

Next, consider the distribution shape of the growth rates of sales and labor productivity. **Figure 6** shows the density estimates of the two growth rates. Both panels show that the distribution of growth rates deviates from a Gaussian and is peaked around its center. This graphical inspection suggests that the distribution of growth rates is close to a Laplace distribution.[19] Based on this observation, we assume that Assumption 2.1 holds (i.e., the microeconomic shocks are drawn from a Laplace distribution). We estimate the scale parameter b of the Laplace distribution by the maximum likelihood method. For 2018 case, the parameter estimates for the growth rates of sales and labor productivity are $\widehat{b} = 0.0674$ and $\widehat{b} = 0.0807$, respectively. These estimates are used in Section 4.

3.2.3 Network information

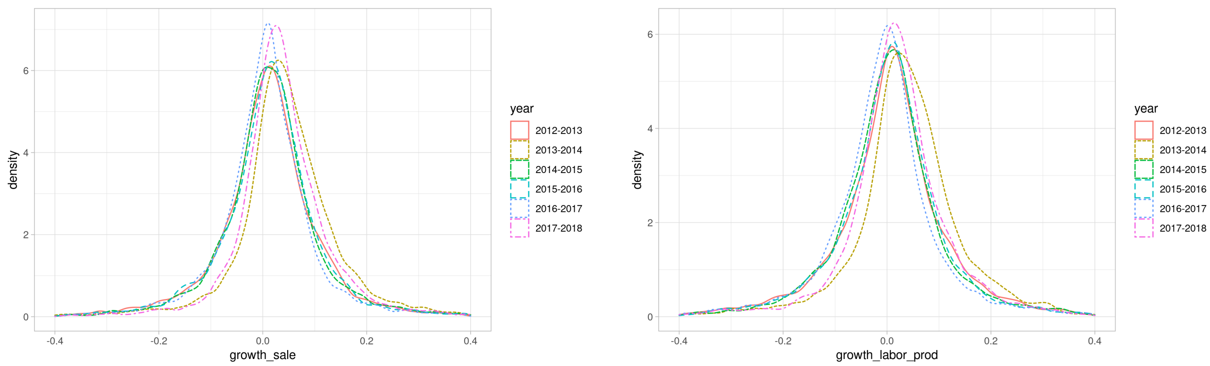
We analyze the basic properties of the input-output network generated by customer/supplier transaction linkages. To get the first glimpse of its network structure, let us consider the network without link weights. **Table 4** gives the summary statistics of the number of links for each firm, that is, the number of suppliers and customers. **Figure 7** shows their CCDFs in the log-log scale. Similar to the firm size distribution, the distribution tail of the number of links is close to a straight line, that is, a Pareto tail. The number of links exhibits high heterogeneity across firms, and in particular, there exist firms with many suppliers and customers.

Next, consider the connectivity of the input-output network. We say that firms A and B belong to a (weakly) connected component if a path (or a sequence of links) exists that connects firms A and B. That

[19] To confirm the deviation from a Gaussian, we use a three-parameter distribution called the generalized Gaussian distribution, whose density is given by

$$d(x;\mu,b,\beta) = \frac{\gamma}{2b\Gamma(1/\gamma)} \exp\left(-\left(\frac{|x-\mu|}{b}\right)^{\gamma}\right)$$

where $\gamma > 0$. Note that this distribution nests the two distributions: when $\gamma = 2$ ($\gamma = 1$), it corresponds to the Gaussian (Laplace) distribution. We estimate parameter γ by the maximum likelihood method. For the growth rates of sales, we find that $\widehat{\gamma} = 0.755(0.016)$, where the standard error is in parenthesis. Although the estimate is significantly smaller than 1 (i.e., the Laplace case), this result shows that the distribution of the growth rate of sales is closer to the Laplace distribution than to a Gaussian.



(a) sale **(b)** labor productivity

Figure 6: The density estimates of the growth rates of sales and labor productivity. The maximum likelihood estimates of the scale parameter of the Laplace distribution are $\widehat{b} = 0.0674$ for (a) and $\widehat{b} = 0.0807$ for (b).

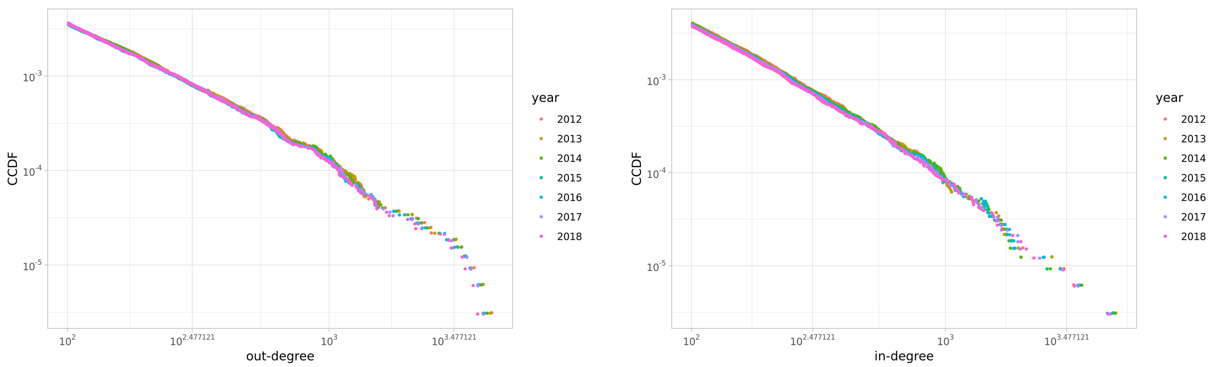
year	var	count	mean	meadian	s.d.	max
2018	out-degree	330426	5.673	3.000	26.646	3692
	in-degree	330426	5.673	2.000	23.825	4363
	in-degree strength	330426	0.647	0.112	4.225	877

Table 4: Summary statistics of the out-degree (the number of suppliers) and in-degree (the number of customers) for each firm. The summary statistics of in-degree strength are also included.

is, two firms are (weakly) connected if they are directly or indirectly connected by links. We find that the input-output network has a unique giant connected component that contains 99.6% of all firms for 2018 data. In other words, the overwhelming majority of firms in the Japanese economy are directly or indirectly connected via customer-supplier relationships.

How connected is the connected component? To measure the degree of connectivity, we calculate the shortest path length for pairs in the connected component, that is, the minimum number of links that connect two firms. The left panel of **Figure 8** shows the histogram of the shortest path length for the pairs of large firms.[20] It shows that in most of the pairs, the path length is 3 or 4, and indeed, the average path length is 3.57. Thus, firms are not only connected either directly or indirectly but the path length of connections tends to be short. Furthermore, we calculate the closeness centrality for each firm, which is defined as the reciprocal of the average path length between the firm and all other firms in the connected component. Intuitively, firms with high closeness centrality lie at the center of the network. The right panel of **Figure 8** depicts the scatter plot of the closeness centrality and the sum of in- and out-degrees. They are positively correlated, and the correlation coefficient is 0.66. As expected, the firms with many customers and suppliers (typically large firms) play the role of a hub in the network, and as a result, the giant network turns out to be well-connected.

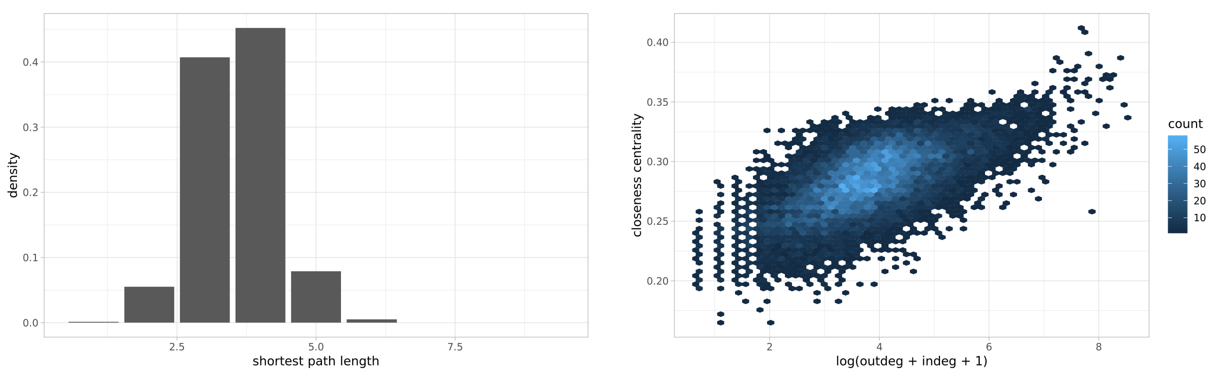
[20]Due to computational infeasibility, we consider the subgraph induced by restricting samples to firms with sales larger than 10 billion yen ($14,212$ firms) to compute the shortest path length and closeness centrality.



(a) Out-degree (the number of suppliers)

(b) In-degree (the number of customers)

Figure 7: The CCDF of the out-degree (the number of suppliers) and in-degree (the number of customers) in the log-log plot.



(a) Histogram of shortest path length

(b) Scatter plot of closeness centrality and degree

Figure 8: Structure of the input-output network. In (b), the horizontal axis is the logarithm of the sum of a firm's in-degree and out-degree plus 1.

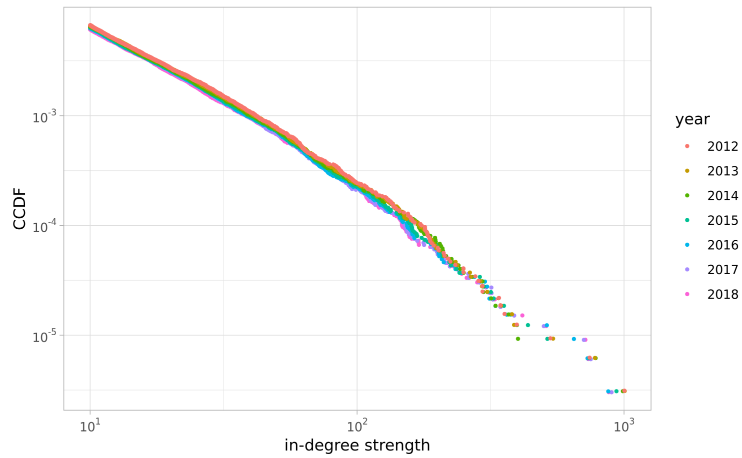


Figure 9: The CCDF of in-degree strength in the log-log scale.

Finally, let us consider the network with link weights, which corresponds to $\widetilde{\Omega}$ in Section 2.2. In TSR data, the link weights are available only for a limited number of pairs of firms. Similar to the case of the labor share, we use a machine learning model to impute these missing link weights (see the Appendix) and construct $\widetilde{\Omega}$. One of the important statistics of $\widetilde{\Omega}$ in the context of the micro-originated aggregate fluctuations is in-degree strength, which is defined by $\widetilde{d}_j := \sum_i \widetilde{\Omega}_{ij}$. Intuitively, this represents the importance of good j used in the production of other firms in the economy. Its summary statistics of in-degree strength are given in **Table 4**. **Figure 9** depicts the CCDF of in-degree strength in the log-log plot, showing that the distribution of in-degree strength has a Pareto tail. Similar to the firm size and the in- and out-degree distributions, the distribution of in-degree strength shows high heterogeneity across firms. This figure also shows that this distribution shape is stable over our sample period.

To summarize, the empirical properties discussed above appear to be in favor of the granular view: The heterogeneity of firm sizes is high, and Zipf's law holds. The distribution of productivity shocks deviates from a Gaussian and is close to a Laplace distribution. The input-output network is well connected with a short path length, and there exist hub firms with many links. These properties are why the granular view has attracted attention in recent years. However, in the next section, we show that this observed heterogeneity is not large enough to generate a large deviation of aggregate output. In fact, despite this heterogeneity, we show that the averaging effect (i.e., the cancellation of microeconomic shocks) is still dominant.

4 Empirical Results

This section provides our main empirical results. We apply the probabilistic methods in Section 2.1 to get implications about the aggregate variance (Section 4.1), the tail probability (Section 4.2), and the closeness to a Gaussian (Section 4.3). Throughout this section, we assume the Assumption 2.1 holds with the standard deviation $\sigma_\epsilon = 5.71\%$ (i.e., with $b = 0.0404$).[21] Given this assumption, we test whether the empirical granularity in Japan is high enough to explain the observed properties of the GDP growth rates.

4.1 Aggregate Variance

Let us consider the size of the aggregate variance driven by microeconomic shocks (i.e., $\sigma_\epsilon^2 \|w\|_2^2$). First, we consider an efficient economy, in which the weights are given by the Domar weights λ (see Eq.(7)). We construct empirical Domar weights for each firm from its sales and Japan's GDP in 2018. The summary statistics and CCDF of the Domar weights are given in **Table 5** and the left panel of **Figure 10**, respectively.

Recall that the tail exponent of the distribution of sales, which is equivalent to that of the Domar weights, is estimated to be $\widehat{\alpha} = 1.17$. Since the estimate is within the range of $(0, 2)$, the aggregate variance decays at a slower rate than $N^{-1/2}$, as suggested by Gabaix (2011). Indeed, **Table 5** shows that $\|\lambda\|_2$ is 0.053 in 2018, and therefore, the standard deviation of aggregate output driven by microeconomic shocks is

$$\sigma_\epsilon \|\lambda\| = 5.71\% \times 0.053 = 0.303\%.$$

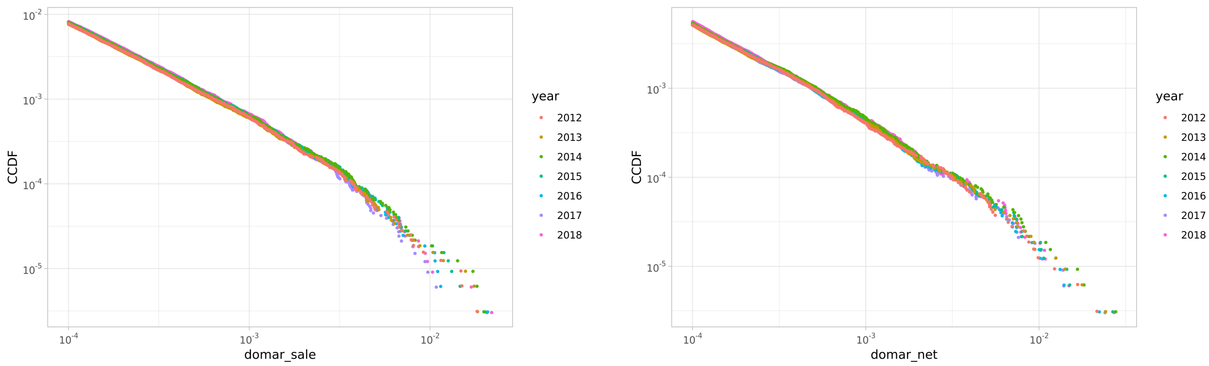
Compared to the empirical counterpart, that is, the standard deviation of the GDP growth rates in **Table 1**, the contribution of microeconomic shocks to the aggregate variance is not negligible. Moreover, the estimate $\widehat{\alpha} = 1.17$ implies that r_{max} does not converge to 0 as $N \to \infty$ by Proposition 2.2. Indeed, **Table 5** shows that $r_{\text{max}} = 0.169$, that is, 16.9% of the micro-originated aggregate variance comes from shocks to the largest firm only. This explains why the CLT does not hold; that is, even when N is large, shocks to the largest firm cannot be canceled out by shocks to other firms, and thus, the decay rate of the micro-originated aggregate variance becomes slower. Therefore, for this case, the empirical results support the granular view.

year	name	count	mean	sd	$\|\cdot\|_\infty$	$\|\cdot\|_2$	$r_{\max}$
2018	λ	330426	6.648e-06	9.270e-05	0.022	0.053	0.169
	λ_{net}	330426	6.112e-06	9.258e-05	0.027	0.053	0.259
	b	330426	2.652e-06	4.297e-05	0.009	0.025	0.135

Table 5: Summary statistics of the weights in the model of an efficient economy. In the calculation, GDP is set to 556 trillion yen. $r_{\max}$ is given by the ratio of the two squared norms, that is, $r_{\max} = \|\cdot\|_\infty^2 / \|\cdot\|_2^2$

How much of the observed heterogeneity of the Domar weights λ can be explained by network het-

[21] Since the frequency of the empirical GDP growth rates given in Section 3.1 is quarterly, we need to consider the quarterly growth rate of microeconomic shocks. In Section 4, assuming that the quarterly growth rate of microeconomic shocks also follows a Laplace distribution, we adjust its parameter so that the observed variance of the annual growth rates is equal to four times the variance of the quarterly growth rates. Thus, since the estimate $\widehat{b} = 0.0807$ for the annual growth rates of labor productivity in Section 3.2.2, we set b equal to $0.0807/2 = 0.0404$ for the quarterly growth rates. Since $\sigma_\epsilon = \sqrt{2}b$ under Assumption 2.1, we have $\sigma_\epsilon = \sqrt{2} \times 0.0404 = 5.71\%$. In the Appendix, to check the robustness of our result, we consider the distribution of the quarterly growth rates of microeconomic shocks whose annual growth rates is a Laplace distribution.



(a) Domar weights **(b)** Network heterogeneity

Figure 10: The CDF of the Domar weights and network heterogeneity in the model of an efficient economy. Hill's estimate of the tail exponent is $\widehat{\alpha} = 1.17$ in Panel (a) and $\widehat{\alpha} = 1.08$ in Panel (b) for 2018.

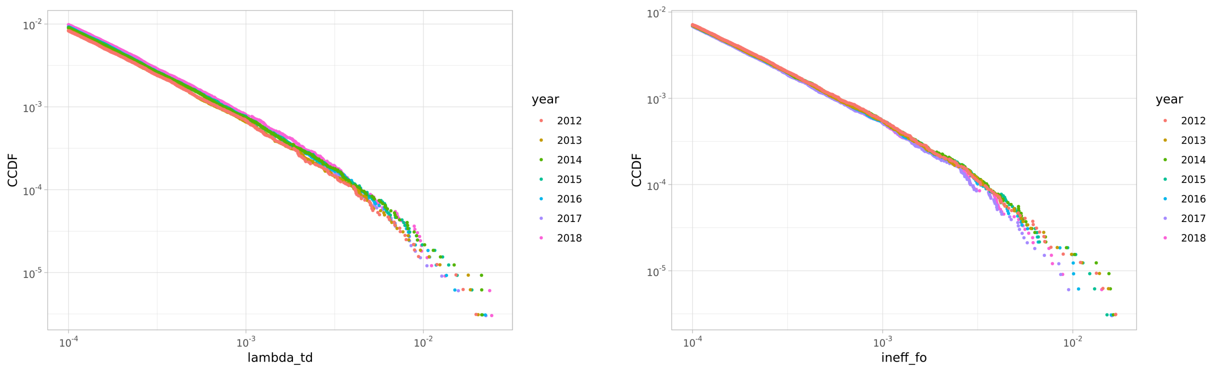
erogeneity λ_{net} ? We calculate the network heterogeneity using Eq.(9). Their summary statistics are given in **Table 5**. The right panel of **Figure 10** depicts the CCDF of the network heterogeneity, showing that the distribution tail of the network heterogeneity is close to a Pareto tail, similar to that of the Domar weights. Hill's estimate of the tail exponent for the network heterogeneity is $\widehat{\alpha} = 1.08$, which is close to the estimate $\widehat{\alpha} = 1.17$ for the Domar weights. Note that the two quantities (i.e., the Domar weights and network heterogeneity) are calculated from two different sources: sales revenue and input-output linkages. Thus, there is no a priori reason to assume that the two distribution tails have a similar tail exponent. The similarity of these two tails suggests that most of the heterogeneity of the Domar weights comes from the network heterogeneity.

To confirm this point further, we compare the norms of three quantities: the Domar weights λ, network heterogeneity λ_{net}, and primitive heterogeneity β.[22] **Table 5** shows that compared to the primitive heterogeneity, both ℓ^2- and ℓ^∞-norms of the network heterogeneity are close to those of the Domar weights. This suggests that the network heterogeneity is the main source of the heterogeneity of the Domar weights. In addition, **Table 5** shows that $r_{\max}$ of the network heterogeneity is 0.259. This means that even when there is no primitive heterogeneity, the single firm with the highest network heterogeneity accounts for a significant part of the aggregate variance driven by microeconomic shocks. Thus, we conclude that the network heterogeneity is the main reason why the CLT does not hold.

[22] To calculate the primitive heterogeneity (i.e., the final expenditure share b), we use the definition of the Domar weights and the accounting identity. That is, given the input-output network (Ω) and Domar weights calculated by firms' sales and GDP, we construct the final expenditure share as follows:

$$b' = \lambda'(I - \Omega)$$

Although the final expenditure shares must be non-negative for all goods, some b take negative values in our calculation. This inconsistency is partly because the missing values of the transaction amount are imputed by a machine learning method, and therefore, there exists an error from the true value. In our empirical analysis, we set b_i equal to 0 when it is negative.



(a) $\widetilde{\lambda}_i$ **(b)** $\lambda_i \frac{\Psi_{iL}}{\Lambda_L}$

Figure 11: The CCDF of the weights in the model of an inefficient economy. Hill's estimate of the tail exponent is $\widehat{\alpha} = 1.15$ in Panel (a) and $\widehat{\alpha} = 1.19$ in Panel (b) for 2018.

Next, let us consider an inefficient economy, in which $\widetilde{\Omega}$ is not equal to Ω. We first construct $\widetilde{\Omega}$ from our empirical data on input-output linkages and then construct Ω by using the relation $\Omega = \text{diag}(\mu)^{-1}\widetilde{\Omega}$. For the calculation of $\widetilde{\lambda}$, we use the following equation:

$$\widetilde{\lambda}' = \beta(I - \widetilde{\Omega})^{-1} = \lambda'(I - \Omega)(I - \widetilde{\Omega})^{-1}$$

where λ is calculated using firms' sales and GDP as in the case of an efficient economy.

Given this setup, let us consider the Cobb-Douglas case $\theta_j = 1$, in which the weight for firm i is given by $\widetilde{\lambda}_i$. Their summary statistics and CCDF are given in **Table 6** and the left panel of **Figure 11**, respectively. Note that $\|\widetilde{\lambda}\|_2 = 0.065$, meaning that for the Cobb-Douglas case, inefficiencies (i.e., markup) increase the aggregate variance induced by microeconomic shocks. Other than this point, the heterogeneity of $\widetilde{\lambda}$ is similar to that of the Domar weights in the model of an efficient economy. The CCDF in **Figure 11** shows that the distribution of $\widetilde{\lambda}$ has a Pareto tail with exponent $\widehat{\alpha} = 1.15$. Consistent with the heavy-tailedness of the distribution, we find that $r_{\max} = 0.138$. That is, the firm with the largest $\widetilde{\lambda}$ alone accounts for 13.8% of the aggregate variance driven by microeconomic shocks. As in the case of an efficient economy, because of the fact that the largest firm has a disproportionate impact on aggregate output, the CLT does not hold, and the effect of microeconomic shocks remains substantial even at the aggregate level.

year	var	count	mean	s.d.	$\|\cdot\|_\infty$	$\|\cdot\|_2$	$r_{\max}$
2018	$\widetilde{\lambda}_i$	330426	8.018e-06	1.133e-04	0.024	0.065	0.138
2018	$\lambda_i \frac{\Psi_{iL}}{\Lambda_L}$	330426	5.558e-06	7.495e-05	0.016	0.043	0.137

Table 6: Summary statistics of the weights in the model of an inefficient economy. $r_{\max}$ is given by the ratio of the two squared norms, that is, $r_{\max} = \|\cdot\|_\infty^2 / \|\cdot\|_2^2$

We conduct the same exercise as above for the Leontief case $\theta_i = 0$ for all i, in which the weight for firm i is given by $\lambda_i \frac{\Psi_{iL}}{\Lambda_L}$. Their summary statistics and CCDF are given in **Table 6** and the right panel of **Figure 11**, respectively. Compared to the case of the Cobb-Douglas case, the aggregate variance driven by microeconomic shocks is lower (i.e., $\|\lambda_i \frac{\Psi_{iL}}{\Lambda_L}\|_2 = 0.043$).[23] Except for the difference in terms of their norms, we find that the heterogeneity of $\lambda_i \frac{\Psi_{iL}}{\Lambda_L}$ is similar to that of $\widetilde{\lambda}$ for the Cobb-Douglas case. As shown in **Figure 11**, the distribution tail is close to a straight line (i.e., a Pareto tail), and the tail exponent is $\widehat{\alpha} = 1.19$. The ratio $r_{\max}$ is 0.137, suggesting that the CLT does not hold. Thus, parameter θ does not affect the overall picture of the heterogeneity of the weights.

In all of the cases considered, microeconomic shocks contribute significantly to the aggregate variance. Because of the high heterogeneity of the weights, the CLT does not hold, and the largest firm accounts for a non-negligible part of the economy in terms of the aggregate variance. Therefore, as far as the aggregate variance is concerned, the granular view holds true.

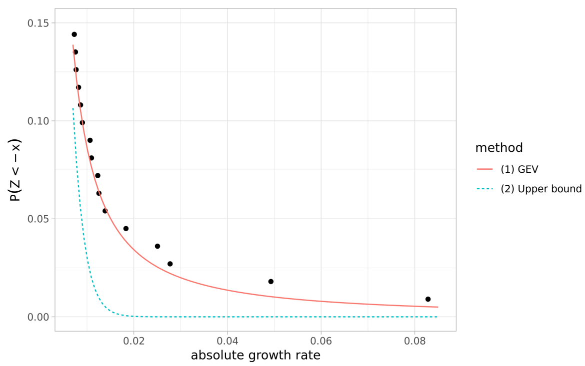
4.2 Tail probability

Let us consider the tail probability of aggregate output driven by microeconomic shocks. Recall that the upper bound of the tail probability (i.e., Eq.(3)) is determined by the two norms of the weights: $\|w\|_2$ and $\|w\|_\infty$. Thus, by plugging the two norms of the empirical weights given in **Table 5** and **Table 6** into Eq.(3), we compare the upper bound and the GEV estimate of the empirical GDP growth rates (see Section 3.1).

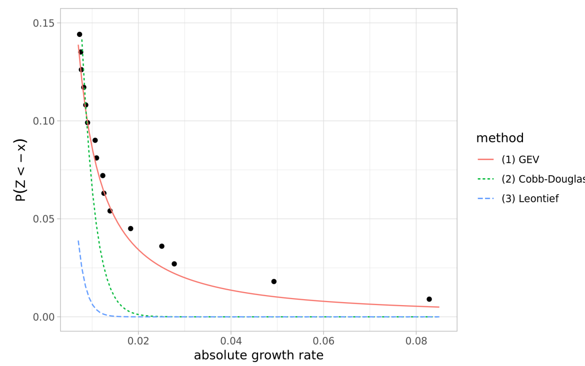
First, consider an efficient economy, in which the weights are given by λ. The left panel of **Figure 12** depicts the upper bound of the tail probability given by Eq.(3) with $\|\lambda\|_2$ and $\|\lambda\|_\infty$ and compares it with the empirical CCDF and GEV estimate of the GDP growth rate. It shows that the upper bound is substantially lower than its empirical counterparts; for example, while the probability that the GDP growth rate is less than -0.02 is estimated to be 3.43% by the GEV estimate, the upper bound of the tail probability yields that $P(Z < -0.02)$ is less than 0.026%. Since the upper bound of the tail probability for a large x is determined by $\|\lambda\|_\infty$, this result suggests that the empirical value of $\|\lambda\|_\infty$ (i.e., the size of the largest firm in terms of sales) is too small to generate a large deviation of aggregate output. In contrast to the aggregate variance discussed in Section 4.1, the microeconomic shocks contribute almost nothing to the tail probability of aggregate output.

Next, we perform the same exercise for an inefficient economy with $\theta = 0, 1$ and find that the results are similar to the case of an efficient economy. The right panel of **Figure 12** depicts the upper bounds of the tail probability for the two cases, showing that they are substantially lower than their empirical counterparts; for example, the upper bounds yield that $P(Z < -0.02) < 0.12\%$ for $\theta = 1$ and $P(Z < -0.02) < 0.01\%$ for $\theta = 0$. Thus, even when inefficiencies are considered, the contribution of microeconomic shocks to the

[23] Intuitively, this is because, given the Leontief technology (i.e., goods are not substitutable), firms cannot increase the amount of an intermediate good produced by another firm that experiences a positive shock. That is, even when a firm experiences a positive shock and decreases its price, this good is not used in the production of other firms as much as in the Cobb-Douglas case, and therefore, the impact of a positive shock on aggregate output is limited.



(a) Efficient economy



(b) Inefficient economy with $\theta = 0, 1$

Figure 12: The upper bound of the tail probability of aggregate output. For (b), Cobb-Douglas and Leontief represent the upper bounds of the tail probability (i.e., Eq.(3)) with $\theta = 1$ and $\theta = 0$, respectively. For comparison, the empirical CCDF and GEV estimate of the GDP growth rates are plotted.

tail probability of aggregate output is negligible.

Why do microeconomic shocks not contribute to the tail probability of aggregate output despite its non-negligible contribution to the aggregate variance? A clue is given by the difference in the convergence rate (see Section 2.1.1 and Section 2.1.2): when the homogeneous economy with $w_i = 1/N$ for all i is considered, the aggregate variance decays at the rate of $1/N$, but the tail probability of aggregate output decays at the rate of $\exp(-cN)$ with some constant c. That is, our result suggests that even when the weights are heterogeneous, the convergence rate of the tail probability is faster than that of the aggregate variance, and given the empirical granularity (i.e., an empirical N is given), the tail probability induced by microeconomic shocks is sufficiently close to 0. This point contrasts with Acemoglu et al. (2017), who show the slow convergence rate of the tail probability when the heterogeneity of the weights is high. The reason for the difference is that while Acemoglu et al. (2017) consider the convergence rate of the tail probability as $N \to \infty$, our analysis considers (the upper bound of) the tail probability with fixed N. It should be noted that our result does not contradict the finding by Acemoglu et al. (2017). Rather, our result means that the convergence rate of the tail probability becomes slower due to the high heterogeneity (an asymptotic result), *and* the tail probability is negligible, given the empirical granularity in Japan (a non-asymptotic result). This is why the distinction between asymptotic and non-asymptotic results is crucial for the assessment of the granular view.

Given these results, one might raise the following question: what is the shape of the distribution with the non-negligible variance and negligible tail probability? To answer this question, we next analyze the closeness of the distribution of aggregate output to a Gaussian using the methods in Section 2.1.3. The result in the following section provides another explanation of why microeconomic shocks cannot generate a large deviation of aggregate output.

4.3 Distribution Approximation

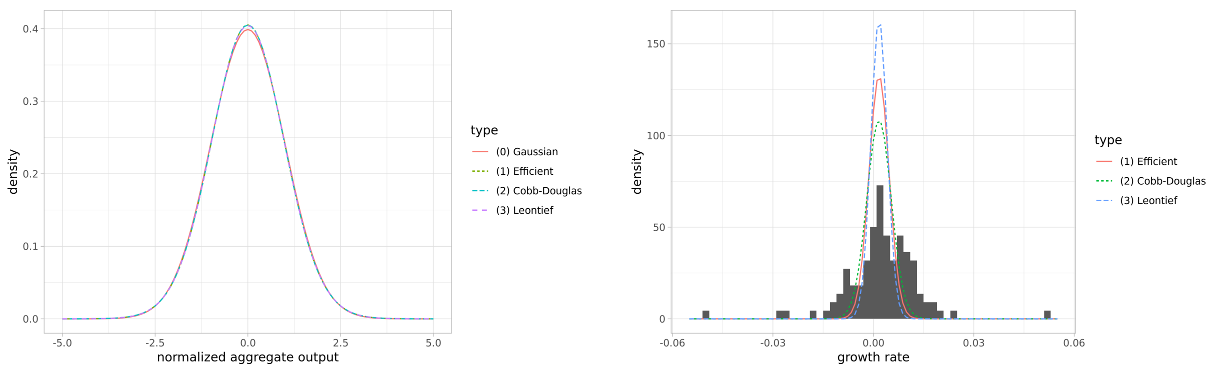
Let us consider the normalized aggregate output Z^* in Eq.(4) and focus on its cumulants. Eq.(5) shows that the jth-order cumulant of the normalized aggregate output $\kappa_j(Z^*)$ is reduced by a factor of $\sum_i w_i^{*j}$. That is, the multiplier factor $\sum_i w_i^{*j}$ summarizes the averaging effect through the summation in Eq.(4). In particular, since $\kappa_4(\epsilon^*) = 3$ under Assumption 2.1, $\kappa_4(Z^*)$ (i.e., the excess kurtosis) is reduced from 3 by a factor of $\sum_i w_i^{*4}$. Using the empirical values of the weights, we calculate the multiplier factor as follows:

$$\sum_i w_i^{*4} = \begin{cases} 0.047 & \text{for an efficient economy} \\ 0.045 & \text{for an inefficient economy with } \theta = 1 \\ 0.039 & \text{for an inefficient economy with } \theta = 0 \end{cases}$$

For example, in the case of an efficient economy, the multiplier factor for the excess kurtosis is equal to 0.047. The excess kurtosis of the normalized aggregate output decreases from 3 to $3 \times 0.047 = 0.141$ by the averaging effect. Although the distribution is not determined solely by the excess kurtosis, this value of the excess kurtosis suggests that the distribution of the normalized aggregate output is very close to a Gaussian; for example, a logistic distribution, which is known to resemble a Gaussian distribution and shown in **Figure 2**, has an excess kurtosis of 1.2. Intuitively, through the summation Eq.(4), the distribution of the normalized aggregate output becomes closer to the standard Gaussian than the logistic distribution. The results are similar for the cases of an inefficient economy. Thus, this substantial reduction in the excess kurtosis suggests that the averaging effect still works, even when the CLT does not hold.

To confirm this point further, we approximate the probability density of the normalized aggregate output using the Edgeworth series. The result is given in the left panel of **Figure 13**, in which we also plot the probability density of the standard Gaussian distribution for comparison. This figure shows that the approximated probability density of the normalized aggregate output is very close to that of the standard Gaussian; only in the center region, a slight deviation from the Gaussian is observed. This is consistent with the small value of the excess kurtosis found above. Thus, given the empirical granularity in Japan, the averaging effect of microeconomic shocks still works and is large enough to make the resultant distribution of aggregate output sufficiently close to a Gaussian. Note that this result holds for both cases of the efficient and inefficient economies, and the approximated probability densities lie on the same curve. This means that when aggregate output is normalized, there is no difference between the models for efficient and inefficient economies. In other words, the difference between these models affects only the aggregate variance but not the distribution shape, that is, the closeness to a Gaussian distribution. This is consistent with the finding in Section 4.1, which shows that the ℓ^2-norm differs across models, but other properties such as $r_{\max}$ and the tail exponent of the weights are similar to each other. Thus, these results confirm the robustness of our finding that the distribution of aggregate output is close to a Gaussian.

It is worth mentioning that our finding does not contradict the fact that the CLT does not hold. As discussed in Section 4.1, the distribution of aggregate output does not converge to the Gaussian distribution



(a) For Z^* **(b)** For Z

Figure 13: Distribution approximation by Edgeworth series. Panel (a) shows the approximation for the normalized aggregate output Z^*. For comparison, the standard Gaussian distribution is plotted. Panel (b) shows the approximation for the normalized aggregate output Z^*. For comparison, the histogram of the GDP growth rate is plotted, which is the same as in **Figure 3**.

as $N \to \infty$. The point is that this is an asymptotic result as $N \to \infty$. The asymptotic result of the non-convergence to a Gaussian does not exclude the possibility that it is sufficiently close to a Gaussian. Our finding shows that with an empirical value of N (i.e., the empirical weights), the latter possibility is important to describe the distribution of aggregate fluctuations driven by microeconomic shocks.

Finally, we approximate the probability density of (not normalized) aggregate output using the Edgeworth series and compare it with the histogram of the GDP growth rate. This is shown in the right panel of **Figure 13**. This figure shows that the approximated probability density cannot explain the observed extremes of the GDP growth rates, for example, the growth rates less than -1%. This is consistent with the finding in Section 3.1; that is, the distribution of the GDP growth rate deviates from a Gaussian in the tail region, while the distribution of aggregate output induced by microeconomic shocks is very close to a Gaussian. Thus, we conclude that microeconomic productivity shocks cannot explain the observed extremes of the GDP growth rates.

The closeness to a Gaussian explains why we cannot find a significant tail probability in Section 4.2, even though we find a non-negligible aggregate variance in Section 4.1. Because of the high heterogeneity across firms, the aggregate variance decays slowly and remains substantial. However, the averaging effect is still dominant, and microeconomic shocks to a substantial extent cancel each other out. For this reason, the resultant distribution of aggregate output induced by microeconomic shocks turns out to be sufficiently close to a Gaussian. In addition, compared to the aggregate variance, the tail probability of aggregate output decays more rapidly, consistent with the light-tailedness of the Gaussian distribution. Therefore, microeconomic shocks do not generate a large deviation but only small fluctuations around its mean, showing the limitation of the granular view.

5 Conclusion

What drives aggregate fluctuations? This old but fundamental topic has attracted renewed interest in the last decade, and the importance of microeconomic shocks as another source of aggregate fluctuations has been widely accepted. However, because of the accessibility of firm-level data, empirical works that quantify the size of the micro-originated aggregate fluctuations are sparse. In particular, little is known about the distribution properties of aggregate output, given an empirical input-output network. Our paper tackles this issue by using probabilistic methods and firm-level input-output data for Japan.

Based on Acemoglu et al. (2012) and Baqaee and Farhi (2020b), we calculate the impact of microeconomic shocks on aggregate output for each firm. We find that these weights are highly heterogeneous and that shocks to the single firm with the largest weight account for a significant part of the aggregate variance. This disproportional impact coming from the firm with the largest weight is the reason why the aggregate variance does not vanish, and the CLT does not hold. In contrast, we also find that microeconomic productivity shocks contribute almost nothing to the tail probability of aggregate output. The empirical granularity (especially the largest weight) is too small to cause a large deviation of aggregate output. This result does not contradict the non-negligible aggregate variance driven by microeconomic shocks and can be explained by analyzing the distribution shape of aggregate output. We find that given the empirical granularity in Japan, the distribution of aggregate output induced by microeconomic shocks is very close to a Gaussian. That is, the averaging effect is still dominant and most microeconomic shocks cancel each other out, leading to the almost Gaussian shape. Since the empirical GDP growth rate deviates from a Gaussian in its tail, microeconomic shocks cannot explain the observed extremes of the GDP growth rate.

Although our analysis is based on the general models widely accepted in the literature, several limitations should be highlighted. First, in our analysis, only the first-order terms (i.e., Hulten's theorem) are considered. Regarding this point, Dew-Becker (2022) and Baqaee and Farhi (2019) go beyond this first-order approximation and analyze the higher-order terms. Another limitation is that the extensive margins such as entry/exit of firms and changes in network linkages in response to shocks are not considered. Some recent studies (e.g., Acemoglu and Tahbaz-Salehi (2020); Baqaee (2018)) introduce these extensive margins and propose a new structural model. Our finding suggests that these extensions are necessary to explain the important features of aggregate fluctuations.

6 Appendix

In Appendix 6.1, we provide the proof of Proposition 2.3. In Appendix 6.2, we explain how to impute the missing data of labor costs in sales and the transaction weight of a supplier/customer for each firm. In Appendix 6.3, we estimate the TFP by using the method proposed by Ackerberg et al. (2015), and the growth rates of the TFP follow a Laplace distribution similar to the one in Section 3.2.2. In Appendix 6.4, we check the robustness of our finding by considering the quarterly microeconomic shocks whose annual distribution follows a Laplace distribution.

6.1 Proof

Here, we provide the proof of Proposition 2.3. We follow the literature on concentration inequality (see Boucheron et al. (2012)) and prove the upper bound of the tail probability from that of the moment generating function.[24]

Definition 6.1. A centered random variable X is called sub-Gamma on the right tail with variance factor v and scale parameter c (denoted by $X \in \text{sub}\Gamma(v,c)$) if

$$\log Ee^{tX} \leq \frac{vt^2}{2(1-ct)}$$

for $0 < t < 1/c$.

Theorem 6.1 (See Chapter 2.4 and Corollary 2.11 in Boucheron et al. (2012))**.** *Suppose that $X \in sub\Gamma(v,c)$. Then, for $x > 0$,*

$$P(X > x) \leq \exp\left(-\frac{x^2}{2(v+cx)}\right)$$

Proof of Proposition 2.3. First, we show that ϵ_i is sub-Gamma. Note that the Laplace assumption of ϵ with parameter b means that $|\epsilon|$ follows an exponential distribution with parameter b. Thus, the kth moments of $|\epsilon|$ is given by

$$E|\epsilon|^k = b^k k! = \frac{1}{2} v_\epsilon c_\epsilon^{k-2} k!$$

where $v_\epsilon := 2b^2$ and $c_\epsilon := b$. The expansion of the moment generating function yields that for $0 < t < 1/c$,

$$Ee^{t\epsilon} \leq Ee^{t|\epsilon|} = 1 + \frac{v_\epsilon}{2}\sum_{k=2}^{\infty} t^k c_\epsilon^{k-2} = 1 + \frac{v_\epsilon t^2}{2(1-c_\epsilon t)} \leq \exp\left(\frac{v_\epsilon t^2}{2(1-c_\epsilon t)}\right)$$

Thus, ϵ is sub-Gamma with variance factor v_ϵ and scale parameter c_ϵ.

Next, let us consider the weighted sum $\sum_i w_i \epsilon_i$. Note that each component $w_i \epsilon_i$ is sub-Gamma with variance factor $v_i := 2w_i^2 b^2$ and common scale parameter $c = b\|w\|_\infty$. Indeed,

$$Ee^{tw_i\epsilon_i} \leq \exp\left(\frac{2b^2 w_i^2 t^2}{2(1-bw_i t)}\right) \leq \exp\left(\frac{v_i t^2}{2(1-ct)}\right)$$

Observe the fact that the sum of sub-Gamma random variables with variance factor v_i and the common scale factor c is also sub-Gamma with $v := \sum_i v_i = 2b^2\|w\|_2^2$ and c. Indeed, by the independence, we have

$$Ee^{t\sum_i w_i\epsilon_i} = \prod_i e^{tw_i\epsilon_i} \leq \exp\left(\frac{vt^2}{2(1-ct)}\right)$$

Finally, by using the concentration inequality for a sub-Gamma random variable above and the symmetry, the desired result follows. □

[24] The proof strategy is essentially the same as the proof of Proposition 5 in Acemoglu et al. (2017), but the coefficients in the upper bound of the tail probability are different. Since the upper bound is used for our empirical analysis and the coefficients are crucial for tight bounds, we show how the coefficients are determined.

year	name	count	mean	sd	mad
2018	growth rate of TFP	3115	0.013	0.153	0.095

Table 7: Summary statistics of the growth rates of the estimated TFP in 2018. The shape parameter of the GGD is estimated to be $0.832(0.025)$, where the standard error is in parenthesis.

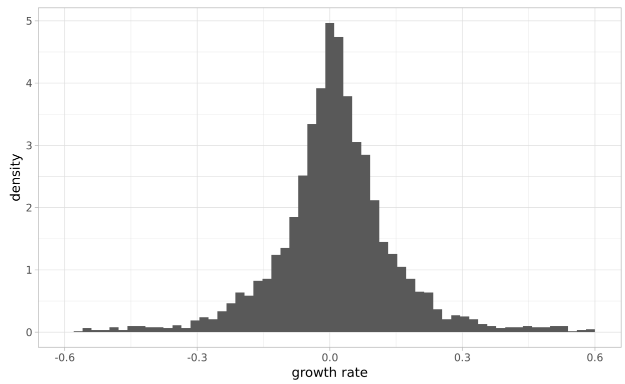


Figure 14: Histogram of the growth rates of the estimated TFP.

6.2 Imputation of Missing Data

TBA

6.3 Estimation of Productivity Shocks

In Section 3, the growth rates of sales and average labor productivity are used as a proxy for microeconomic productivity shocks. Here, to confirm that this choice does not affect our findings, we estimate the TFP shocks by using the method proposed by Ackerberg et al. (2015). That is, we estimate the production function at the sector-level by taking account of the endogeneity concern (i.e., some unobservable shocks might be associated with inputs of the production) and measure TFP as the residual of the change in value-added minus the contributions of the changes in labor and capital.[25] Then, we define the TFP shock as the log difference in the estimated TFP between two successive years for each firm.

The summary statistics and histogram of the estimated TFP shocks are given in **Table 7** and **Figure 14**, respectively. Similar to **Figure 6**, the histogram of the estimated TFP shocks deviates from a Gaussian but is close to a Laplace distribution. Indeed, as in Section 3.2.2, we estimate the parameters of the generalized Gaussian distribution and find that the estimate of the shape parameter γ is $0.832(0.025)$, which is relatively close to 1 (i.e., a Laplace distribution).[26] This shows that the Laplace shape observed is not due to our

[25] Following the extant study using the TSR data to estimate TFP under the ACF approach (Ito and Miyakawa (2022)), we employ capital investment as a proxy to control for unobservable productivity.

[26] We find that the estimated TFP shocks have relatively many outliers. In **Table 3.2.2** and the estimation of the generalized Gaussian

choice of the proxy but is observed when the estimated TFP shocks are used. Thus, since Assumption 2.1 is considered a reasonable assumption, our finding about the closeness to a Gaussian still is not affected by the choice of the proxy of microeconomic shocks.

6.4 Robustness

In Section 4, we used Assumption 2.1 on a quarterly basis by adjusting the parameter b because the empirical counterpart of aggregate output (i.e., the GDP growth rate) is on a quarterly basis. However, since the Laplace shape observed in Section 3.2.2 is on an annual basis, it would be preferable to consider quarterly growth rates whose annual distribution is the Laplace distribution for the underlying microeconomic shocks. Here, we consider these quarterly growth rates and apply the Edgeworth series approximation to check whether our finding in Section 4.3 (i.e., the closeness of the distribution of Z to a Gaussian) still holds in this case. In particular, this case provides another example showing how the averaging effect makes the distribution of aggregate output closer to a Gaussian.

Consider the distribution of quarterly growth rates from its moment generating function. Let $\epsilon_{q,k}$ be the kth quarterly growth rate for $k = 1, 2, 3, 4$ and let ϵ_a be the annual growth rate, that is, $\epsilon_a = \sum_{k=1}^{4} \epsilon_{q,k}$ If the distribution of ϵ_a is the Laplace distribution with parameter b, its moment generating function is

$$Ee^{t\epsilon_a} = (1 - b^2t^2)^{-1}$$

If $\epsilon_{q,k}$ are independent random variables drawn from a common distribution, then we have $Ee^{t\epsilon_a} = Ee^{t\sum_{k=1}^{4}\epsilon_{q,k}} = (Ee^{t\epsilon_{q,k}})^4$. This yields that

$$Ee^{t\epsilon_{q,k}} = (1 - b^2t^2)^{-1/4}$$

Using this equation, we derive the series of cumulants as in Section 2.1.3. In particular, the cumulant generating functions for the two variables are related as follows:

$$K_{\epsilon_{q,k}}(t) = \frac{1}{4}K_{\epsilon_a}(t)$$

For example, the second and fourth cumulants are given by $\kappa_2(\epsilon_{q,k}) = \frac{1}{4}\kappa_2(\epsilon_a)$ and $\kappa_4(\epsilon_{q,k}) = \frac{1}{4}\kappa_4(\epsilon_a)$, respectively. Thus, the excess kurtosis for the two variables are related as follows:

$$\text{excess kurtosis of } \epsilon_{q,k} = \frac{\kappa_4(\epsilon_{q,k})}{\kappa_2^2(\epsilon_{q,k})} = 4 \times \text{excess kurtosis of } \epsilon_a$$

Since the excess kurtosis for a Laplace distribution is 3, the excess kurtosis of the quarterly growth rate is $4 \times 3 = 12$. That is, the distribution of the quarterly growth rate features a higher excess kurtosis than a Laplace.

As in Section 4.3, let us consider how much the excess kurtosis decreases through the summation Eq.(1). Since the excess kurtosis of the normalized aggregate output is given by Eq.(5), we find that under this condition, the excess kurtosis of the normalized aggregate output is $0.047 \times 12 = 0.564$ for the case of an

distribution, we exclude the observations of TFP shocks with their absolute values larger than 1.0.

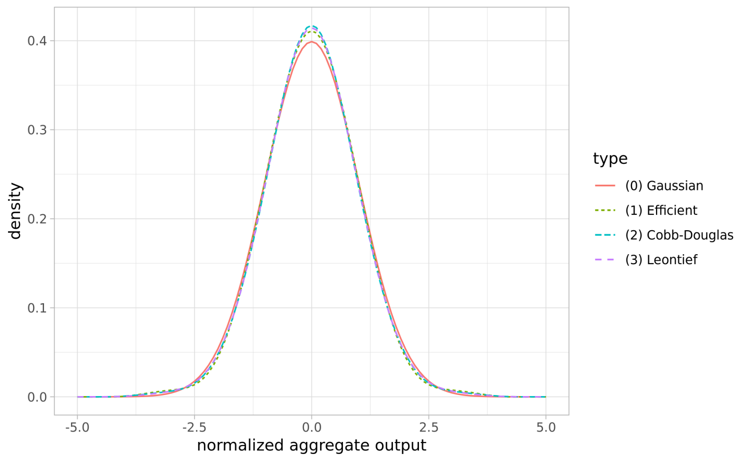


Figure 15: Distribution approximation by Edgeworth series. For comparison, the standard Gaussian distribution is also plotted.

efficient economy. This value is sufficiently close to 0, suggesting that the resultant distribution of aggregate output is close to a Gaussian (recall that the excess kurtosis is equal to 1.2 for a logistic distribution, which is known to resemble a Gaussian).

This point can be confirmed by the Edgeworth series approximation.[27] The result is given in **Figure 15**, comparing the approximated distribution of the normalized aggregate output with the standard Gaussian distribution. While a slight deviation at the center can be observed, the approximated distribution is sufficiently close to a Gaussian, consistent with our analysis of the excess kurtosis above. That is, this suggests that even when the distribution with a higher excess kurtosis than a Laplace distribution is considered, the averaging effect still dominates, making the resultant distribution of aggregate output sufficiently close to a Gaussian. Thus, the empirical granularity is not large enough to compensate the averaging effect, confirming the robustness of our finding in Section 4.3.

[27]In the Edgeworth series approximation here, we use the first ten terms in Eq.(6) to approximate the distribution.

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. (2017). Microeconomic Origins of Macroeconomic Tail Risks. *American Economic Review*, 107(1):54–108.

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

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Altinoglu, L. (2021). The origins of aggregate fluctuations in a credit network economy. *Journal of Monetary Economics*, 117:316–334.

Arata, Y. (2021). The Role of Granularity in the Variance and Tail Probability of Aggregate Output. *RIETI Discussion 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. (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica*, 87(4):1155–1203.

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

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

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

Boehm, C. E., Flaaen, A., and Pandalai-Nayar, N. (2019). Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tohoku Earthquake. *Review of Economics and Statistics*, 101(1):60–75.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 121:1–14.

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

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

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

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

Cramér, H. (1936). Über eine eigenschaft der normalen verteilungsfunktion. *Mathematische Zeitschrift*, 41(1):405–414.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American economic review*, 102(6):2437–71.

Dew-Becker, I. (2022). Tail risk in production networks. Technical report, 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., Kll Uppelberg, C., and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Technical report.

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.

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*.

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media.

Ito, Y. and Miyakawa, D. (2022). Performance of exiting smes in japan: An empirical analysis using exit mode data. Technical report, IMES Discussion Paper Series.

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.

Linnik, Y. V. and Ostrovskii, I. (1977). *Decomposition of random variables and vectors*.

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

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

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

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

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

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