Granular Firms and Aggregate Fluctuations*

Omar Blanco-Arroyo† Simone Alfarano‡

January 25, 2024

Abstract

We propose a framework that is able to explain the behavior of the contribution of idiosyncratic shocks to GDP fluctuations. We show that, when the firm size distribution and the relationship between size and volatility is Pareto distributed, GDP volatility depends on the Domar weight of the largest firm, the relative volatility of largest firms, the tail exponent and the size-volatility elasticity. We find that the maximum contribution of the granular channel can rationalize 23% of GDP volatility in Spain. Finally, we estimate that the number of granular firms in the Spanish economy is approximately 50 firms.

JEL Classifications: D22, D24, E32, L11, L25

Keywords: granularity, aggregate fluctuations, idiosyncratic shocks, firm-level volatility, firm size distribution

*This paper is a revised version of Chapter 2 of Omar Blanco-Arroyo's Ph.D. thesis. Omar Blanco-Arroyo thanks Simone Alfarano and Gabriele Tedeschi for making possible this dissertation. For helpful comments and suggestions, we thank Annarita Colasante, Philipp Mundt, Maria Cristina Recchioni, Alberto Russo, Andrea Teglio and seminar participants at various institutions. Blanco-Arroyo acknowledges the financial support of the Universitat Jaume I (grant PREDOC/2017). All remaining errors are our own

†Department of Applied Economics II, Universitat de València (omar.blanco@uv.es).

‡Department of Economics, Universitat Jaume I (alfarano@uji.es).

1 Introduction

Traditionally, business cycles theories have dismiss the possibility that *microeconomic* shocks may originate aggregate fluctuations due to a "diversification" argument (Lucas, 1977). Gabaix (2011) seminal work challenges the convention by introducing the "granular" hypothesis: in the presence of significant heterogeneity at the micro level, the behavior of macroeconomic aggregates is attributable to the incompressible "grains" of economic activity, the large firms.[1] In this view, idiosyncratic shocks to the *granular* firms play a crucial role in shaping aggregate fluctuations. And yet, we lack a framework that provides a theoretically founded method for identifying the number of granular firms that populate a granular economy.

The first attempt to quantify the granular size of the economy (i.e., the number of granular firms) is made by Blanco-Arroyo et al. (2018) empirical work, who find that the contribution of idiosyncratic shocks to gross domestic product (GDP) fluctuations increases rapidly when the very top firms are taken into account and an almost steady value from a given number of firms onwards. They refer to this behavior as "granular curve".[2] The granular curve clearly shows two well differentiated regimes: the granular regime, which is composed of a small number of large firms whose idiosyncratic perturbations can lead to aggregate fluctuations, and the atomistic regime, which is composed of those firms whose effect on the aggregate is negligible. Blanco-Arroyo et al. (2018) propose an empirical method to estimate the granular size based on replacing large firms by smaller ones and comparing the resulting granular curve with the counterfactual case in which all firms are of equal size. The granular size is then determined by the number of large firms that, once removed, cause the empirical curve to converge to the counterfactual case. However, this procedure seems to be too conservative, as the convergence point is much larger than point that visually represents the change of regime in the granular curve.

This paper seeks to shed light on the determinants of the granular curve behavior and to quantify the granular size of a granular economy more precisely. Building on the models developed by Gabaix (2009a) and Carvalho and Gabaix (2013), we setup a conceptual framework that traces back the volatility of GDP growth to large firms' idiosyncratic shocks. We show that, when the distribution of firm size is power law (see, e.g., Axtell (2001), Luttmer (2007), di Giovanni and Levchenko (2013)) and the firms' idiosyncratic volatility depends on size as a power law (see, e.g., Stanley et al. (1996), Koren and Tenreyro (2013), Yeh (2017)), GDP fluctuations are shaped by five parameters that capture the large firms dynamics: (i) productivity multiplier, (ii) Domar weight (Domar, 1961) of the largest firm, (iii) volatility of the largest firm, (iv) tail index of firm size distribution, and (v) size-volatility elasticity.

Theoretically, our framework provides three key results. First, the largest firm contains a great deal of information on the characteristics of the economy and plays a crucial role in driving aggregate fluctuations. This result is in line with Carvalho and Grassi (2019), who develop a heterogeneous firm dynamics setup in which aggregate fluctuations are caused by firm-level disturbances alone and conclude that business cycles have a "small sample" origin.

Second, the granular contribution to aggregate fluctuations is bounded. When the firm size distribution is power law, the contribution of idiosyncratic shocks to aggregate fluctuations exhibits an asymptotic value. This finding is in line with Yeh (2021), who explores the effect the contribution of idiosyncratic shocks

[1] Recent contributions that also seek to understand the microeconomic underpinnings of aggregate fluctuations are Acemoglu et al. (2012), di Giovanni and Levchenko (2012), Carvalho and Gabaix (2013), di Giovanni et al. (2014), **?**), Baqaee (2018), Baqaee and Farhi (2019), Carvalho and Grassi (2019).

[2] This type of behavior has also been documented in Brazil (Silva and Da Silva, 2020) and Kazakhstan (Konings et al., 2021).

to aggregate fluctuations when including the size-volatility relationship. The fact that exists a maximum contribution leaves room to traditional alternative factors, such as oil and monetary policy shocks, and amplification mechanism, such as "cascade effects" propagated throughout the input-output network (Acemoglu et al., 2012).

Third, the effect of the size-volatility relationship in shaping aggregate fluctuations is non-negligible. The literature that studies the granular origins of aggregate fluctuations has typically downplayed the effect of the size-volatility relationship by arguing that the estimates come from biased and non-representative samples (Gabaix, 2011).[3] Recently, Yeh (2021) estimates the relationship using the universe of U.S. firms and concludes that it is statistically different from zero even when taking the large firms only. We show it is incompatible to assume that the weak form of Gibrat (1931) law for volatilities holds and use the volatility of the largest firms (e.g., Gabaix (2011), Carvalho and Grassi (2019)) and that changes in the size-volatility relationship have greater impact on aggregate fluctuations than tail index changes, which have been the main object of study.

We then employ our setup to study the granular curve behavior observed in the data. As in Blanco-Arroyo et al. (2018), we focus on the top 1000 Spanish firms. The estimation of the parameters support the hypotheses on which it is based: the distribution of firm size and the size-volatility relationship follow a power law behavior.

Empirically, we show that the granular curve is well characterized by our framework and find that the average maximum contribution of top Spanish firms's idiosyncratic shocks to the GDP fluctuations is 23%. This estimate is in line with previous empirical estimations that are purely econometric (see, e.g., Gabaix (2011), Blanco-Arroyo et al. (2018), Fornaro and Luomaranta (2018), Miranda-Pinto and Shen (2019), Silva and Da Silva (2020)). Then, we propose a set of measures that allow to quantify the granular size of the economy more precisely than the empirical method initially proposed by Blanco-Arroyo et al. (2018), as the estimated size is closer to the point that visually represents the change from the granular to the atomistic regime than the empirical method initially proposed. In particularly, we estimate that the granular size of the Spanish economy is approximately 50 firms.

The results are robust to changes in the number of firms and to time-varying parameters. Our baseline estimation focuses on the largest 1000 firms and considers the entire period available. In an alternative approach, we increase the number of firms to 2500 in steps of 500 and find that our framework continues to characterize the empirical granular curve and the calibrated number of firms remains in the region that we visually identify as the change of regime. We also explore the granular curve behavior in smaller time windows. After calibrating the volatility of the largest firm, we show that the framework provides a good characterization of the changes observed in the empirical granular curve through the business cycle. Finally, we find that the granular size exhibits a cyclical behavior: the number of firms whose idiosyncratic shocks have an impact on the aggregate grows in expansion phases and shrinks in recession phases.

Related literature Our paper draws on, and contributes to, two strands of literature: the granular origins of aggregate fluctuations and the empirical industrial dynamics literature. Our conceptual framework

[3] This critique stems from the fact that the estimation has typically been carried out using firms in *Compustat* database. As argued Gabaix (2011), Compustat only comprises large traded firms that are expected to be more volatile than non-traded firms, as small volatile firms are more prone to seek outside equity financing, while large firms are in any case very likely to be listed in the stock market.

sheds light on the components that drive Gabaix (2011) "granular residual" and, hence, relates to the recent empirical literature that investigates the proportion of aggregate shocks that can be accounted for by idiosyncratic to the large firms (see, e.g., Gabaix (2011), di Giovanni et al. (2014), Stella (2015), Magerman et al. (2016), Yeh (2017)).[4] It also relates to the scarce theoretical literature that studies how large firms dynamics shape aggregate fluctuations (di Giovanni and Levchenko, 2012, Carvalho and Grassi, 2019, Daniele and Stüber, 2020, Gaubert and Itskhoki, 2021). Although, unlike the literature, our framework takes the firm size distribution and the size-volatility relationship as exogenously given.

This paper is also related to the empirical industrial dynamics literature that studies the firm size distribution (see, e.g., Axtell (2001), Gaffeo et al. (2003), Fujiwara et al. (2004), Luttmer (2007), di Giovanni and Levchenko (2013)) and the size-volatility relationship (e.g., Stanley et al. (1996), Lee et al. (1998), Sutton (2002), Koren and Tenreyro (2013), Calvino et al. (2018), Yeh (2021)). In line with the bulk of the recent literature, we find that the upper tail of the firm size distribution follows a power law with exponent larger than one and that the weak form of Gibrat (1931) law for volatilities, typically assumed in the granular literature, does not hold for the largest firms in the economy.

Outline The remainder of this paper is organized as follows. Section 2 discusses the conceptual framework that traces back the origins of business cycles fluctuations to large firm's dynamics. Section 3 presents the data and estimates the variables that constitute our model. Section 4 characterizes the behavior of the empirical contribution of idiosyncratic shocks to large firms to aggregate fluctuations, quantifies the granular size of the economy and explores its cyclical behavior. Section 5 concludes. Derivations and robustness checks can be found in the Appendix.

2 Conceptual framework and motivation

This section builds on the models develop by Gabaix (2009a) and Carvalho and Gabaix (2013) to shed light on how idiosyncratic firm shocks shape aggregate fluctuations. We follow the literature and assume that the upper tail of the firm size distribution and the size-volatility relationship follow a power law. Under these assumptions, GDP growth volatility is driven by five components: (i) productivity multiplier, (ii) "Domar" weight of the largest firm, (iii) volatility of the largest firm, (iv) tail index of firm size distribution, and (v) size-volatility elasticity. Furthermore, changes in the size-volatility elasticity have a larger impact on aggregate volatility than changes in the tail index.

2.1 Conceptual framework

Consider an economy populated by n competitive firms that produce intermediate and final goods using capital, labor and intermediate inputs supplied from one another. According to Hulten (1978), after a Hicks-neutral idiosyncratic productivity shock $\varepsilon_i = dA_i/A_i$ to firm i, the shock to aggregate total factor productivity (TFP) Λ is

$$\frac{d\Lambda}{\Lambda} = \sum_{i=1}^{n} \frac{S_i}{Y} \varepsilon_i, \tag{1}$$

[4] Previous literature that seeks to understand the microeconomic underpinnings of aggregate fluctuations includes Jovanovic (1987), Durlauf (1993), Bak et al. (1993), Nirei (2006).

where S_i is firm i's value of sales (gross output) and Y is GDP (aggregate value added). S_i/Y is the so-called "Domar" weight (Domar, 1961). The sum of the Domar weights in (1) can be greater than one. This reflects the fact that the change in factor efficiency creates extra output, which serves to increase final demand and intermediate inputs—see Carvalho and Gabaix (2013) for an intuition.[5] The weighted sum of idiosyncratic productivity shocks is none other than Gabaix (2011) "granular residual" (see Section 4.1).

Gabaix (2009a) and Carvalho and Gabaix (2013) show that, in absence of other disturbances, GDP growth dY/Y is proportional to TFP growth $d\Lambda/\Lambda$: $dY/Y = \mu d\Lambda/\Lambda$, for some productivity multiplier $\mu \geq 1$. Thus, GDP growth is equal to

$$\frac{dY}{Y} = \mu \sum_{i=1}^{n} \frac{S_i}{Y} \varepsilon_i. \tag{2}$$

Assume that productivity shocks are uncorrelated across firms (i.e., $\mathrm{cov}\left(\varepsilon_i, \varepsilon_i\right) = 0 \; \forall i$) and firm i's has a variance of shocks $\sigma_i = \mathrm{var}\left(\varepsilon_i\right)$.[6] Then, we have that the volatility of GDP growth is

$$\sigma_Y = \mu \sqrt{\sum_{i=1}^{n} \left(\frac{S_i}{Y}\right)^2 \sigma_i^2}. \tag{3}$$

The square root of the weighted sum is Gabaix (2011) "granular" volatility, Carvalho and Gabaix (2013) "fundamental" volatility and di Giovanni et al. (2014) "direct effect".

Gabaix (2011) seminal work introduces the "granular" hypothesis: in the presence of significant heterogeneity at the firm-level, economic fluctuations are attributable to the incompressible "grains" of economic activity, the large firms. The intuition is as follows. When the distribution of firm size in Equation (3) is sufficiently fat-tailed, idiosyncratic shocks to the granular firms do not die out in the aggregate, because they do not cancel out with shocks to smaller firms. Thus, the origins of aggregate fluctuations can be traced back to the dynamics of the *granular* firms.

Our first goal is to shed light on the industrial dynamics factors that drive the volatility of GDP growth. To this end, we first study the distribution of firm size—measured by the value of sales—and then the volatility of idiosyncratic productivity shocks.

2.2 Firms size distribution

A plethora of empirical evidence finds that the the entire firm size distribution, or at least its upper tail, is well approximated by a power law (see, e.g., Axtell (2001), Fujiwara et al. (2004), Luttmer (2007), di Giovanni and Levchenko (2013), among many others).[7,8] Given our focus on the large firms, the evidence put forth by the literature in favor of the power law distribution makes this a natural baseline to consider. Therefore,

[5] Hulten (1978) first-order approximation for frictionless, efficient economies has recently been extended by Baqaee and Farhi (2019, 2020) to study the role played by second-order effects, such as complementarity, substitutability, returns to scale, factor reallocation, and network structure.

[6] Throughout this section, we drop the time subscript for the sake of simplicity.

[7] See Gabaix (2009b) for a review of power laws in economics and finance.

[8] Gibrat (1931) and the literature that followed (see Sutton (1997) for a review) describe the firm size distribution by a lognormal. Recent studies using census data conclude that the lognormal behavior emerges in non-representative samples (Axtell, 2001).

we assume that the counter cumulative distribution function (CCDF) of sales S is characterized by

$$\mathbb{P}\left(\text{firms} > S_i\right) = \left(\frac{S_{\min}}{S_i}\right)^{\zeta}, \tag{4}$$

for $S_i > S_{\min}^{1/\zeta}$, with $\zeta \in [1, 2)$. The CCDF (4) corresponds to a density $p\left(S_i\right) = \zeta S_{\min}^{\zeta} S_i^{-(\zeta+1)}$. We introduce introduce the cut-off $S_{\min}$ to account for the fact that only the upper tail of the sales distribution could display a power-law behavior (Fujiwara et al., 2004). We bound the tail index ζ in the range $[1, 2)$ to ensure that the distribution is fat-tailed and, hence, the economy is granular. Traditionally, business cycle theories have discarded the possibility that aggregate fluctuations may originate from microeconomic shocks to firms due to a "diversification argument" (Lucas, 1977). In particular, in an economy populated by a large number n of firms hit by independent shocks, the law of large numbers applies and, hence, GDP volatility would be roughly proportional to $1/\sqrt{n}$—a negligible effect. As shown by Gabaix (2011), this would be the case if $\zeta \geq 2$. However, when ζ lies in the range $[1, 2)$, as estimated by the literature above, the law of large numbers does not apply and GDP volatility decays much slower. For instance, when $\zeta = 1$, known as Zipf's law (Zipf, 1949), the rate of decay is $1/\ln n$. Thus, shocks to individual large firms may translate into aggregate fluctuations.

2.3 Size-volatility relationship

The works of Meyer and Kuh (1957) and Hymer and Pashigian (1962) are the first to document the negative relationship between firm's volatility, measured by the standard deviation of firm's sales growth rate, and its size, measured by the average value of sales. Additional contributions find that this relationship is described by a power law (see, e.g., Stanley et al. (1996), Lee et al. (1998), Sutton (2002), Koren and Tenreyro (2013), Calvino et al. (2018), Yeh (2021)). We follow the literature and assume that the power-law behavior also holds for the relationship between size and volatility of shocks. Thus, the relationship between the volatility of the idiosyncratic productivity shock σ_i and the value of sales S is described by the law

$$\sigma_i\left(S\right) = \sigma_{\min}\left(\frac{S_{\min}}{S_i}\right)^{\alpha}, \tag{5}$$

with $\alpha \in [0, 1/2]$. As in Equation (4), we introduce the cut-off $S_{\min}$ and its corresponding volatility $\sigma_{\min}$. The intuition typically provided to explain the limiting cases $\alpha = 0$ and $\alpha = 1/2$ is based on a diversification argument. As argued by Amaral et al. (1997), in a firm made up of many units, which are of identical size and grow independently of one another, fluctuations as a function of size decay as a power law with an exponent $\alpha = 1/2$ because the law of large numbers applies. On the contrary, if there are very strong correlations between the units, the growth dynamics are indistinguishable from the dynamics of structureless organizations and, hence, $\alpha = 0$. The latter is the case predicted by Gibrat (1931) weak law, namely, there is no size dependence of σ. Thus, the average volatility is well captured the volatility of all firms (i.e., $\sigma_i = \overline{\sigma}\ \forall i$, where $\overline{\sigma}$ is the average volatility).

The literature above estimates α between the two limiting cases even for large firms.[9] The most common

[9] Some exceptions are Hall (1987) and Haltiwanger et al. (2013), who find that Gibrat's law holds for large firms, as deviations observed in the data are attributable to the dynamics of small entrants. Yeh (2021) estimates the relationship using the universe of U.S. firms and finds a strong size-variance relationship even when excluding entrant firms.

mechanisms proposed to explain the size-volatility relationship are based on output (Klette and Kortum, 2004) and establishment (Foster et al., 2001, 2006) diversification. Recently, Yeh (2017) rules out these mechanisms and concludes that large firms face smaller price elasticities and therefore respond less to a given-sized productivity shock than small firms do, as implied by Decker et al. (2020). Despite the lack of consensus, it is important to emphasize that our results do not hinge on a particular microfoundation.

2.4 Aggregate fluctuations

Proposition 1 (GDP fluctuations)**.** *If the firm size distribution and the relationship between size and volatility are power-law, then GDP fluctuations have the following form. If $\zeta' \neq 1$,*

$$\sigma_Y = \mu \frac{S_{\max}}{Y} \sigma_{\max} \left\{ \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) n^{1-2/\zeta'} \right] \right\}^{1/2}, \tag{6}$$

where $S_{\max}/Y$ and $\sigma_{\max}$ are, respectively, the Domar weight and volatility of the largest firm, $\Gamma(\cdot)$ is the Gamma function, n is the number of firms that populate the economy and the tail index $\zeta' \equiv \zeta/(1-\alpha)$ consists in the tail index of the firm size distribution ζ and the size-volatility elasticity α. If $\zeta' = 1$,

$$\sigma_Y = \mu \frac{S_{\max}}{Y} \overline{\sigma} \frac{\pi}{\sqrt{6}}, \tag{7}$$

where $\overline{\sigma}$ is a representative volatility.

Proof. See Appendix A. □

According to Equation (7), when the firm size distribution is Zipf (1949) (namely, the tail index ζ is equal to 1) and the weak form of Gibrat (1931) law for variances holds (namely, the elasticity α is equal to 0), the volatility of GDP growth caused by idiosyncratic shocks alone is determined by the following firm dynamics variables: productivity multiplier, Domar weight of the largest firm and representative volatility. On the other hand, when deviations from Zipf law (i.e., $\zeta \in (1,2)$) and/or Gibrat law (i.e., $\alpha \in (0,1/2]$) exist, Equation (6) shows that the volatility of GDP growth is driven instead by the following variables: productivity multiplier, Domar weight and volatility of the largest firm, number of firms in the economy, tail index and size-volatility elasticity.

We follow Gabaix (2011) and quantify the contribution of idiosyncratic shocks to the volatility of GDP growth using the R^2 statistic. If $\zeta' \neq 1$, then

$$R^2 = \mu^2 \left(\frac{S_{\max}}{Y}\right)^2 \left(\frac{\sigma_{\max}}{\sigma_y}\right)^2 \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) n^{1-2/\zeta'} \right]. \tag{8}$$

Given the large number of firms that populate an economy, the contribution exhibits an upper bound:

$$\mathcal{A} = \mu^2 \left(\frac{S_{\max}}{Y}\right)^2 \left(\frac{\sigma_{\max}}{\sigma_Y}\right)^2 \frac{2}{2-\zeta'} \tag{9}$$

According to Equation (9), $\mathcal{A}$ increases when any of the following changes take place: the share of economic activity commanded by the largest firm increases, the idiosyncratic volatility of the largest firm with respect to GDP volatility increases, the firm size distribution becomes more homogeneous (i.e., ζ increases) and the elasticity of volatility to size (i.e., α increases). Note that changes in ζ impact on $S_{\max}$ and, in turn, on $\sigma_{\max}$. Appendix B discusses how they are related.

If $\zeta' = 1$, then the contribution is

$$R^2 = \mu^2 \left(\frac{S_{\max}}{Y}\right)^2 \left(\frac{\overline{\sigma}}{\sigma_Y}\right)^2 \frac{\pi^2}{6}, \tag{10}$$

which coincides with the asymptotic contribution (i.e., $R^2 = \mathcal{A}$). Equation (10) represents the maximum contribution of idiosyncratic shocks to GDP volatility because the size of the largest firm in a Zipf distribution is larger than in a power law with tail index greater than 1 (Newman, 2005) and the average volatility is larger than the volatility of the largest firm (see, e.g., Comin and Philippon (2005), Comin and Mulani (2006)).

As mentioned above, the literature (and also our estimates in Section 3) shows that there are deviations from Zipf and Gibrat laws in data. Thus, throughout this paper, we characterize GDP volatility using Equation (6). Interestingly, this characterization suggests to channels through which the contribution decreases: sizes become more homogeneous and volatility is less elastic to changes in size. To which of the two channels the contribution is more sensitive is answered by the following proposition.

Proposition 2. *Let δ be deviations from Zipf's law baseline case, i.e., tail index of the firm size distribution is $\zeta = 1 + \delta$. Then, the excess of sensitivity of the contribution to changes in the size-volatility elasticity with respect to changes in the tail index is*

$$\frac{\partial R^2/\partial \alpha}{\partial R^2/\partial \delta} = \zeta' \geq 1. \tag{11}$$

Proof. See Appendix B. □

The role of the size-volatility relationship in shaping aggregate fluctuations has been omitted in the granularity literature, as leading modeling assumption is that Gibrat's law holds (Gabaix, 2011, Carvalho and Grassi, 2019). Equation (11) extends Yeh (2021) results by showing that deviations from Gibrat's law, not only attenuate significantly the impact of the granular contribution, but also that they play an even more important role than changes in the cross-sectional dispersion in firm size, which is the main object of study in the literature.

3 Data and measurement

In this section, we present the data set on which the results of this work are based and the procedure followed to estimate the parameters that drive aggregate volatility. Our findings are as follows. First, the time average aggregate productivity multiplier is well approximated by its micro-founded version. Second, the Domar weight of the largest has remained rather stable through time. Third, in line with the bulk of the literature presented above, the upper tail of the firm size distribution is well characterized by a power law. Fourth, the size-volatility relationship estimated for the top 1000 firms follows a power law.

3.1 Data and summary statistics

Firm-level data come from SABI (*Sistema de Análisis de Balances Ibéricos*) database. The database is compiled by Bureau Van Dijk Electronic Publishing (BvD). SABI includes information on both listed and unlisted Spanish firms collected from various sources, such as national registers and annual reports. The fact that the data set provides information on unlisted firms is a crucial to avoid strong selection bias, as some of the largest Spanish firms are privately held. The main variables used in the analysis are net sales and number of employees for each firm. The time period is 1994-2018. During this lapse of time, the Spanish economy experienced a rapid economic growth, followed by a double recession (2008:II-2009:IV and 2010IV-2013:II).[10,11]

Given our focus on large firm dynamics, we build our dataset using the largest 200,000 Spanish firms in SABI. We attenuate the impact of exogenous shocks by excluding those firms that are engaged in oil, oil-related and energy activities because their sales come mostly from worldwide commodity prices, rather than real productivity shocks. We also exclude financial and public firms because their sales do not mesh well with the meaning used.[12] Recently, Cravino and Levchenko (2017) and di Giovanni et al. (2018, 2020) find evidence suggesting that foreign shocks are transmitted to the domestic economy through the largest firms and its affiliates. We mitigate the impact of foreign shocks by restricting the sample to those firms whose "global ultimate owner" is based in Spain.[13] We use unconsolidated sales denominated in euros, since sales that are consolidated across the multiple firms that comprise the corporation overestimate the impact of multinational firms and do not provide a reliable picture of the evolution of large firms (Gutiérrez and Philippon, 2019).[14] The resulting sample comprises the top 75,000 Spanish firms.

SABI, as well as other BvD products,[15] has a low coverage for years previous to 1995 and a reporting lag of roughly two years. This particularly affects the years 1994 and 2018 in our sample. We try to overcome these limitations by interpolating missing values with a maximum gap of two consecutive periods. This procedure does not change our conclusions and allows us to increase the representativeness of the sample substantially.[16]

The contant GDP expressed in 2015 euros and GDP deflator come from the OECD's *National Accounts Statistics* (SNA) database (OECD, 2020a). GDP per capita is calculated using total population coming from OECD's SNA database (OECD, 2020b). Total factor productivity (index 100 in 2015) is obtained from the Bank of Spain.[17]

Table 1 presents summary statistics for firm-level growth rates for the whole sample. The average growth

[10]See Fernandez-Villaverde et al. (2013) and Royo (2013), respectively, for a detailed explanation of the causes and consequences of the economic boom in Spain.

[11]Recession dates are taken from *Asociación Española de Economía* (AEE).

[12]Firms are filtered our using the four-digit SIC primary code. See Appendix C in Gabaix (2011).

[13]A more suitable approach would be to retain those firms whose headquarters are located in Spain. Unfortunately, SABI does not provide this information. We use the global ultimate owner to identify whether the firm is a parent or an affiliate. In the case of individuals and families, the country reported is the country of residence. In the case of firms, it is the country where the firm is based.

[14]In particular, we downloaded companies with unconsolidated accounts only (consolidation code U1) and companies that present both consolidated and unconsolidated accounts (consolidation code C2/U2).

[15]Kalemli-Ozcan et al. (2023) discuss in detail how to use ORBIS and AMADEUS (the Global and European supersets of SABI, repectively) to construct representative firm-level datasets.

[16]Figure A.3 shows the number of firms affected by the linear interpolation procedure through time and the number of firms with valid observations.

[17]The time series can be found in the summary indicators table "Structural Indicators of the Spanish economy and of the European Union" (Table 1.4).

Table 1: Summary statistics.

	Weight	Sales	Employees	Productivity
Average aggregate growth rate		0.065	0.063	0.069
Average individual growth rate		0.126	0.074	0.030
		Standard deviation of growth rate		
Sample	0.367	0.415	0.323	0.429
0 - 20 size percentile	1.39×10^{-6}	0.694	0.427	0.681
21 - 40 size percentile	0.001	0.328	0.307	0.388
41 - 60 size percentile	0.008	0.352	0.299	0.387
61 - 80 size percentile	0.040	0.306	0.252	0.321
81 - 100 size percentile	0.317	0.282	0.221	0.292
Top 1000	0.229	0.296	0.227	0.305
Top 100	0.124	0.340	0.252	0.363
Top 10	0.057	0.276	0.223	0.316
Average $\sqrt{\mathcal{H}}$		0.065	0.068	

Notes: "Weight" refers to the sum of the Domar weights. "Productivity" refers to labor productivity proxied by the log of the sales per employee ratio, as in Gabaix (2011). "Standard deviation of growth rate" reports the time average standard deviation of growth rates within a percentile category. $\mathcal{H}$ is the Herfindahl index of the total firm shares.

rate of aggregate sales and employees is lower than the unweighted average of firm-level growth rate. The reasoning is because smaller firms tend to grow faster than larger firms, conditional on survival. On the contrary, firm-level productivity, which is defined as the log of the sales per employee ratio (see Section 3.2), in smaller firms tend to grow slower than larger firms. This is to be expected, as smaller firms are, on average, less efficient than larger firms (Taymaz, 2005). The table also reports the sum of Domar weights and the averages of firm volatility, measured by the standard deviation, for each size quintile. The results show that exist a high degree of heterogeneity and that smaller firms are more volatile than large firms. Finally, the square root of the Herfindahl index of sales and employees shares have an order of magnitude consistent with that reported by Gabaix (2011) and suggest that the economy is "granular".

3.2 Idiosyncratic shocks

Following Gabaix (2011), we focus on the labor productivity shocks.[18] We proxy firm-level labor productivity using the log of its sales per worker ratio: $z_{it} := \text{Sales}_{it}/\text{Employees}_{it}$.[19] The growth rate is then defined simply as $g_{it} = \Delta \ln z_{it}$, where Δ denotes the difference between years t and $t-1$. Firm-level growth rates are computed using only firms present in the dataset in both years, so that it captures the *intensive*

[18] Gnocato and Rondinelli (2018) estimate the granular residual with labor productivity shocks and firm-level TFP shocks. They show that both proxies for productivity shocks are highly correlated. See also Syverson (2004).

[19] We use this revenue-based productivity measure because it is not data intensive and is widely used in the literature. An important caveat is that it confounds idiosyncratic demand and factor price affects with efficiency differences (Foster et al., 2008). Therefore, it is not a clear measure of productivity shock, as it would be a measure based on quantities of physical output. Empirically, however, both measures are strongly correlated (see Foster et al. (2008)).

margin growth rates.[20] Suppose that innovations to g_{it} evolve according to the following one-factor model: $g_{it} = \eta_t + \varepsilon_{it}$, where η_t is a common shock and ε_{it} is an idiosyncratic shock. We make the identification assumption that $\mathbb{E}[\eta_t \varepsilon_{it}] = 0$. Firm i's labor productivity idiosyncratic shock in year t can be estimated as the deviation of its growth rate from the common shock to the top Q firms:

$$\varepsilon_{it}(Q) = g_{it} - \eta_t(Q). \tag{12}$$

Thus, ε_{it} captures the residual unexplained by the common shock. This approach to identifying firm-specific shocks is standard in macroeconomics (see, e.g., Koren and Tenreyro (2007), Gabaix (2011) and di Giovanni et al. (2014)).

To estimate Equation (12), we first need to estimate the common shock $\eta_t(Q)$. Since our goal is to assess the contribution of idiosyncratic shocks to the largest firms to aggregate fluctuations, we restrict our attention to the top $Q = 1,000$ firms, as in Gabaix (2011) robustness exercise. This choice is based on the fact that the one-factor model employed to extract ε_{it} implicitly assumes a certain degree of homogeneity among firms, which is likely to be a less good approximation for a large Q. The growth rate of productivity is expected to depend on firm characteristics and factors which, in turn, depend on size. If we consider a large number of firms with very heterogeneous size, firm characteristics can also be very heterogeneous and thus the implicit assumption $\eta_{it} = \eta_t \; \forall i$ may be a poor approximation. With this in mind, Section 4.3 shows the robustness of the results to alternative Qs. Once the number of potentially granular firms is set, we estimate the common shock to the top Q as the cross-sectional median productivity growth rate, as in Blanco-Arroyo et al. (2018). Given that the time dimension is somewhat limited and that the Great Recession was particularly severe in Spain, the median growth rate seems a more suitable estimate of the common shock that hit the largest firms during these years.[21]

The dataset contains some large outliers, which may be due to mergers, acquisitions or simply measurement errors. We follow the convention in the literature and mitigate their impact by *winsorizing* extreme shocks at 50%.[22] Recently, the winsoring procedure to handle extreme values and outliers has been criticized by Dosi et al. (2018), who argue that it is not necessary when analyzing granularity because large firms have more accurate accounting information and, therefore, do not suffer from large jumps. In addition, they show that Gabaix (2011) results are heavily influenced by such cleaning procedure. Taking into account Dosi et al. (2018) critique, Appendix C re-estimates the idiosyncratic shocks using the arc-elasticity proposed by Davis et al. (1996). The main advantage of this measure is that it allows us to avoid any winsorizing or trimming procedure. We show that our results do not depend on the definition of the productivity growth rate or the data cleaning strategy.

[20] In SABI, the *extensive margin* of entry and exit of firms cannot be calculated because it cannot be distinguished whether the newly observed firms are a genuine entry or an entry into the database. Using the universe of French firms, di Giovanni et al. (2014) show that the extensive margin plays no role in shaping aggregate fluctuations. Osotimehin (2019) finds that it contributes little to the variability of French aggregate productivity.

[21] The conclusions reached in the present paper remain unchanged if we use the mean growth rate to estimate the common shock instead.

[22] More precisely, we set $\hat{\varepsilon}_{it} = \text{sign}(\hat{\varepsilon}_{it})\, 0.5$ if $|\hat{\varepsilon}_{it}| > 0.5$. The winsorizing procedure affects 5% of the top Q Spanish firms in the time period 1995-2018. Results are not materially sensitive to the choice of that threshold.

3.3 Productivity multiplier

The frameworks set up by Gabaix (2009a) and Carvalho and Gabaix (2013), among many others, predict that GDP growth volatility is proportional to TFP growth by a factor μ that represents the productivity multiplier (see Equation (2)). Therefore, μ can be directly estimated by the following *relative standard deviations*:

$$\mu = \sigma_Y / \sigma_\Lambda, \tag{13}$$

where σ_Y and σ_Λ are the standard deviation of GDP per capita growth and TFP growth, respectively. According to (13), the estimated multiplier over the period 1994-2018 is 2.92. However, μ is expected to change through time, that is:

$$\mu_t = \sigma_{Yt} / \sigma_{\Lambda t}. \tag{14}$$

We compute the relative standard deviations at year t using a centered rolling window of 10 years. Alternatively, we obtain deviations from the Hodrick-Prescott trend of log GDP per capita and log TFP using a smoothing parameter 6.25 and compute the rolling window. Panel A in Figure 1 shows that the multiplier exhibits a clear cyclical behavior. The time average is equal to 3.41 in the two cases.

Additionally, we use the "granular" instrumental variable (GIV) methodology proposed by Gabaix and Koijen (2020) to estimate the productivity multiplier. Our IV is the granular residual (20), which is constructed using the estimated shocks from (12). The granular residual is a consistent and powerful IV because shocks are idiosyncratic and the firm size distribution presents a high degree of heterogeneity (see Table 1 and Section 3.4). We run the following ordinary least squares (OLS) regression:

$$g_{Yt} = \text{constant}\,(K) + \mu_{\text{GIV}}\,(K)\,\mathcal{E}_t\,(K) + u_t\,(K)\,, \tag{15}$$

for $K = 1, 2, \ldots, Q$, where $Q = 500, 1000, \ldots, 10,000$, g_{Yt} is the growth rate of GDP per capita, $\mathcal{E}_t$ is the granular residual and u_t is the error term. We estimate the productivity multiplier μ as the coefficient on the GIV $\mathcal{E}_t$. Equation (15) is also estimated by Gabaix (2011) to quantify the contribution of the idiosyncratic shocks to the top 100 U.S. firms (i.e., $K = 100$) to GDP growth fluctuations.

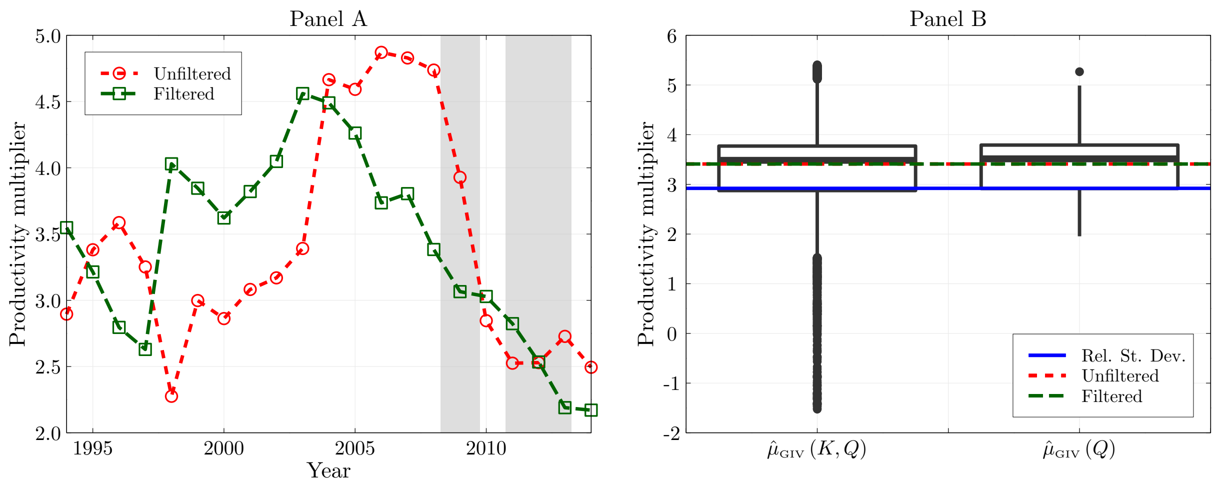
Panel B in Figure 1 shows the estimated productivity multiplier $\hat{\mu}_{\text{GIV}}$. We find that the median value (3.48) is almost identical to the time average multiplier estimated using (14). The multiplier estimated by Equation (13) can be seen as a lower bound. As the box plot of $\hat{\mu}_{\text{GIV}}\,(Q)$ renders clear, outliers are produced when the granular residual is constructed with a small number of firms, (i.e., small K).

In order to simplify the analysis and to be able to clearly identify the contribution of idiosyncratic shocks to aggregate fluctuations, in what follows, we follow the model presented in Section 2.1 and assume μ is constant thought time. Furthermore, we assume that μ does not depend on the number of large firms K. In line with the estimation provided by Blanco-Arroyo et al. (2018), we set $\mu = 3.5\ \forall t, K$.

3.4 Firm size distribution

The CCDF (4) implies that the probability of the largest firm $(S_{\min}/S_{\max})^\zeta$ has a frequency $1/n_{\text{tail}}$, where n_{tail} is the number of firms whose volume of sales is above the threshold for which the power law behavior holds (i.e., $S_i \geq S_{\min}$). Thus, the size of the largest firm is $S_{\max} = n_{\text{tail}}^{1/\zeta} S_{\min}$. Likewise, the size of the ith largest firm is approximately $S_i = (n_{\text{tail}}/i)^{1/\zeta}\, S_{\min}$ (see Newman (2005) for a rigorous proof). Taking logs

Figure 1: Productivity multiplier.



Notes: Panel A shows the productivity multiplier $\mu_t = \sigma_{Yt}/\sigma_{\Lambda t}$, computed using a centered rolling window of 10 years. "Unfiltered" computes the standard deviation of growth rates. "Filtered" computes the standard deviation of deviations from the Hodrick-Prescott trend of the time series in logs (the smoothing parameter is 6.25). Shaded lines indicate recession dates, defined using the Spanish Economic Association data. Panel B shows the relative standard deviations calculated using the time period 1994-2018, the average "unfiltered" productivity multiplier and the average "filtered" productivity multiplier. $\hat{\mu}_{\mathrm{GIV}}$ refers to the estimated coefficient in (15) when the granular residual (20) is calculated using the top $K = 1, 2, \ldots, Q$ firms and $K = Q$, where $Q = 500, 1000, \ldots, 10,000$.

and rearranging, the "Zipf" plot for the power law distribution is characterized by

$$\ln i = c - \zeta \ln S_i, \tag{16}$$

where i is the rank of firm i and $c \equiv \ln n_{\mathrm{tail}} + \zeta \ln S_{\min}$. According to Equation (16), if the upper tail of the firm size distribution is power law, then the log-log plot should display a straight line.

A popular way to estimate the tail index ζ is to run an OLS using (16) as the econometric specification. However, Gabaix and Ibragimov (2011) show that this method, known as "log-log rank-size regression", delivers strongly biased estimates in small samples, and suggest the following modification:

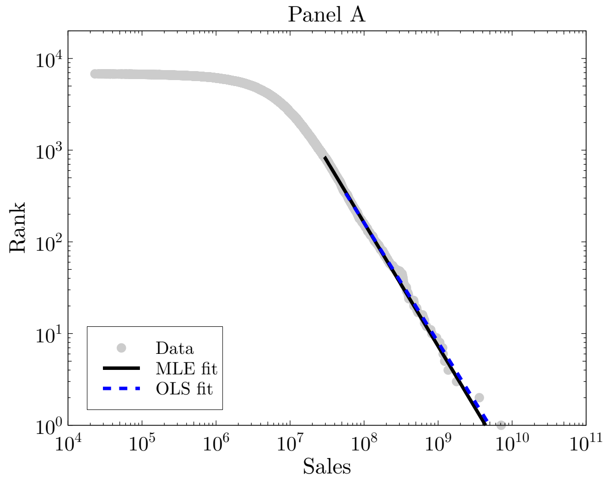
$$\ln(i - 1/2) = c - \hat{\zeta}_{\mathrm{OLS}} \ln S_i + u_i, \tag{17}$$

with asymptotic standard error $\hat{\zeta}^{\mathrm{OLS}}\sqrt{2/n_{\mathrm{tail}}}$. We estimate specification (17) using two different cut-off points. First, we take the tail that corresponds to 5% of the samples in each year. Note that the size of the tail is arbitrarily chosen following the literature.[23] Second, we take the tail that corresponds to those values of sales above the top Q largest firm. That is, we set $S_{\min} = S_Q$, and, hence, $n_{\mathrm{tail}} = Q$.

Although the log-log rank-size regression method is commonly used in the literature, it has numerous pitfalls (see Clauset et al. (2009) for a detailed explanation). As a cross-check, we also calculate the tail exponent from the density associated to the CCDF (4) by using *maximum likelihood estimation* (MLE). The

[23] It is also a standard approach in the literature to determine the threshold through visual inspection of the empirical distribution. If the distribution has a truncation point, then the threshold is typically set equal to the truncation point. As yet another alternative, we use this approach and find that the estimates are very close to those using the 5% cut-off.

Figure 2: Firm size distribution.



Notes: Panel A and Panel B show the double logarithmic plot of rank against sales in year 1994 and 2018, respectively. "MLE fit" denotes the power-law fit from (17) when taking the tail that corresponds to those values above the endogenous threshold determined by Clauset et al. (2009) procedure. "OLS fit" denotes the power-law fit from (18) when taking the tail that corresponds to 5% of the sample.

estimator for ζ is

$$\hat{\zeta}_{\text{MLE}} = \hat{n}_{\text{tail}} \left(\sum_{i=1}^{\hat{n}_{\text{tail}}} \frac{S_i}{S_{\min}} \right)^{-1}, \tag{18}$$

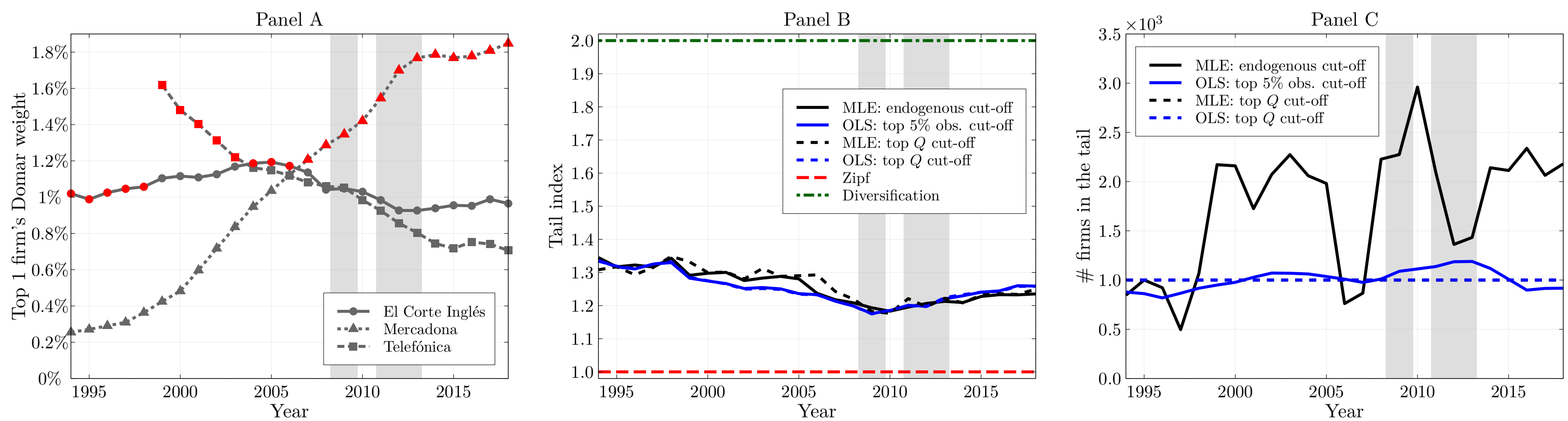
with standard error $\hat{\zeta}_{\text{MLE}}/\sqrt{\hat{n}_{\text{tail}}}$ (see Newman (2005)). We follow Clauset et al. (2009) and estimate $\hat{S}_{\min}$ as the value of sales that minimizes the distance (measured by the Kolmogorov–Smirnov statistic) between the probability distribution of size and the best-fit power-law model above $\hat{S}_{\min}$. Thereofore, $\hat{n}_{\text{tail}}$ is the number of firms whose sales are in the range $\left[\hat{S}_{\min}, S_{\max}\right]$. Additionally, we set $S_{\min} = S_Q$.

Panel A and B in Figure 2 show the empirical firm size distribution in year 1994 and 2018, respectively. Particularly, we follow the intuition provided by (16) and plot the double logarithmic plot of rank vs. sales. The distribution is characterized by a truncation point and an upper tail that displays a straight line characteristic of the power law distribution. This visual identification is confirmed by the fits provided by (17) and (18). Panel B and Panel C in Figure 3 present, respectively, the estimates for the tail index ζ and the cut-offs used in the estimation through time. Despite the fact that the sample coverage grows over time (see Panel B in Figure A.3), the estimates exhibit an almost steady value equal to 1.255 and are not sensitive to the choice of the cut-off. The average tail index is closer to Zipf (1949) law (i.e., $\zeta = 1$) than the diversification argument (i.e., $\zeta \geq 2$), which implies that the firm size distribution is sufficiently fat-tailed for idiosyncratic shocks to individual firms do not wash out at the aggregate level, because the idiosyncratic shocks to large firms do not cancel out with shocks to smaller firms (Gabaix, 2011).

As discussed by Mitzenmacher (2004) and Newman (2005), the log-normal distribution can behave as a power law.[24] As an alternative, we fit a log-normal distribution on the firm size distribution using MLE. We

[24] See Saichev et al. (2009) for a lengthy discussion on the ongoing debate between power law and log-normal in firm size distribution.

Figure 3: Heterogeneity in firm size.



Notes: Panel A shows the Domar weight of those firms that alternate in the top 1. Observations in red denote the weight of the top 1 in year t. Panel B shows the estimates for the tail index ζ using maximum likelihood estimation (MLE) and ordinary least squares (OLS). "Endogenous cut-offs" takes the tail that corresponds to those values above the endogenous threshold determined by Clauset et al. (2009) procedure. "Top 5% observations cut-off" takes the tail that corresponds to 5% of the samples in each year. "Top Q cut-off" takes the tail that corresponds to the top 1000 firms. Panel C shows the number of firms in the tail used in each estimation.

impose the same cut-offs for these estimations as in the power law estimations and perform Vuong (1989) *likelihood ratio test* $\mathcal{R}$ to compare the fits of both models.[25] The sign of $\mathcal{R}$ indicates which model is closer to the true model: if $\mathcal{R}$ is statistically greater than zero, then the test statistic presents evidence in favor of power-law model. Figure A.4 shows that the ratio alternates positive and negative values that are not statistically different from zero. Thus, we cannot conclude which candidate distribution provides a better fit. As argued in Section 2.2, we follow the bulk of the literature and assume that the underlying theoretical distribution is power law.

Finally, Panel A in Figure 2 shows the Domar weight of the largest firm through time. In our sample, three firms alternate in the top 1: El Corte Inglés (general merchandise store), Telefónica (communications) and Mercadona (food store). The fact that El Corte Íngles is the largest firm in our sample during the period 1994-1998 is consequence of the low coverage in SABI database, as discussed in Section 3.1. The reason behind the jump observed between years 1998 and 1999 is that Telefónica enters the sample 1999. In line with Gutiérrez and Philippon (2019), we find that the largest firm's Domar weight has not increased through time.[26] The relative size of Mercadona in 2018 is similar to that of Telefónica in 1999. As baseline, we assume that the average Domar weight of the largest firm (1.4%) captures the evolution in time. In Section 4.3, we relax this assumption in order to study how the granular size of the economy changes over the business cycle.

3.5 Idiosyncratic shocks volatility

We estimate the size-shock relationship using the methodology proposed by Koren and Tenreyro (2013), which allows for variation within firms.[27] The volatility of firm-level shocks $\sigma_{i\tau}$ is defined as the standard deviation of idiosyncratic shocks ε_{it} to firm i over a time block τ. The measure of size $\tilde{S}_{i\tau}$ is the average normalized sales within τ. Normalized sales are defined as $S_{it}/S_{\min,t}$, where $S_{\min,t}$ is the value of sales of the Qth firm in year t. To use every year in our sample, we calculate $\sigma_{i\tau}$ and $\tilde{S}_{i\tau}$ in a four-year time window. The sample is divided into 6 time blocks (i.e., $\tau = 6$). The econometric specification is

$$\ln \sigma_{i\tau} = \text{constant} + \alpha \ln \tilde{S}_{i\tau} + \varphi_\tau + \varphi_i + u_{i\tau}, \tag{19}$$

where φ_τ and φ_i control for time blocks and firm fixed effects, respectively. Some firms enter and leave the top Q, so they have few observations per block. To reduce the estimated volatility, we only consider those firms that have at least 3 of the 4 years that constitute a block.

Table 2 shows that the estimated size-volatility elasticity is statistically different from zero. Therefore, there are clear deviations from Gibrat's law. When we include firms fixed effects, our estimate is within the range 0.1–0.25, previously estimated in the literature (see, e.g, Stanley et al. (1996), Sutton (2002), Koren and Tenreyro (2013), Yeh (2017), Calvino et al. (2018)).

[25] We use the normalized log-likelihood ratio: $n_{\text{tail}}^{-1/2}\mathcal{R}/\sigma_{\mathcal{R}}$. The likelihood ratio is $\mathcal{R} = L(\theta_1|x)/L(\theta_2|x)$, where L is the likelihood function and θ_1 and θ_2 are, respectively, a vector of parameters for the power-law model and log-normal model. The standard deviation associated to $\mathcal{R}$ is $\sigma_{\mathcal{R}}$.

[26] Gutiérrez and Philippon (2019) find that the top 20 U.S. firms have not become larger relative to the economy. We also calculate the Domar weight of the largest firm in the U.S. using cite Gabaix (2011) data set from Compustat North America database and find that General Motors and Walmart alternate in the top 1. During the period 1951-2008, the relative size of the largest has not increased.

[27] See also Yeh (2021), who estimates more systematically the size-variance relationship using the universe of U.S. firms, and quantifies its impact on the explanatory power of the granular residual.

Table 2: Idiosyncratic shocks volatility and size.

	$\ln \sigma_{i\tau}$	
Constant	-1.996^{***}	-1.813^{***}
	(0.029)	(0.064)
$\ln \tilde{S}_{i\tau}$	-0.062^{**}	-0.150^{***}
	(0.020)	(0.053)
φ_τ	✓	✓
φ_i		✓
Observations	5,430	5,430
R^2	0.027	0.552
Number of clusters		1,861

Notes: The specifications use the four-year standard deviation of annual productivity shocks to the $Q = 1000$ largest firms in the time period 1995-2018. The size is computed at its mean value over the four-year window. Clustered (by firm) standard errors in parentheses. *, ** and *** indicate significance at 10% , 5% and 1%.

As a robustness check, Appendix C estimates the elasticity α following Stanley et al. (1996) cross-sectional methodology. For a given cross-section of idiosyncratic shocks and size binds, we calculate the standard deviation of shocks and average value of sales within each bin. Then, the elasticity is estimated by running log-log OLS regression of standard deviations on average size. The estimate is very similar to the baseline specification.

Finally, we need to estimate the value for $\sigma_{\max}$. For each firm, we average the standard deviation of shocks $\sigma_{i\tau}$ and normalized sales $\tilde{S}_{i\tau}$ used in the estimation of specification (19). As mentioned in Section 3.1, SABI has some limitations that makes firm-level volatility estimation quite volatile.[28] We choose $\sigma_{\max}$ to match a standard deviation of 11.6%, corresponding to that of average volatility of the top 30 firms in SABI. This number is comparable to the volatility of the growth rates of sales per employee ratio reported by Gabaix (2011) and is in agreement with previously reported estimates (Comin and Philippon, 2005, Davis et al., 2007, Foster et al., 2008, Haltiwanger, 2011, Bachmann and Bayer, 2014, Castro et al., 2015).

4 Quantifying the granular size of the economy

In this section, we use the conceptual framework introduced in Section 2 and the estimated parameters in Section 3 to characterize the behavior of the granular curve first observed by Blanco-Arroyo et al. (2018). In addition, we use our framework to propose a set of measures to quantify the granular size of the economy and find that approximately the top 50 Spanish firms are granular, i.e., idiosyncratic shocks to these firms may translate into aggregate fluctuations. We show that our results are robust to alternative Qs and time varying parameters. When we allow the parameters to change through time, we observe that the granular curve and, therefore, the number of granular firms changes with the business cycle. The average contribution

[28] As noted by Gabaix (2011), measuring firm volatility is also difficult because various frictions and identifying assumptions provide conflicting predictions about links between changes in total factor productivity and changes in observable quantities such as sales and employment.

of idiosyncratic shocks over the cycle coincides with that observed for the entire time period studied.

4.1 Granular curve

Building on Hulten (1978) result (see Equation (1)), Gabaix (2011) constructs the "granular residual" $\mathcal{E}_t$, which is a parsimonious measure of the idiosyncratic shocks to the top K firms:

$$\mathcal{E}_t = \sum_{i=1}^{K} \frac{S_{it-1}}{Y_{t-1}} \varepsilon_{it}, \tag{20}$$

where firm i's idiosyncratic shocks ε_{it} in year t are estimated using (12). Gabaix (2011) and the empirical literature that followed estimate the contribution of the idiosyncratic shocks to the top K firms to GDP fluctuations by regressing the growth rate of GDP g_{Yt} on the granular residual. As noted by Blanco-Arroyo et al. (2018), the estimation is based on an exogenous choice for the number of large firms. Such "pointwise" estimation does not provide information on the extent of the granular size of the economy (i.e., those top firms whose idiosyncratic shocks may translate into aggregate fluctuations), as the number of firms is arbitrarily chosen. Therefore, the contribution of the granular term to the GDP fluctuations may underestimated or overestimated depending on the choice of K. Blanco-Arroyo et al. (2018) construct (20) for $K = 1, 2, \ldots, Q$ and evaluate the behavior of R^2 as $K \to Q$. We follow this approach and estimate the R^2 as

$$R^2(K) = \mu^2 \frac{\sigma_{\mathcal{E}}^2(K)}{\sigma_Y^2}, \tag{21}$$

where $\sigma_{\mathcal{E}}^2$ is the variance of the granular residual (20) and σ_Y^2 is the variance of the growth rate of GDP per capita. Note that the behavior of R^2 will reflect only changes in $\sigma_{\mathcal{E}}^2$, as the productivity multiplier μ is held constant through K (i.e., $\mu(K) = \mu\ \forall K$). This choice is based on the stability of the multiplier to changes in K (see Panel B in Figure 1).

As argued in Section 3, our baseline case assumes that firm i's Domar weight and shock volatility is constant through time. Hence, the variance of the granular residual is

$$\sigma_{\mathcal{E}}^2 = \sum_{i=1}^{K} \left(\frac{S_i}{Y}\right)^2 \sigma_i^2, \tag{22}$$

where S_i/Y is the time average Domar weight and the variance of shock σ_i^2 is held constant through time. Section 3.4 shows that S_i is well described by a power law distribution with exponent 1.255. Thus, we can use (8) to characterize the contribution of idiosyncratic shocks to aggregate fluctuations as $K \to Q$. Replacing the total number of firms in the economy n by the largest K firms, the explanatory power of the granular residual is

$$R^2(K) = \mu^2 \left(\frac{S_{\max}}{Y}\right)^2 \left(\frac{\sigma_{\max}}{\sigma_Y}\right)^2 \frac{2}{2-\zeta'} \left[1 - \Gamma(2/\zeta')\, K^{1-2/\zeta'}\right]. \tag{23}$$

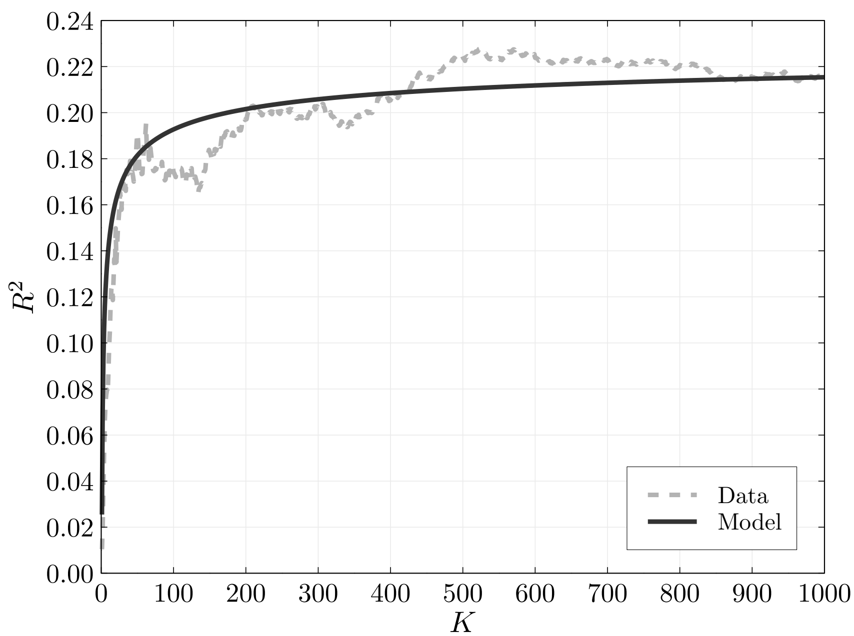
Figure 4 shows the behavior of the empirical explanatory power calculated (Equation (21)). In line with Blanco-Arroyo et al. (2018), we observe the "granular curve" behavior: a rapid increase of R^2 when a small number of top firms are included in the granular residual and slow increase after a given number of firms. We also include the predicted behavior by (23) when we use the parameters estimated in Section 3 (see

Table 3: Parameters.

Parameters	Description	Value
μ	productivity multiplier	
$S_{\max}/Y$	top 1 firm's weight	
$\sigma_{\max}$	top 1 firm's volatility	
$\sigma_{\min}$	top Q firm's volatility	
ζ	scaling parameter size	
α	scaling parameter volatility	
ρ	correlation	
N	# firms in the economy	
m_ℓ	mean of $\ln S$ ($\ln \mathcal{S}$)	
σ_ℓ	st. dev. of $\ln S$ ($\ln \mathcal{S}$)	

Notes: μ is the average value of the fraction σ_Y/σ_Λ. $S_{\max}/Y$ is the average Domar weights of top 1 firm. $\sigma_{\max}$ is the average standard deviation of labor productivity shocks among the top 30 Spanish firms. σ_Y is the standard deviation of GDP per capita growth. ζ is the average tail index.

Figure 4: Granular curve.



Notes: "Data" refers to $R^2(K)$ calculated using (21) and "Model" refers to $R^2(K)$ calculated using (23). The parameters used to calculate (23) are presented in Table 3.

Table 3). As the figure renders clear, our framework is able to characterize the dynamics of the empirical explanatory power of the granular residual.

4.2 Granular size measurement

As shown in Figure 4, the model developed in Section 2 describes well the behavior of the contribution of idiosyncratic shocks to GDP growth fluctuation. Therefore, we can employ our model to provide a set of measures that quantify the granular size of the economy, i.e., how many granular firms populate the economy. We propose three measures based on the following three definitions of granular firms:

1. Those firms whose marginal contribution is above a constant contribution.
2. Those firms that account for 75% of the maximum granular contribution.
3. Those firms whose marginal contribution is above the marginal contribution in the equally-weighted firms scenario.

To grasp the intuition of definition 1, let us focus on the top $Q = 1000$ and consider a constant contribution between the largest firm and Q. This constant contribution is captured by the secant between firm 1 and Q, which is given by

$$\mathcal{M} = \frac{R^2(Q) - R^2(1)}{Q-1},$$

where R^2 is determined by (23). Definition 1 seeks to find the number of firms whose marginal contribution to aggregate fluctuations is above the secant, that is: $\partial R^2(K)/\partial K = \mathcal{M}$. This is none other than the *mean value theorem*, which states that for a given planar arc between two endpoints, there is at least one point at which the tangent to the arc is parallel to the secant through its endpoints. According to Definition 1, the number of granular firms is given by

$$K^*_{\mathcal{M}} = \left[\frac{Q}{1-Q^{1-2/\zeta'}}\left(\frac{2}{\zeta'}-1\right)\right]^{\zeta'/2}. \tag{24}$$

Using the estimated parameters presented in Table 3, we find that $K^*_{\mathcal{M}} = 82$. The drawback of this measure is its dependence on the number of firms used to compute the granular curve Q.

Definition 2 relies on the existence of a maximum granular contribution when the underlying theoretical distribution of firm size is power law (see Equation (9)). The parameters presented in Table 3 give a maximum granular contribution of $\mathcal{A} = 23\%$. We propose an arbitrarily chosen threshold of 75% of this value. Thus, the granular firms are those firms whose accumulated contribution is equal to 17.25%. According to Definition 2 the number of granular firms is given by the following expression

$$K^*_{\mathcal{A}} = \left[\frac{1-\mathcal{T}}{\Gamma(2/\zeta')}\right]^{\zeta'/(\zeta'-2)}, \tag{25}$$

which is determined by the equation $R^2(K)/\mathcal{A} = \mathcal{T}$, where $\mathcal{T}$ is set to 0.75. We find that $K^*_{\mathcal{A}} = 36$. The drawback of this measure is the fact that depends on the exogenous threshold $\mathcal{T}$.

In the spirit of Blanco-Arroyo et al. (2018), Definition 3 uses the counterfactual in which all firms are of equal size. Let us assume a representative firm size for all firms ($S_i = \overline{S}\ \forall i$, where $\overline{S}$ is the representative size). According to Equation (5), the volatility of shocks is identical across firms (i.e., $\sigma_i = \overline{\sigma}\ \forall i$, where $\overline{\sigma}$ is the representative volatility). In this scenario, GDP growth volatility (3) becomes

$$\sigma_Y = \mu \frac{\overline{S}}{Y} \overline{\sigma} \sqrt{n}.$$

Assume that the economy is made only up of the top Q (i.e., $n = Q$), the representative size across the top Q firms necessary to match the empirical σ_Y is given by

$$\overline{S} = \frac{1}{\mu} \frac{\sigma_Y}{\overline{\sigma}} \frac{Y}{\sqrt{Q}}.$$

Plugging this size into the predicted explanatory power of the equal-weight scenario (i.e., $R^2 = \mu^2 \left(\overline{S}/Y\right)^2 \left(\overline{\sigma}/\sigma_Y\right)^2 K$) we find that, when all firms are of equal size, the explanatory power is simply the number of top K firms to total number Q of top firms ratio:

$$R^2_{\overline{S}}(K) = \frac{K}{Q},$$

where $K = 1, 2, \ldots, Q$. Definition 3 seeks to find the number of firms whose marginal contribution is above the marginal contribution in the equal-weight counterfactual. That is, $\partial R^2(K)/\partial K = \partial R^2_{\overline{S}}(K)/\partial K$. According to Definition 3, the number of granular firms is given by

$$K^*_{\overline{S}} = \left(\mu \frac{S_{\max}}{Y} \frac{\sigma_{\max}}{\sigma_Y}\right)^{\zeta'} \left[Q \Gamma\left(\frac{2}{\zeta'} + 1\right)\right]^{\zeta'/2}. \tag{26}$$

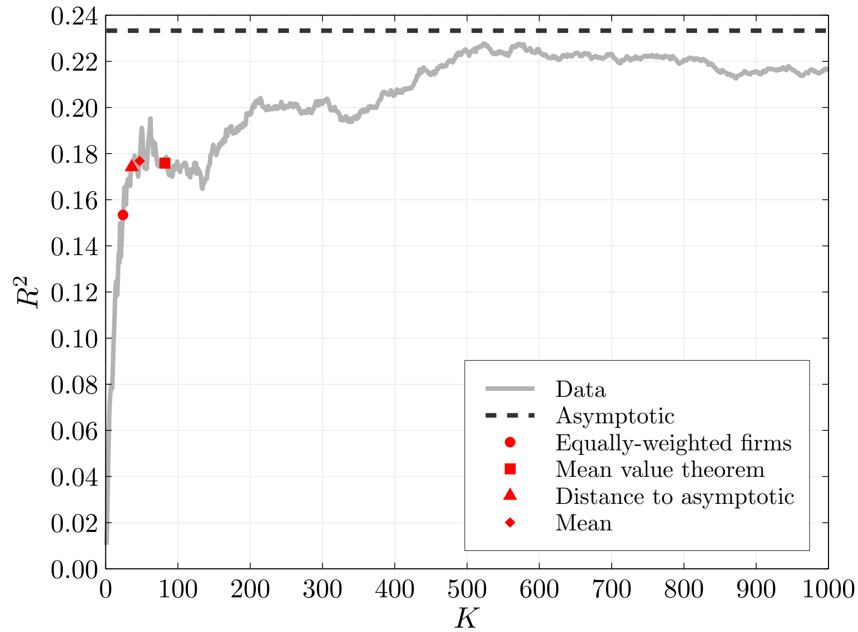
We find that $K^*_{\overline{S}} = 24$. As in Definition 1, the drawback of this measure is its dependence on the number of firms used to compute the granular curve Q.

Figure 5 shows the empirical behavior of the explanatory power of the granular residual, the maximum contribution of idiosyncratic shocks to GDP growth volatility and the calibrated granular size for each definition above. We also include the mean value of the three measures ($K^* = 47$), which appears to be closer to the point that visually represents the change from the granular to the atomistic regime. Therefore, we conclude that the granular size of the Spanish economy is approximately 50 firms. In other words, if the largest 50 firms did not exist, the Spanish economy would not be granular.

A potential concern with our baseline calibration is that two out of the three proposed measures depend on the number of firms Q. To address this concern, in Section 4.3, we increase Q from 1000 to 2500 in steps of 500 and re-estimate the granular size of the economy. We find that the mean value of the measurements is still within the region that we visually identify as the regime change

Blanco-Arroyo et al. (2018) propose a methodology to calibrate the number of granular firms that consists in gradually replacing the top firms by smaller firms. In each replacement, we compute the empirical explanatory of the granular residual as the number of top firms increases. We observe that the average explanatory power gradually decreases until it converges to a benchmark in which all firms are equally weighted. The number of granular is then the point of convergence, which is approximately 450. However, the decrease in the average explanatory power is not constant. We identify a “inner granular structure” around firm 50 that is left unexplained. The measures proposed in the present work indicate that the number of granular firms previously estimated by Blanco-Arroyo et al. (2018) is too conservative and it is the inner granular structure which determines the granular size of the economy. Hence, this result indicates

Figure 5: Granular size of the economy.



Notes: "Data" refers to $R^2(K)$ calculated using (21). "Asymptotic" refers to the maximum contribution of idiosyncratic shocks to the volatility of GDP per capita growth (see (9)). "Equally-weighted firms" refers to (26). "Mean value theorem" refers to (24). "Distance to asymptotic" refers to (25). "Mean" refers to the value resulting from averaging the three measurements.

that we should not consider the transition phase between granular and atomistic regime when quantifying the number of large firms whose idiosyncratic shocks have a non-negligible impact on the aggregate fluctuations.

4.3 Extensions

4.3.1 Alternative Qs

As argued in Section 3.2, we focus on the top $Q = 1000$ firms because the homogeneity assumption used to estimate idiosyncratic shocks is likely to be a less good approximation when taking a large Q. We now assess the robustness of our results to alternative Qs. Specifically, we reestimate Equation (12) for $Q = 1000, 1500, 2000, 2500$ and the contribution of idiosyncratic shocks to aggregate (21) as $K \to Q$. Regarding the approximation (23), the only parameters that potentially depend on Q are the volatility of the largest firms $\sigma_{\max}$ and the size-volatility elasticity α. We observe that $\sigma_{\max}$, computed as the average standard deviation of the top 30 firms remains unchanged as Q grows large. Therefore, any changes are attributable to the elasticity α. The estimation of α using the specification (19) is challenging because the introduction of a large number of small firms impacts heavily on the estimate. To attenuate this impact, we average firm i's productivity shock volatility in log $\ln \sigma_{i\tau}$ and normalized sales in log $\ln \tilde{S}_{i\tau}$ over τ time blocks, and divide them into $\mathcal{B} = 25$ bins using the average normalized sales in log. Then, we compute the average volatility σ_{B} and size $\tilde{S}_{\mathrm{B}}$ within each bin B, with $\mathrm{B} = 1, \ldots, \mathcal{B}$. Finally, we use the following specification to estimate

Table 4: Elasticity.

	σ_{B}			
	$Q = 1000$	$Q = 1500$	$Q = 2000$	$Q = 2500$
Constant	-1.933^{***}	-1.882^{***}	-1.885^{***}	-1.889^{***}
	(0.088)	(0.125)	(0.078)	(0.100)
$\tilde{S}_{\mathrm{B}}$	-0.169^{***}	-0.193^{***}	-0.165^{***}	-0.150^{***}
	(0.032)	(0.043)	(0.025)	(0.031)
Observations	25	25	25	25
R^2	0.585	0.506	0.677	0.535

Notes: The specifications use the average standard deviation and average normalized sales within each size bin. The number of observations corresponds to the number of bins. *, ** and *** indicate significance at 10% , 5% and 1%

the size-volatility elasticity:

$$\sigma_{\mathrm{B}} = \text{constant} + \alpha \tilde{S}_{\mathrm{B}} + u_{\mathrm{B}}. \tag{27}$$

Table 4 shows the estimates when Q increases from 1000 to 2500 in steps of 500 firms. The estimates are in line with those previously estimated by specification (19). We use these values to calibrate the elasticity that best captures the dynamics of the contribution of idiosyncratic shocks to large firms to aggregate fluctuation. We chose the following values: 0.15 for $Q = 1000$, 0.2 for $Q = 1500$, and 0.18 for $Q = 2000$ and $Q = 2500$. Recall that the rest of the parameters remain as presented in Table 3, as they do not depend on Q. Figure 6 shows that the behavior of R^2 is quite stable to changes in the number of large firms taken to estimate the idiosyncratic shocks. We also include the estimated granular size of the economy. In particular, we show the mean value of the three measure proposed in Section 4.2. These are $K^* = 47$ for $Q = 1000$, $K^* = 90$ for $Q = 1500$, $K^* = 85$ for $Q = 2000$ and $K^* = 97$ for $Q = 2500$. As expected, the estimated granular size grows as Q grows large, but the estimated values remain within the region that we can visually identify as the change from the granular to the atomistic regime.

4.3.2 Time-varying parameters

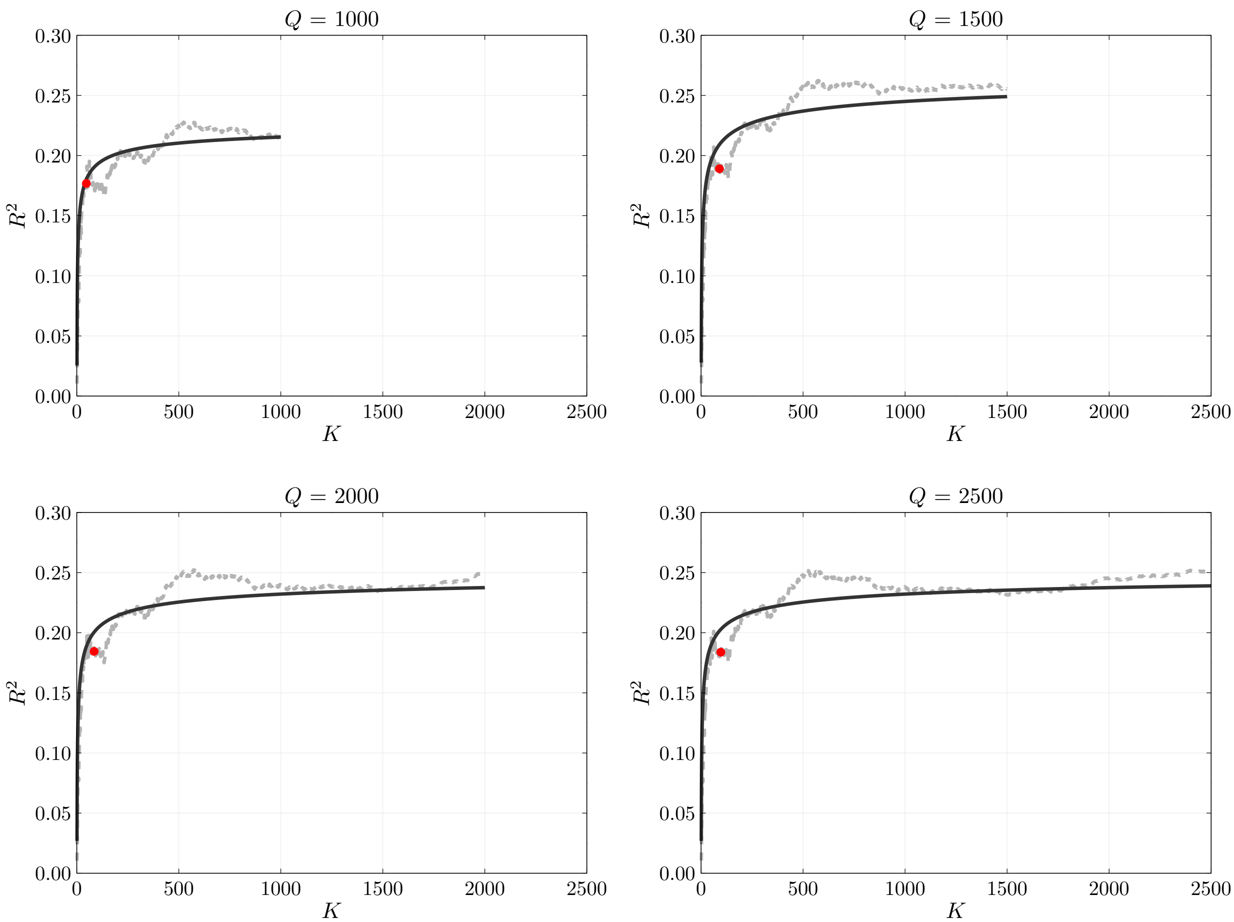
Based on the stability displayed by the share of economic activity commanded by the largest firm and the cross-sectional dispersion of firm sizes through time, our baseline specification characterizes the granular curve behavior and quantifies the granular size of the economy assuming that the determinants remain constant during entire period. We now relax this assumption to study how idiosyncratic shocks to the large firms contribute to GDP growth volatility through time.

The empirical contribution in time window τ is

$$R^2_{\tau}\left(K\right) = \mu^2 \frac{\sigma^2_{\mathcal{E}\tau}\left(K\right)}{\sigma^2_{Y\tau}}, \tag{28}$$

where $\sigma^2_{\mathcal{E}\tau}$ is the variance of the granular residual (20) in τ and $\sigma^2_{Y\tau}$ is the variance of the growth rate of GDP per capita in τ. We follow Carvalho and Gabaix (2013) and chose $\tau = 10$ years. As in (21), the evolution

Figure 6: Granular curves.



Notes: The grey dashed line is $R^2(K)$ calculated using (21). The solid black line is $R^2(K)$ calculated using (23). The red dot is the estimated granular size of the economy.

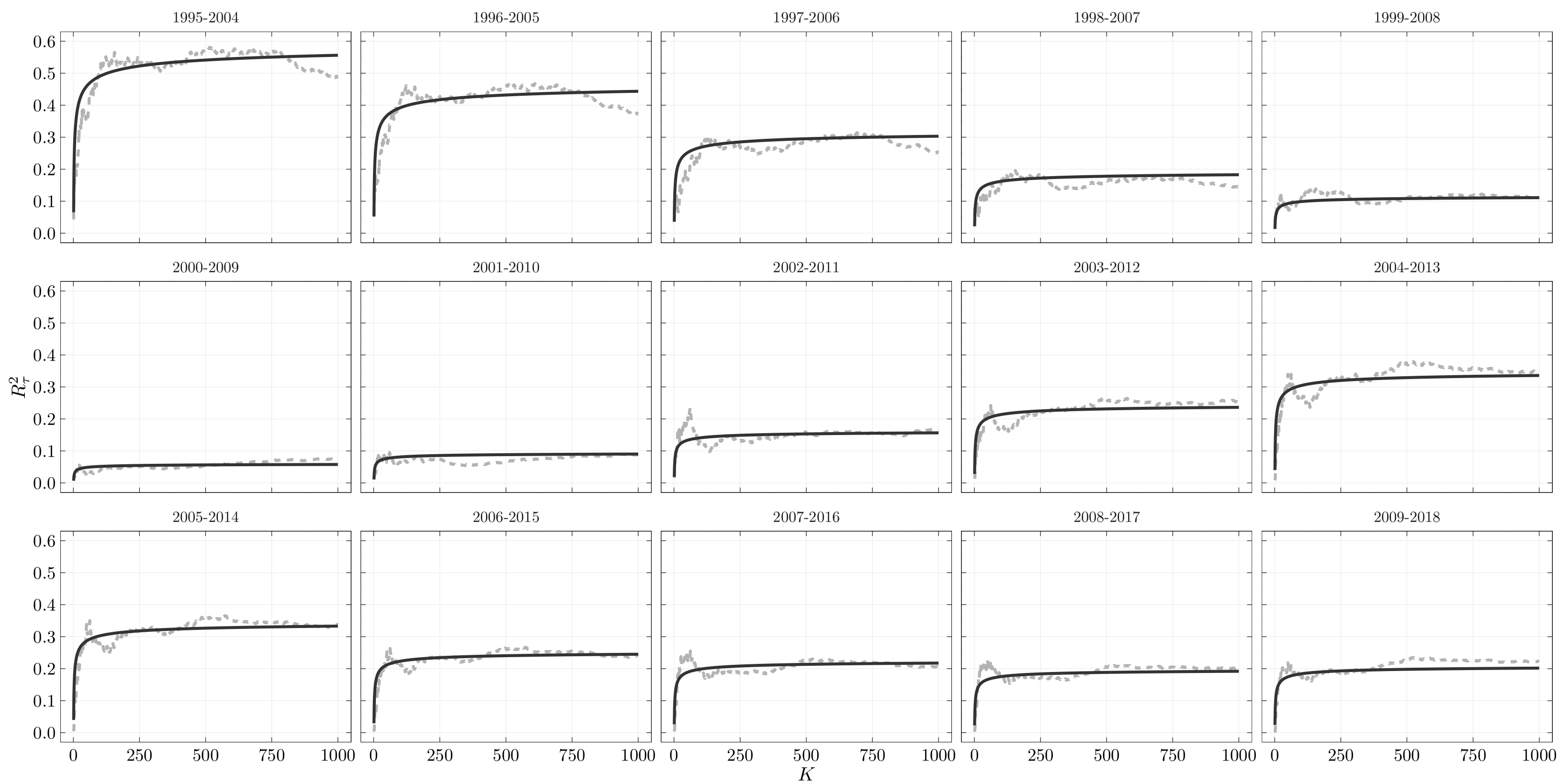
of $R^2_\tau(K)$ will reflect only changes in the relative variance, as the productivity multiplier is held constant at 3.5 through time.

The approximation of contribution (28) is

$$R^2_\tau(K) = \mu^2 \left(\frac{S_{\max,\tau}}{Y_\tau}\right)^2 \left(\frac{\sigma_{\max,\tau}}{\sigma_{Y\tau}}\right)^2 \frac{2}{2-\zeta'_\tau} \left[1 - \Gamma\left(2/\zeta'_\tau\right) K^{1-2/\zeta'_\tau}\right], \tag{29}$$

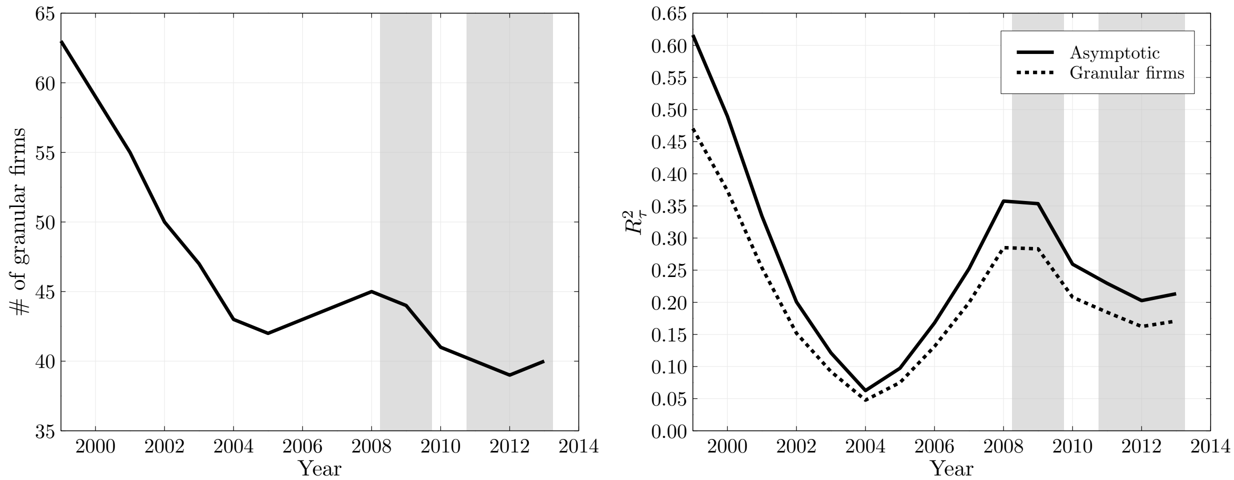
where $S_{\max,\tau}/Y_\tau$ is the average Domar weight in τ, $\sigma_{\max,\tau}$ and $\sigma_{Y\tau}$ are, respectively, the volatility of shocks to large firms and growth rate of GDP per capita in τ and $\zeta'_\tau \equiv \zeta_\tau/(1-\alpha)$. We hold the elasticity α constant at 0.15 because of limited time dimension does not allow us to use Koren and Tenreyro (2013) methodology. Finally, given the difficulty of measuring the volatility of the largest firms in the data, we choose $\sigma_{\max,\tau}$ that approximates the behavior of $R^2_\tau(K)$ in the time window τ. The values chosen range between 0.062 and 0.155, which are still within the range of estimates reported by the literature (see Section 3.5).

Figure 7: Granular curves through time.



Notes: The dashed grey line is the empirical R^2_τ (Equation (28)). The solid black line is the analytical approximation (Equation (29)). The estimation of the parameters constituting the approximation is explained in the main text. Each time window τ consist in 10 years.

Figure 8: Granular size and its contribution over time.



Notes: Panel A shows the average value of the granular sizes resulting from the proposed measures. Panel B shows the maximum contribution of idiosyncratic shocks to volatility of GDP growth (Equation (9)) and the contribution of granular firms. Year is the centered year in the window τ.

Figure 7 plots the empirical contribution of idiosyncratic shocks to GDP fluctuations (Equation (28)) and its analytical approximation (Equation (29)). Two results are worth noting. First, the empirical contribution exhibits the granular curve behavior in all the time windows in which the sample is divided. Second, the granular curve exhibits a cyclical behavior: the contribution of idiosyncratic shocks to GDP fluctuations shrinks when taking into account recession years and grows in expansion years. During the period under study, the dynamics of the large firms play a crucial role in shaping aggregate fluctuations in times of relative stability. Nevertheless, when taking into account the years in which the financial crisis (an exogenous shock) hit the Spanish economy, the impact of the dynamics of large firms at the aggregate level becomes almost negligible. In line with this intuition, Figure 8 shows that the calibrated number of granular firms (Panel A) and its contribution to aggregate fluctuations (Panel B) increased during the Spanish economic boom and decreased during the burst of the housing bubble.

5 Conclusion

The emergent literature on the granular origins of aggregate fluctuations challenges the tradition in macroeconomics by arguing that, in the presence of significant heterogeneity at the firm level, idiosyncratic shocks to the granular (large) firms do not cancel out with shocks to smaller firms and, thus, translate into aggregate fluctuations. The literature quantifies the contribution of idiosyncratic shocks to aggregate fluctuations using an exogenous given number of large firms, which does not provide information on the granular size of the economy, namely, the number of granular firms. We provide a conceptual to quantify the number granular size of an economy.

The first part of our analysis characterizes, analytically, the contribution of idiosyncratic shocks to the large firms to GDP growth volatility and shows that it is driven by the share of economic activity commanded by the largest firm, the volatility of the largest firm with respect to GDP volatility and two summary statistics for large firm dynamics: the tail index of firm size distribution and the size-volatility elasticity. Additionally, we show that changes size-volatility relationship have greater impact on aggregate fluctuations than tail index changes, which have been the main object of study.

In the second part of the paper, we show that the granular curve is well characterized by our framework and find that the average maximum contribution of top Spanish firms's idiosyncratic shocks to the GDP fluctuations is 23%. We estimate that the granular size of the Spanish economy is approximately 50 firms. Finally, we find that the granular size exhibits a cyclical behavior: the number of firms whose idiosyncratic shocks have an impact on the aggregate grows in expansion phases and shrinks in recession phases.

In future research, we plan to extend our framework to other countries and to include amplification mechanisms of shocks such as the input-output network. These two additional dimensions could provide greater insights on the observed differences in output volatility across countries.

Appendix A Proof of Proposition 1

We first plug the relationship between size and shock volatility (5) into the variance of the granular residual (i.e., the squared granular volatility (3)). Rearranging, we have that

$$\sigma_Y^2 = \mu^2 \frac{S_{\min}^{2\alpha}}{Y^2} \sigma_{\min}^2 \sum_{i=1}^{n} \mathcal{S}_i^2,$$

where $\mathcal{S}_i \equiv S_i^{1-\alpha}$. If S_i is drawn from power law distribution (4), then $\mathcal{S}_i$ is a power law with exponent $\zeta' \equiv \zeta / (1-\alpha)$.[29] Zaliapin et al. (2005) show that the sum of *i.i.d.* power-law summands can be replaced by the maximum summand. If $\zeta' \neq 1$, then

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathcal{S}_i\right] = \mathcal{S}_{\max} \frac{1}{1-\zeta'} \left[1 - nB\left(n, 1/\zeta'\right)\right] \cong \mathcal{S}_{\max} \frac{1}{1-\zeta'} \left[1 - \Gamma\left(1/\zeta'\right) n^{1-1/\zeta'}\right],$$

where $B(\cdot,\cdot)$ and $\Gamma(\cdot)$ are, respectively, the Beta and Gamma distributions.[30] In our case, we use the following approximation:

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathcal{S}_i^2\right] = \mathcal{S}_{\max}^2 \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) n^{1-2/\zeta'}\right].$$

Then, the variance of GDP growth is

$$\sigma_Y^2 = \mu^2 \left(\frac{S_{\max}}{Y}\right)^2 \sigma_{\min}^2 \left(\frac{S_{\min}}{S_{\max}}\right)^{2\alpha} \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) n^{1-2/\zeta'}\right]. \tag{A.1}$$

[29] Recall that the CCDF (4) corresponds to a density $p(S) = \zeta S_{\min}^{\zeta} S^{-(\zeta+1)}$, then the density of $\mathcal{S}$ is

$$p(\mathcal{S}) = \frac{p(S)}{d\mathcal{S}/dS} = \frac{\zeta S_{\min}^{\zeta}}{(1-\alpha) S^{1+\zeta-\alpha}} = \frac{\zeta \mathcal{S}_{\min}^{\zeta/(1-\alpha)}}{(1-\alpha) \mathcal{S}^{(1+\zeta-\alpha)/(1-\alpha)}} = \frac{\zeta' \mathcal{S}_{\min}^{\zeta'}}{\mathcal{S}^{1+\zeta'}}.$$

[30] The approximation $B(n, 1/\zeta') \sim \Gamma(1/\zeta') n^{-1/\zeta'}$ is valid because n is large and $1/\zeta'$ is a constant.

We use the relationship (5) and define $\sigma_{\max} \equiv \sigma_{\min} \left(S_{\min}/S_{\max} \right)^{\alpha}$. Equation (6) is the square root of (A.1) with $\sigma_{\max}$.

In the case of $\zeta' = 2$, the expression

$$\frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) n^{1-2/\zeta'} \right]$$

can be approximated using the Euler–Mascheroni constant γ, which is defined as:

$$\gamma = \lim_{x \to 0} \left[\frac{1}{x} - \Gamma\left(x\right) \right].$$

To see this, let $2/\zeta' = 1 + x$, then

$$\lim_{\zeta' \to 2^-} \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) \right] = \lim_{x \to 0^-} \left(1+x\right) \left[\frac{1}{x} - \Gamma\left(x\right) \right] = \gamma.$$

And let $2/\zeta' = 1 - x$, then

$$\lim_{\zeta' \to 2^+} \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) \right] = \lim_{x \to 0^+} \left[\frac{1}{x} - \Gamma\left(x\right) \right] = \gamma.$$

Alternatively, $\zeta' = 1$ only if $\zeta = 1$ and $\alpha = 0$. Thus, Gibrat's law holds true, and the behavior of the volatility of shocks is well characterized by the average volatility $\overline{\sigma}$. In this case, the variance of GDP growth is

$$\begin{aligned}
\sigma_Y^2 &= \mu^2 \frac{1}{Y^2} \overline{\sigma}^2 \sum_{i=1}^{n} S_i^2 \\
&= \mu^2 \left(\frac{S_{\max}}{Y} \right)^2 \overline{\sigma}^2 \sum_{i=1}^{n} \frac{1}{i^2} \\
&= \mu^2 \left(\frac{S_{\max}}{Y} \right)^2 \overline{\sigma}^2 \frac{\pi^2}{6}.
\end{aligned}$$

Equation (7) is the square root of the above expression.

Appendix B Proof of Proposition 2

We cannot assess de impact of δ and α on R^2 using (8) because we have to take into their impact on $S_{\max}$ and $\sigma_{\max}$ first. As discussed in Section 3.4, $S_{\max} = n^{1/(1+\delta)} S_{\min}$, thus an increase in δ decreases $S_{\max}$, *ceteris paribus*. Using the definition of $\sigma_{\max}$ and the relationship between δ and $S_{\max}$, we have that $\sigma_{\max} = \sigma_{\min} n^{-\alpha/(1+\delta)}$. Therefore, *ceteris paribus*, $\sigma_{\max}$ increases when δ increases and decreases when α increases. Plugging these expressions into (A.1) and rearranging, we get that

$$R^2 = \mu^2 \left(\frac{S_{\min}}{Y} \right)^2 \left(\frac{\sigma_{\min}}{\sigma_Y} \right)^2 n^{2/\zeta'} \frac{2}{2-\zeta'} \left[1 - \Gamma\left(2/\zeta'\right) n^{1-2/\zeta'} \right].$$

The impact of δ and α on R^2 is then

$$\frac{\partial R^2}{\partial \delta} = \frac{\partial \zeta'}{\partial \delta}\frac{\partial R^2}{\partial \zeta'} = \frac{1}{1-\alpha}\frac{\partial R^2}{\partial \zeta'} = \frac{\zeta'}{1+\delta}\frac{\partial R^2}{\partial \zeta'} < 0$$
$$\frac{\partial R^2}{\partial \alpha} = \frac{\partial \zeta'}{\partial \alpha}\frac{\partial R^2}{\partial \zeta'} = \frac{1+\delta}{(1-\alpha)^2}\frac{\partial R^2}{\partial \zeta'} = \frac{\zeta'}{1-\alpha}\frac{\partial R^2}{\partial \zeta'} < 0,$$

where

$$\frac{\partial R^2}{\partial \zeta'} = \mu^2 \left(\frac{S_{\min}}{Y}\right)^2 \left(\frac{\sigma_{\min}}{\sigma_Y}\right)^2 \frac{2}{\zeta'^2 (2-\zeta')^2} \left\{ n^{2/\zeta'} \left[\zeta'^2 + 2(2-\zeta')\ln n\right] - n\Gamma(2/\zeta')\left[\zeta'^2 + 2(2-\zeta')\frac{\Gamma'(2/\zeta')}{\Gamma(2/\zeta')}\right]\right\} < 0$$

As one would expect, the more homogeneity in the firm size distribution (larger δ) and the more inelastic is volatility to size (larger α), the lower the contribution of idiosyncratic shocks to GDP fluctuation (smaller R^2). The ratio between the two expressions is the elasticity (11).

Appendix C Alternative construction of the granular residual

In our baseline specification (see Section 3.2), firm-level labor productivity growth rates are defined as yearly natural log differences. Due to the existence of mergers, acquisitions or measurement errors, we observe some large jumps. We follow the convention in the literature and attenuate the impact of these outliers by winsorizing them. This technique has recently been criticized by Dosi et al. (2018), who argue that supply-driven granular shocks play no role when this cleaning procedure is not carried out. In this section, we show that our results are robust to the data cleaning strategy.

C.1 Alternative productivity growth rates

As an alternative approach, we now calculate the firm-level labor productivity growth rates using the arc-elasticity adopted by Davis et al. (1996):

$$g'_{it} \equiv 2\left(\frac{z_{it} - z_{it-1}}{z_{it} + z_{it-1}}\right). \tag{C.1}$$

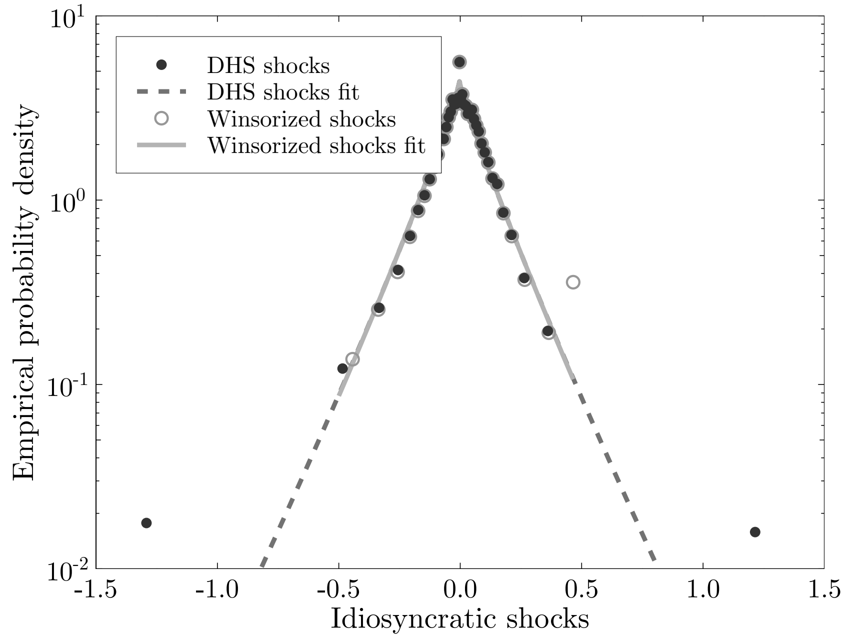
That is, the denominator is the average of the beginning and end period levels, rather than the beginning period level. This growth rate, which we label DHS, has two main advantages compared with the log difference. First, it ranges from -2 to 2 and thus limits the impact of outliers. Second, it avoids pitfalls associated with temporary transitory shocks and measurement errors (Neumark et al., 2011).

The idiosyncratic productivity shocks are estimated as in (12), namely the deviation of g'_{it} from the common shock $\eta'_t(Q)$:

$$\varepsilon'_{it} = g'_{it} - \eta'_t(\mathcal{Q}). \tag{C.2}$$

where the common shock to the top $Q = 1000$ firms is estimated as the median productivity growth rate for Spain. The key difference with respect to (12) is that shocks are already bounded. Thus, these estimated

Figure A.1: Empirical probability density of idiosyncratic productivity shocks.



Notes: Pooled empirical densities on semi-log scale of idiosyncratic productivity shocks to the top $Q = 1000$ largest firms in Spain during the period 1995-2018. Winsorized shocks and DHS shocks refer to the estimated idiosyncratic shocks according to Equation (12) (see Section 3.2) and (C.2), respectively. The solid and dashed lines show the exponential power distribution fit (C.3) obtained by maximum likelihood estimation of the scale (a), shape (b) and location (m) parameters. The resulting estimates are shown in Table 2.

shocks do not present extreme values that need to be winsorized to an exogenously determined threshold, such us 50% in Section 3.2.

We now assess the existing differences between specification (12) and (C.2) by constructing the empirical probability density of the idiosyncratic shocks. Figure A.1 presents the pooled empirical densities of idiosyncratic productivity shocks to the top Q largest firms. We show the pooled densities rather than the year-by-year densities because they are quite stable through time (see Figure A.5). They exhibit a markedly "tent-shape" form on semi-log scale. This is a well-known behavior of the distribution of firm size growth rates since the seminal work of Stanley et al. (1996). They found that the distribution of the growth rates of sales and employees is well characterized by an *Laplace distribution*. Delli Gatti et al. (2005) show that this behavior is caused by the fact that both variables exhibit a power-law behavior. In particular, they show that when the logarithm of a power-law random variable follows an exponential distribution, the difference of two exponential random variables becomes a Laplace distribution. Therefore, given that our measure of labor productivity is the sales per employee ratio, it is not surprising that the distributions of the productivity growth rate and idiosyncratic shock display also such behavior.[31]

[31] Recall that firm i's labor productivity growth rate is simply the difference between i's sales and employees growth rates: $g_{it} = \Delta \ln z_{it} = g_{it}^{\text{sales}} - g_{it}^{\text{employees}}$, where $g_{it}^{\text{sales}} = \Delta \ln \text{sales}_{it}$ and $g_{it}^{\text{employees}} = \Delta \ln \text{employees}_{it}$. If both g_{it}^{sales} and $g_{it}^{\text{employees}}$ are distributed as a Laplace, then so is g_{it}. In addition, if $\varepsilon_{it} = g_{it} - \eta_t$, where the shock η_t is common to all firms in year t, then ε_{it} is also distributed as a Laplace.

Table A.1: Maximum likelihood estimates of the exponential power distribution.

	ε	ε'
a	0.122^{***}	0.122^{***}
	(0.004)	(0.004)
b	0.893^{***}	0.886^{***}
	(0.069)	(0.067)
m	-0.003^{**}	-0.003^{**}
	(0.001)	(0.001)

Notes: Maximum likelihood estimates of the exponential power distribution (C.3). Winsorized shocks ε are estimated using (12) and DHS shocks ε' are estimated using (C.2). Standard errors in parentheses. *, ** and *** indicate significance at 10% , 5% and 1%.

In the literature that studies the empirical distribution of growth rates (see, e.g., Bottazzi et al. (2002), Bottazzi and Secchi (2006), Bottazzi et al. (2011)), it is standard to use the *exponential power distribution*, also known as *Subbotin distribution* (Subbotin, 1923):

$$f_S\left(x; a, b, m\right) = \frac{1}{2ab^{1/b}\Gamma\left(1 + 1/b\right)} \exp\left(-\frac{1}{b}\left|\frac{x - m}{a}\right|^b\right), \tag{C.3}$$

where a, $b \in \mathbb{R}^+$, $m \in \mathbb{R}$, $\Gamma\left(\cdot\right)$ denotes the Gamma function and $x \in \{\varepsilon_{it}, \varepsilon'_{it}\}$. The distribution is characterized by the scale parameter a, the shape parameter b and the location parameter m. It comprises the Laplace ($b = 1$) and the normal ($b = 2$) distributions as special cases. As expected, the empirical distributions are well approximated by (C.3) (see Figure A.1). The resulting estimates are shown in Table A.1, along with the corresponding standard errors. Notice that all parameters are virtually unaffected by winsorization. Contrary to Dosi et al. (2018), the distribution of productivity shocks remains fat-tailed after the winsorizing procedure.

The fact that the distribution of the idiosyncratic shocks is fat-tailed is in direct contradiction with the prediction of Gibrat's law that the distribution should be normal. Therefore, models that consider Gibrat's law as a baseline (e.g., Gabaix (2011), di Giovanni and Levchenko (2012) and Carvalho and Grassi (2019)) not only omit the negative relationship between size and volatility but also implicitly impose a degree of homogeneity in the size of idiosyncratic shocks to large firms that is at odds with the piece of evidence presented.

C.2 Alternative estimation of the size-volatility relationship

Before assessing the potential impact of the winsorizing process on the granular curve, we re-estimate the relationship between size and the volatility of DHS shocks using our baseline methodology (see Section 3.5). Table A.2 presents the estimates for the elasticity α. Compared to our baseline estimation, these results show that the deviation from Gibrat's law is even greater when we do not resort to the winsorization process to handle outliers. Yeh (2019) also finds that the relationship between size and the volatility of growth rates becomes steeper when using DHS growth rates rather than log-difference growth rates.

Table A.2: Alternative idiosyncratic shocks volatility and size.

	$\ln \sigma'_{i\tau}$	
Constant	-2.000^{***}	-1.792^{***}
	(0.032)	(0.062)
$\ln \tilde{S}_{i\tau}$	-0.037^{*}	-0.164^{***}
	0.022	0.062
φ_τ	✓	✓
φ_i		✓
Observations	5,435	5,435
R^2	0.028	0.557
Number of clusters		1,870

Notes: The specifications use the four-year standard deviation of annual productivity shocks, calculated using DHS growth rates (C.2), to the $Q = 1000$ largest firms in the time period 1995-2018. The size is computed at its mean value over the four-year window. Clustered (by firm) standard errors in parentheses. *, ** and *** indicate significance at 10% , 5% and 1%.

As an alternative to the Koren and Tenreyro (2013)'s methodology, we also estimate the relationship using the cross-sectional methodology employed by Stanley et al. (1996) and Sutton (2002). Although this methodology captures the degree of dispersion in idiosyncratic shocks rather than firm's shock volatility over time, we decide to take it into account because it has long been used by the literature. The procedure is as follows. First, we pool the normalized sales (i.e., $S_{it}/S_{\min,t}$) and productivity shocks to the top $Q = 1000$ firms over time and divide them into $\mathcal{B} = 16$ bins using normalized sales. Second, we fit the empirical distribution of shocks in each bin B, with $\mathrm{B} = 1, \dots, \mathcal{B}$, using the distribution (C.3). Thus, we have an estimate for the scale parameter a_{B}, shape parameter b_{B} and location parameter m_{B} for each bin B. We compute the cross-sectional standard deviation[32]

$$\sigma_{\mathrm{B}} = a_{\mathrm{B}} b_{\mathrm{B}}^{1/b_{\mathrm{B}}} \frac{\sqrt{\Gamma(3/b_{\mathrm{B}})}}{\sqrt{\Gamma(1/b_{\mathrm{B}})}}$$

and the average normalized sales $\tilde{S}_{\mathrm{B}}$ within each bin. Finally, we run the following OLS regression:

$$\ln \sigma_{\mathrm{B}} = \text{constant} + \alpha \ln \tilde{S}_{\mathrm{B}} + u_{\mathrm{B}}. \tag{C.4}$$

The estimated coefficient from (C.4) reflects the relationship between size and dispersion. As noted by Thesmar and Thoenig (2011), the main disadvantage of using cross-sectional dispersion is that it does not remove the average growth rate of the firm and, hence, it does not eliminate the bias in the evolution of firm volatility caused by a change in the distribution of firm's growth potential. However, the results are similar to the baseline specification (see Table A.3).

[32] Due to its symmetry, the Subbotin density has all central moments of odd order equal to zero. The central moment of order $2l$ reads

$$M_{2l} = \left(ab^{1/b}\right)^{2l} \frac{\Gamma((2l+1)/b)}{\Gamma(1/b)}.$$

Table A.3: Idiosyncratic shocks dispersion and size.

	$\ln \sigma_{\mathrm{B}}$			
	Subbotin fit		Standard deviation	
	Winsorized	DHS	Winsorized	DHS
Constant	-1.633^{***}	-1.550^{***}	-1.418^{***}	-0.983^{***}
	(0.107)	(0.267)	(0.129)	(0.186)
$\ln \tilde{S}_{\mathrm{B}}$	-0.102^{***}	-0.174^{*}	-0.187^{***}	-0.247^{***}
	(0.033)	(0.083)	(0.040)	(0.058)
Observations	16	16	16	16
R^2	0.401	0.239	0.608	0.566

Notes: Winsorized refers to growth rates calculated using (12). DHS refers to growth rates calculated using (C.2). Subbotin fit estimates the shocks volatility using (C.3). Standard deviation estimates the shocks volatility using the standard deviation of the shocks within bin B. The number of observations corresponds to number of bins $\mathcal{B}$. Standard errors in parentheses. *, ** and *** indicate significance at 10% , 5% and 1%.

C.3 Alternative granular residual

Under DHS definition of growth rates (C.1), the correct weights for aggregation are

$$w'_{it} \equiv \frac{S_{it} + S_{it-1}}{Y_t + Y_{it-1}}, \tag{C.5}$$

the granular residual is

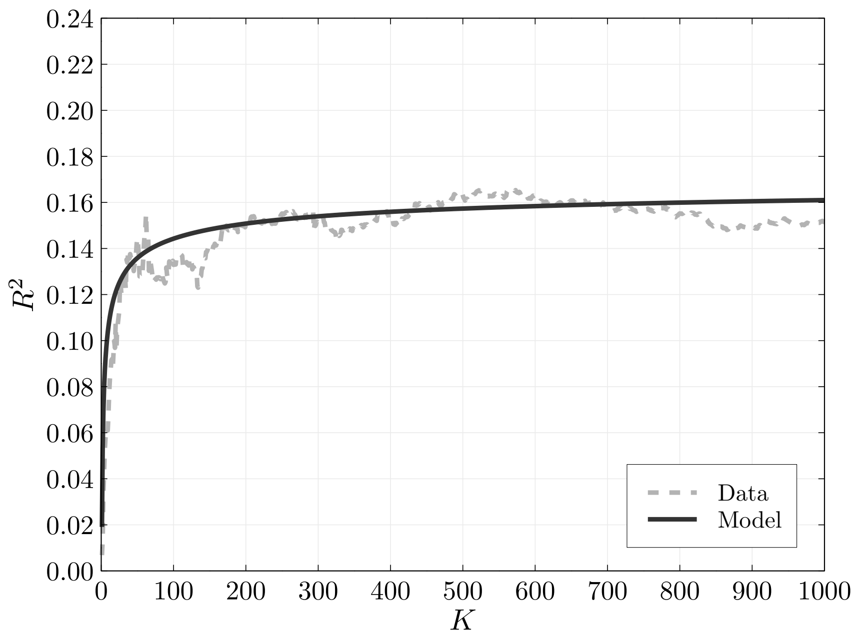
$$\mathcal{E}'_t = \sum_{i=1}^{K} w'_{it} \varepsilon'_{it} \tag{C.6}$$

and GDP growth is

$$g'_Y \equiv 2\left(\frac{Y_t - Y_{t-1}}{Y_t + Y_{t-1}}\right). \tag{C.7}$$

Following the procedure described in Section 4, we compute the empirical contribution of idiosyncratic shocks to GDP growth fluctuations, its approximation and calibrate de granular size of the economy. Figure A.2 shows that the granular curve behavior is still present and the approximation characterizes such behavior. The calibrated number of granular firms is 45.

Figure A.2: Granular curve with alternative shocks.



Notes: The parameters used in the approximation are: $\mu = 3.5$, $S_{\max/Y} = 0.0138$, $\sigma_{\max} = 0.102$, $\sigma_Y = 0.023$, $\zeta = 1.252$ and $\alpha = 0.15$.

Appendix D Additional figures

Figure A.3: Sample characteristics.

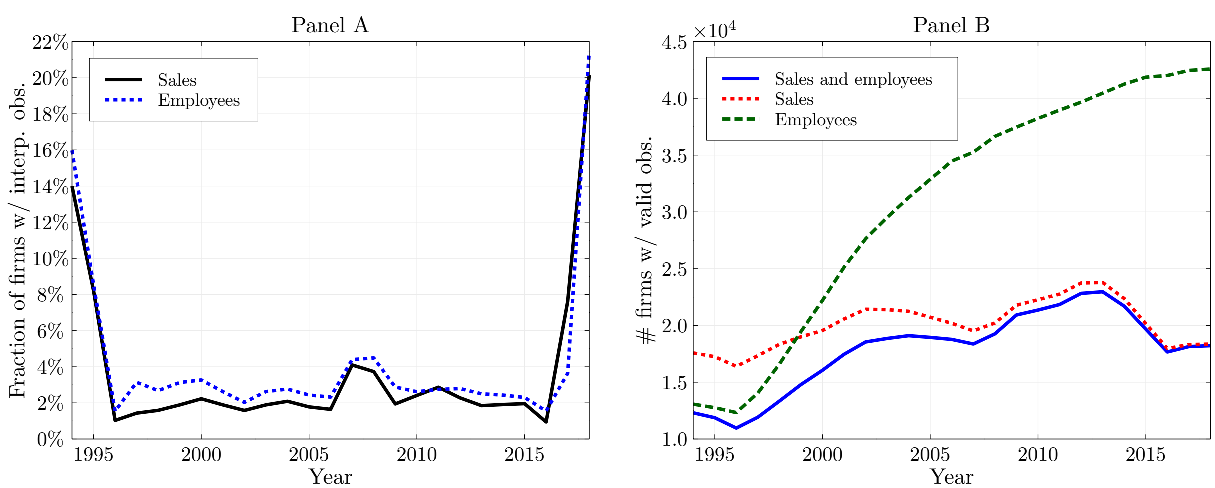
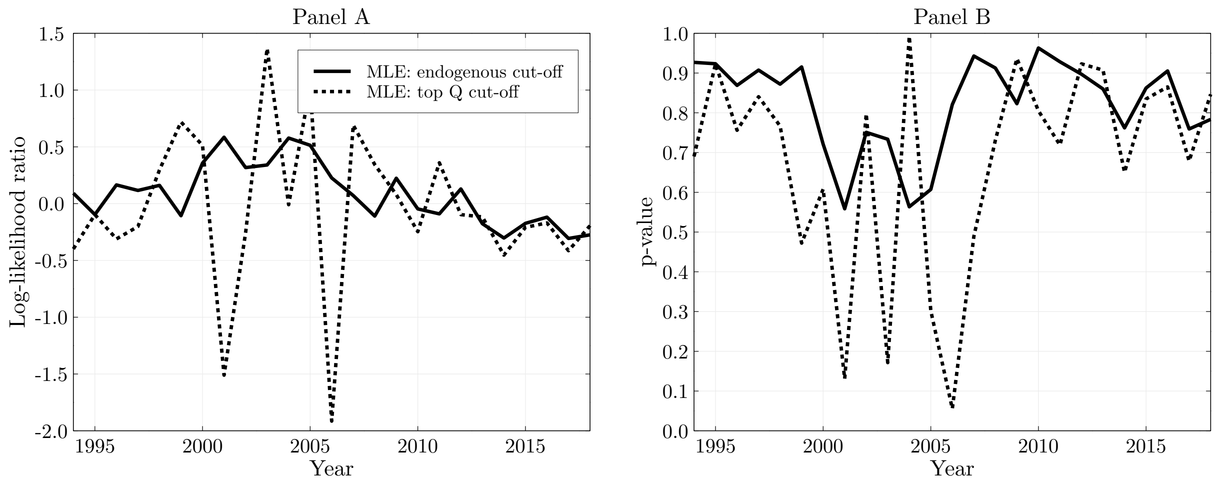
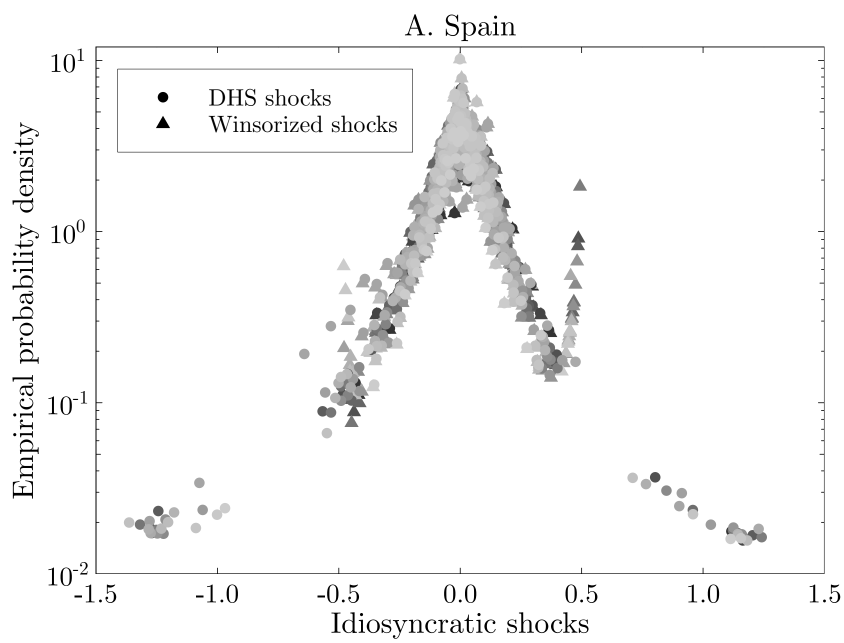


Figure A.4: Comparison of candidate distributions.



Notes: Panel A compares the power-law and log-normal candidate models using Vuong (1989) normalized likelihood ratio $\mathcal{R}$. We fit both models on the firm size distribution using MLE. Panel B shows the associated p-value.

Figure A.5: Year-by-year empirical probability density of idiosyncratic productivity shocks.



Notes: Year-by-year empirical densities on semi-log scale of idiosyncratic productivity shocks to the top $Q = 1000$ largest firms in Spain during the period 1995-2018. Winsorized shocks and DHS shocks refer to the estimated idiosyncratic shocks according to (12) (see Section 3.2) and (C.2), respectively.

References

Acemoglu, Daron, Vasco Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Amaral, Luis A Nunes, Sergey V Buldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael A Salinger, H Eugene Stanley, and Michael HR Stanley, "Scaling behavior in economics: I. Empirical results for company growth," *Journal de Physique I*, 1997, *7* (4), 621–633.

Axtell, Robert L, "Zipf distribution of US firm sizes," *Science*, 2001, *293* (5536), 1818–1820.

Bachmann, Rüdiger and Christian Bayer, "Investment dispersion and the business cycle," *American Economic Review*, 2014, *104* (4), 1392–1416.

Bak, Per, Kan Chen, José Scheinkman, and Michael Woodford, "Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics," *Ricerche Economiche*, 1993, *47* (1), 3–30.

Baqaee, David Rezza, "Cascading failures in production networks," *Econometrica*, 2018, *86* (5), 1819–1838.

_ **and Emmanuel Farhi**, "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

_ **and** _ , "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 2020, *135* (1), 105–163.

Blanco-Arroyo, Omar, Alba Ruiz-Buforn, David Vidal-Tomás, and Simone Alfarano, "On the determination of the granular size of the economy," *Economics Letters*, 2018, *173*, 35–38.

Bottazzi, Giulio, Alex Coad, Nadia Jacoby, and Angelo Secchi, "Corporate growth and industrial dynamics: Evidence from French manufacturing," *Applied Economics*, 2011, *43* (1), 103–116.

_ **and Angelo Secchi**, "Explaining the distribution of firm growth rates," *The RAND Journal of Economics*, 2006, *37* (2), 235–256.

_ **, Elena Cefis, and Giovanni Dosi**, "Corporate growth and industrial structures: some evidence from the Italian manufacturing industry," *Industrial and Corporate Change*, 2002, *11* (4), 705–723.

Calvino, Flavio, Chiara Criscuolo, Carlo Menon, and Angelo Secchi, "Growth volatility and size: A firm-level study," *Journal of Economic Dynamics and Control*, 2018, *90*, 390–407.

Carvalho, Vasco and Basile Grassi, "Large firm dynamics and the business cycle," *American Economic Review*, 2019, *109* (4), 1375–1425.

_ **and Xavier Gabaix**, "The great diversification and its undoing," *American Economic Review*, 2013, *103* (5), 1697–1727.

Castro, Rui, Gian Luca Clementi, and Yoonsoo Lee, "Cross sectoral variation in the volatility of plant level idiosyncratic shocks," *The Journal of Industrial Economics*, 2015, *63* (1), 1–29.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman, "Power-law distributions in empirical data," *SIAM review*, 2009, *51* (4), 661–703.

Comin, Diego and Sunil Mulani, "Diverging trends in aggregate and firm volatility," *The Review of Economics and Statistics*, 2006, *88* (2), 374–383.

_ **and Thomas Philippon**, "The rise in firm-level volatility: Causes and consequences," *NBER Macroeconomics Annual*, 2005, *20*, 167–201.

Cravino, Javier and Andrei A Levchenko, "Multinational firms and international business cycle transmission," *The Quarterly Journal of Economics*, 2017, *132* (2), 921–962.

Daniele, Federica and Heiko Stüber, "The Micro-Origins of Business Cycles: Evidence from German Metropolitan Areas," *The Review of Economics and Statistics*, 2020, pp. 1–48.

Davis, Steven J, John C Haltiwanger, and Scott Schuh, *Job creation and destruction*, MIT Press, 1996.

_ , John Haltiwanger, Ron Jarmin, and Javier Miranda, "Volatility and dispersion in business growth rates: Publicly traded versus privately held firms," *NBER Macroeconomics Annual*, 2007, *21*, 107–179.

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda, "Changing business dynamism and productivity: Shocks versus responsiveness," *American Economic Review*, 2020, *110* (12), 3952–90.

Delli Gatti, Domenico, Corrado Di Guilmi, Edoardo Gaffeo, Gianfranco Giulioni, Mauro Gallegati, and Antonio Palestrini, "A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility," *Journal of Economic Behavior & Organization*, 2005, *56* (4), 489–512.

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

_ and _ , "Firm entry, trade, and welfare in Zipf's world," *Journal of International Economics*, 2013, *89* (2), 283–296.

_ , _ , and Isabelle Mejean, "Firms, destinations, and aggregate fluctuations," *Econometrica*, 2014, *82* (4), 1303–1340.

_ , _ , and _ , "The micro origins of international business cycle comovement," *American Economic Review*, 2018, *108* (1), 82–108.

_ , _ , and _ , "Foreign shocks as granular fluctuations," Working Paper, NBER 2020.

Domar, Evsey D, "On the measurement of technological change," *The Economic Journal*, 1961, *71* (284), 709–729.

Dosi, Giovanni, Mauro Napoletano, Andrea Roventini, and Tania Treibich, "Debunking the granular origins of aggregate fluctuations: from real business cycles back to Keynes," *Journal of Evolutionary Economics*, 2018, pp. 1–24.

Durlauf, Steven N, "Nonergodic economic growth," *The Review of Economic Studies*, 1993, *60* (2), 349–366.

Fernandez-Villaverde, Jesus, Luis Garicano, and Tano Santos, "Political credit cycles: the case of the Eurozone," *The Journal of Economic Perspectives*, 2013, *27* (3), 145–166.

Fornaro, Paolo and Henri Luomaranta, "Aggregate fluctuations and the effect of large corporations: Evidence from Finnish monthly data," *Economic Modelling*, 2018, *70*, 245–258.

Foster, Lucia, John C Haltiwanger, and Cornell John Krizan, "Aggregate productivity growth: lessons from microeconomic evidence," in "New developments in productivity analysis," University of Chicago Press, 2001, pp. 303–372.

_ , John Haltiwanger, and Chad Syverson, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

__ , __ , **and Cornell J Krizan**, "Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s," *The Review of Economics and Statistics*, 2006, *88* (4), 748–758.

Fujiwara, Yoshi, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, and Wataru Souma, "Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms," *Physica A: Statistical Mechanics and its Applications*, 2004, *335* (1), 197–216.

Gabaix, Xavier, "The granular origins of aggregate fluctuations," Working paper, NBER 2009.

__ , "Power laws in economics and finance," *Annual Review of Economics*, 2009, *1* (1), 255–294.

__ , "The granular origins of aggregate fluctuations," *Econometrica*, 2011, *79*, 733–772.

__ **and Ralph SJ Koijen**, "Granular instrumental variables," Working Paper 28204, NBER 2020.

__ **and Rustam Ibragimov**, "Rank- 1/2: a simple way to improve the OLS estimation of tail exponents," *Journal of Business & Economic Statistics*, 2011, *29* (1), 24–39.

Gaffeo, Edoardo, Mauro Gallegati, and Antonio Palestrini, "On the size distribution of firms: additional evidence from the G7 countries," *Physica A: Statistical Mechanics and its Applications*, 2003, *324* (1-2), 117–123.

Gaubert, Cecile and Oleg Itskhoki, "Granular comparative advantage," *Journal of Political Economy*, 2021, *129* (3), 871–939.

Gibrat, Robert, *Les inégalités économiques*, Recueil Sirey, 1931.

Gnocato, Nicolò and Concetta Rondinelli, "Granular sources of the Italian business cycle," Bank of Italy Working Paper 2018.

Gutiérrez, Germán and Thomas Philippon, "Fading Stars," *AEA Papers and Proceedings*, May 2019, *109*, 312–16.

Hall, Bronwyn H, "The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector," *Journal of Industrial Economics*, 1987, *35*, 583–606.

Haltiwanger, John, "Firm dynamics and productivity growth," *European Investment Bank Papers*, 2011, *16* (1), 116–136.

__ **, Ron S Jarmin, and Javier Miranda**, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hulten, Charles R, "Growth accounting with intermediate inputs," *The Review of Economic Studies*, 1978, *45* (3), 511–518.

Hymer, Stephen and Peter Pashigian, "Firm size and rate of growth," *Journal of political economy*, 1962, *70* (6), 556–569.

Jovanovic, Boyan, "Micro shocks and aggregate risk," *The Quarterly Journal of Economics*, 1987, pp. 395–409.

Kalemli-Ozcan, Sebnem, Bent Sorensen, Carolina Villegas-Sánchez, Vadym Volosovych, and Sevcan Yesiltas, "How to Construct Nationally Representative Firm Level Data from the Orbis Global Database: New Facts on SMEs and Aggregate Implications for Industry Concentration," *American Economic Journal: Macroeconomics (Forthcoming)*, 2023.

Klette, Tor Jakob and Samuel Kortum, "Innovating firms and aggregate innovation," *Journal of political economy*, 2004, *112* (5), 986–1018.

Konings, Jozef, Galiya Sagyndykova, Venkat Subramanian, Astrid Volckaert et al., "The granular economy of Kazakhstan," Technical Report 2021.

Koren, Miklós and Silvana Tenreyro, "Volatility and development," *The Quarterly Journal of Economics*, 2007, *122* (1), 243–287.

_ **and** _ , "Technological diversification," *American Economic Review*, 2013, *103* (1), 378–414.

Lee, Youngki, Luís A Nunes Amaral, David Canning, Martin Meyer, and H Eugene Stanley, "Universal features in the growth dynamics of complex organizations," *Physical Review Letters*, 1998, *81* (15), 3275.

Lucas, Robert E, "Understanding business cycles," in "Carnegie-Rochester conference series on public policy," Vol. 5 North-Holland 1977, pp. 7–29.

Luttmer, Erzo G J, "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 2007, pp. 1103–1144.

Magerman, Glenn, Karolien De Bruyne, Emmanuel Dhyne, and Jan Van Hove, "Heterogeneous firms and the micro origins of aggregate fluctuations," NBB Working Paper 2016.

Meyer, J.R. and E. Kuh, *The Investment Decision: An Empirical Study*, Harvard University Press, Cambridge, Massachusetts, 1957.

Miranda-Pinto, Jorge and Yuanting Shen, "A Granular View of the Australian Business Cycle," *Economic Record*, 2019, *95* (311), 407–424.

Mitzenmacher, Michael, "A brief history of generative models for power law and lognormal distributions," *Internet mathematics*, 2004, *1* (2), 226–251.

Neumark, David, Brandon Wall, and Junfu Zhang, "Do small businesses create more jobs? New evidence for the United States from the National Establishment Time Series," *The Review of Economics and Statistics*, 2011, *93* (1), 16–29.

Newman, Mark EJ, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, 2005, *46* (5), 323–351.

Nirei, Makoto, "Threshold behavior and aggregate fluctuation," *Journal of Economic Theory*, 2006, *127* (1), 309–322.

OECD, "Aggregate National Accounts, SNA 2008: Gross domestic product (Edition 2020)," "https://www.oecd-ilibrary.org/content/data/f950d838-en" 2020.

_ , "Aggregate National Accounts, SNA 2008: Population and employment by main activity (Edition 2020)," "https://www.oecd-ilibrary.org/content/data/39003e99-en" 2020.

Osotimehin, Sophie, "Aggregate productivity and the allocation of resources over the business cycle," *Review of Economic Dynamics*, 2019, *32*, 180–205.

Royo, Sebastián, "How did the Spanish financial system survive the first stage of the global crisis?," *Governance*, 2013, *26* (4), 631–656.

Saichev, Alexander I, Yannick Malevergne, and Didier Sornette, *Theory of Zipf's law and beyond*, Vol. 632, Springer, 2009.

Silva, Murilo and Sergio Da Silva, "The Brazilian granular business cycle," *Economics Bulletin*, 2020, *40* (1), 463–472.

Stanley, Michael HR, Luis AN Amaral, Sergey V Buldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael A Salinger, and H Eugene Stanley, "Scaling behaviour in the growth of companies," *Nature*, 1996, *379* (6568), 804.

Stella, Andrea, "Firm dynamics and the origins of aggregate fluctuations," *Journal of Economic Dynamics and Control*, 2015, *55*, 71–88.

Subbotin, Mikhail, "On the law of frequency of error," *Matematicheskii Sbornik*, 1923, *31* (2), 296–301.

Sutton, John, "Gibrat's legacy," *Journal of Economic Literature*, 1997, *35* (1), 40–59.

_ , "The variance of firm growth rates: the 'scaling'puzzle," *Physica A: statistical mechanics and its applications*, 2002, *312* (3-4), 577–590.

Syverson, Chad, "Product substitutability and productivity dispersion," *Review of Economics and Statistics*, 2004, *86* (2), 534–550.

Taymaz, Erol, "Are small firms really less productive?," *Small Business Economics*, 2005, *25* (5), 429–445.

Thesmar, David and Mathias Thoenig, "Contrasting trends in firm volatility," *American Economic Journal: Macroeconomics*, 2011, *3* (4), 143–80.

Vuong, Quang H, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, 1989, pp. 307–333.

Yeh, Chen, "Are firm-level idiosyncratic shocks important for US aggregate volatility?," *US Census Bureau Center for Economic Studies Paper No. CES-WP-17-23*, 2017.

_ , "Revisiting the origins of business cycles with the size-variance relationship," Working Paper 2019.

_ , "Revisiting the origins of business cycles with the size-variance relationship," Working Paper 2021.

Zaliapin, IV, Yan Y Kagan, and Federic P Schoenberg, "Approximating the distribution of Pareto sums," *Pure and Applied geophysics*, 2005, *162* (6-7), 1187–1228.

Zipf, George Kingsley, "Human behavior and the principle of least effort.," *Addison-Wesley press*, 1949.