Idiosyncratic Volatility Shocks and Aggregate Fluctuations: An Appraisal of the Spanish Case ‡

Omar Blanco Arroyo†

Simone Alfarano

Bachelor's Degree in Economics

2015/2016

Abstract

The aim of this work is to find empirical evidence, in case it exists, about the granularity of the Spanish economy. The Granular Hypothesis (Gabaix 2011) has been tested for Spain between the years 1990 and 2014. For this purpose I have used a sample made up 31477 firms, for which I have taken three variables for each company. These are: sales, number of employees and the SIC code.

First, I show that the firm's size distribution is fat tailed. This fact is essential in order to show that the idiosyncratic shocks do not cancel out. Therefore, the Law of Large Numbers does not apply in this case. Second, I calculate the granular residual in the sense of Gabaix (2011).

The main result obtained is that the top 100 largest firms in Spain account for one third of the aggregate output variability, in line with the original result or Gabaix for american firms.

Keywords: granularity, Granular Residual, idiosyncratic shocks, aggregate uctuations, business cycle, power-laws.

JEL Codes: E32, C16.

---

‡I would like to thank Simone Alfarano for his implication and help from the first steps of this project. Moreover, I would like to mention the help provided by Marko Petrovic, who has given me the database employed and has explained me the most complex part of the code. Finally, to acknowledge to David Vidal for his comments to improve this work.

†al185060@uji.es; Group C

Contents

1 Introduction **1**

2 Related literature **3**

3 Data set **5**

3.1 Descriptive statistics 11

3.2 Empirical distributions 14

4 Power law **16**

4.1 Empirical specification 16

4.2 Evidence of power law 19

5 Idiosyncratic shocks with power law rms **23**

5.1 The model and its calibration 23

5.2 Testing the diversification argument 25

5.3 Approximation to the granular effects 26

6 Granular Residual **27**

6.1 Definitions and methodology 28

6.2 Empirical Granular Residual 29

6.3 Robustness 31

7 Conclusion **32**

8 References **33**

A R codes **36**

A.1 Power law R code 36

A.2 Granular R code 38

B Granular Residual Graphs **40**

List of Figures

1 Sum of the sales of the top 50 and 100 firms in SABI 2
2 Hodrick-Prescott cyclical component of the real GDP 6
3 Total Factor Productivity in Spain 7
4 Spanish production network 10
5 Descriptive statistics for the variable Sales 12
6 Descriptive statistics for the variable Employees 13
7 Firm size distribution 14
8 Zipf's plot 17
9 Power law plot 22
10 Scaling parameter comparison 23
11 Granular Residual graphs 40

List of Tables

1 Sector definitions and number of firms in each sector 9
2 Skewness and kurtosis of ln(sales) by year 15
3 Results from OLS method 20
4 Results from MLE method 21
5 Granular Residual Coefficients OLS Regression 30
6 Granular Residual with industry demeaning coefficients OLS regression . . 30
7 Robustness coefficients OLS regression 31

Listings

1 log-log Regression 36
2 poweRlaw package for MLE estimation 37
3 Input Granular Residual 38
4 Granular Residual 38
5 Granular Residual Industry Demeaning 39

Idiosyncratic Volatility Shocks and Aggregate Fluctuations: An Appraisal of the Spanish Case

Omar Blanco Arroyo

1 Introduction

THIS WORK AIMS at replicating the analysis proposed by Gabaix (2011) for the Spanish economy. This author shows that aggregate uctuations can be attributed to the largest firms. Gabaix interprets that the economy is composed by grains, which are the firms. Due to firms are heterogeneous, its impact on the economy is also heterogenous. Based on this approach the author constructs what he calls the granular residual, which is an aggregate measure of the idiosyncratic shocks (see equation 6.1). In his work he constructs this Granular Residual using the 100 largest firms in United Estates. In order to test if the Granular Hypothesis is significative he employs the OLS regression. In particular he regresses the GDP per capita on the granular. The resulting R^2 is the measured used to show if the impact of the largest firms is large enough.

This type of analysis based on the Granular Hypothesis takes as backing the existence of very large firms that account for a significant fraction of the Gross Domestic Product (GDP). Therefore, it seems possible that shocks in one of these large firms could account for aggregate uctuations. Figure 1 is shown as an illustrative example to quantify the fraction accounted by the largest firms, in the sample, of Spanish GDP. As one can see, the top 100 firms in the sample account for 32% of Spanish GDP, from 1999 to 2014, and the top 50 firms is 26%.

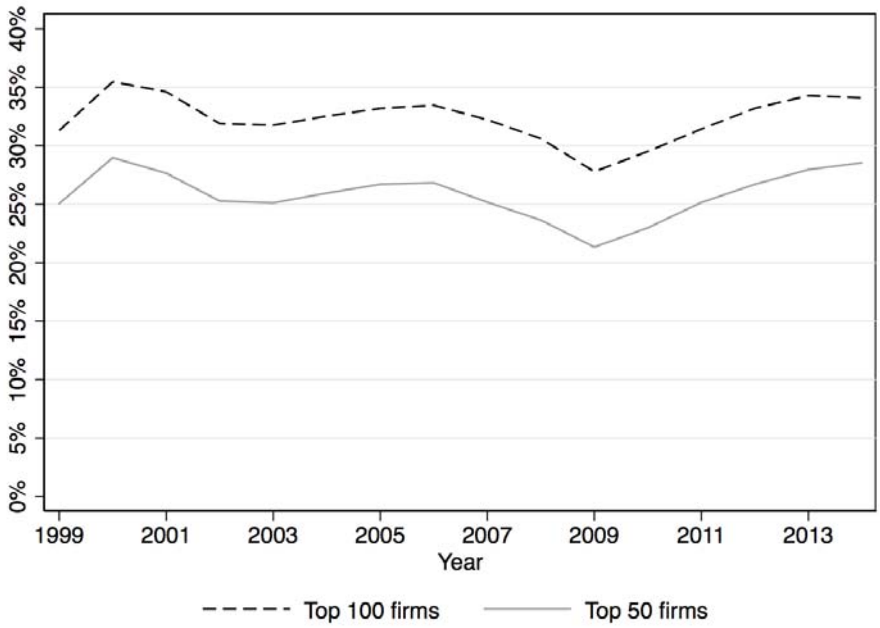

Fig. 1: Sum of the sales of the top 50 and 100 firms in SABI, as a fraction of GDP. *Source*: own elaboration from SABI database.

The Granular Hypothesis has been employed in several empirical works in order to study the volatility in the economic activity. Blank et al. (2009), using an early approach of the granular residual (Gabaix 2009), show that fat-tailed distribution in the German banking sector affects its own stability. On their behalf, Wagner (2012) test Granular Hypothesis on the manufacturer industry in Germany. They found that the Granular Residual is able to explain a significant amount of the sales growth in this industry, the R^2 is 45% approximately. Yet other works have focused on studying variations in exports through the granular approach. This is the case of Di Giovanni & Levchenko (2012) and del Rosal (2013). The first authors showed that trade openness can increase the explanatory power of the granular on the aggregate volatility, whereas the second author reaches the conclusion that the granular behaviour can be present in the exports of all UE countries, causing significant impact on the aggregate output.

The results of this work are, in some sense, in line with the ones introduced by Gabaix (2011), in so far as regressing the GDP growth on granular residual yields 36%. Furthermore, if the industry is taken into account (industry demeaning), this yields to an increase of R^2 of 56%. But, even though this results can be encouraging, there is a significant amount of disparities. First, and most important, when I regress Total Factor productivity on granular, the coefficients are all negative. Besides that, there is a lose of explanatory power when the industry demeaning is done, decreasing from R^2 equal to 30% to 7.5%.

Second, the last lag has a negative coefficient when I regress GDP growth on granular. And, finally, there is an enormous difference between R^2 and $R^2\ adj.$, in both granular residual OLS regression and granular residual industry demeaning OLS regression.

The estructure of this work is as follows. In section 2 a brief review of the literature related to the analysis of aggregate uctuations based on idiosyncratic shocks is done. In section 3, I analyse in detail the data used to perform further analysis and put in context the time period in which the analysis is applied. Section 4 is dedicated to show that firm's size distribution is fat-tailed. Section 5 tries to explain the approach outlined by Gabaix (2011), both for an economy without linkages and with them. In section 6 I construct the granular residual and regress the GDP per capita on it with the aim of quantifying its explanatory power. Finally, Section 7 presents the conclusions I have reached, and I also expose some lines of future work.

2 Related literature

Idiosyncratic shocks, either sector level or firm level, as a source of aggregate uctuations were not considered for the aggregate uctuations. Conventional wisdom argued that idiosyncratic shocks cancel out, hence their impact on the aggregate cannot be an explanation. This argument was built taking into account the Law of Large Numbers, which holds that when the number of independent random variables tends to infinity, the average of the random variables converge to the average population (Wooldridge 2015).[1] This argument, known as "diversification argument" (Acemoglu et al. 2012), is exposed clearly by Lucas (1977), who argues that:

> A new technology, reducing costs of producing an old good or making possible the production of a new one, will draw resources into the good which benefits, and away from the production of other goods. [...] in a complex modern economy, there will be a large number of such shifts in any given period, each small in importance relative to the total output. There will be much "averaging out" of such effects across markets.

[1] Mathematically the Law of Large Numbers for a number of independent random variables, $Y_1, Y_2, \ldots, Y_n$, would be

$$plim\left(\overline{Y}_n\right) = \mu$$

where $\overline{Y}_n$ is the average of the variables and μ is the population average.

Lucas (1977) also mentions that all business cycle are similar, since they are not restricted to a particular type of countries or a certain period of time. Based on these statements Long Jr & Plosser (1983) decide to develop one of the first business cycle models based on multiple sectors.[2] Specifically, the model consists of six perfectly competitive sectors that have links to each other through the intermediate input and no correlation in the aggregate shocks. The authors are able to show how sectors that are suppliers of many other sectors, i.e. have direct links, if they are hit by a shock this is not quickly diluted, as they have the incentive to transmit the benefit of shock, and thus these shocks can become aggregate uctuations, namely, there comovement in the output of the sectors.[3] However, although they were able to generate comovement in the sectors' output, and thus make the shocks more lasting, this occurs when there was a low level of disaggregation. For high levels of disaggregation such shocks succumbed to law of large numbers.

Durlauf (1993) also exposes a sectoral model with linkages between local companies, and argues that complementarities between the companies over time can affect the aggregate behaviour. Moreover, Bak et al. (1993) argue that idiosyncratic shocks are not canceled due to the nonlinearity of interactions. They use a multi-sector and multi-stage model in the production process in which there are intermediate linkages, but only with neighboring companies.

Following the line of works that focus on the propagation of shocks to industry level through the input-output matrix, the work of done by Horvath (1998) showed that in order to idiosyncratic shocks not disappear in the aggregate, i.e. Law of Large Numbers does not apply, it is necessary that (1) sectors (key sectors) which serve as input to many others were few, and (2) there were a lack of substitutability between the intermediate inputs of the sectors. In this way supply sectors must react to shocks the same way as the key sector, which provides them intermediate input. Therefore, Horvath argues that the ratio to which is applied the law of large numbers is not proportional to the number of sectors but the number key sectors. Subsequently, Horvath (2000), based on the approach already exposed, shows that the results are generalizable. To this end he develops a multi-sector general equilibrium model, which is able to explain persistents business cycles and sectoral comovement.

The alternative view to Horvath (1998, 2000) comes from Dupor (1999), who shows,

[2] Kydland & Prescott (1982) also try to give explanation to the business cycle through technology aggregate shocks, receiving criticism that such technological shocks should be much higher than observed to produce the effects they exposed (Summers 1986).

[3] Long Jr & Plosser (1983) define comovement as cross-sectoral correlation.

from a set conditions, that the way in which sectors are related is not relevant to the behavior of the aggregate volatility generated by idiosyncratic shocks, since as sectors are more disaggregated, aggregate volatility converges to zero at the ratio predicted by the law of large numbers, $1/\sqrt{N}$.

More recently, taking the input-output matrix as a medium of propagation of shocks, Acemoglu et al. (2012) claims that depending on the structure of the interconnections that exist in the network, which is a representation of the input-output matrix, idiosyncratic shocks can not stay where they were originated, but spread across the economy, thus affecting the output of other sectors.[4] Thereby the ratio at which aggregate volatility decays depends on the network structure. This approach is empirically demonstrated by Acemoglu et al. (2015), who, through the study of four different types of shocks, conclude, inter alia, that the shocks propagated through the network are significant quantitatively, and more important than the direct effect of shock, up to five times more important.

So far all cited authors place their approach in a closed economy, that is why di Giovanni et al. (2014) studied whether large enterprises imported in cross-border comovement, since "the ow of international trade is dominated by only a handful of large companies". That is, the authors, with this approach, intended to study whether shocks are transferred between countries via trade of large enterprises. The result of their study is that large French companies, which have direct trade linkages, are important in the aggregate comovement. Specifically, the large companies in the sample account for 20% of aggregate volatility.

Finally, I include the recent work done by Carvalho & Grassi (2015), that based on existing evidence on the importance of large companies on the aggregate outcome (Gabaix 2011, di Giovanni et al. 2014), decide to develop a framework to assess the linkages between micro decisions and volatility of macro-level products. The conclusion reached by the authors is that a large fraction of the aggregate dynamics can be rationalized by the dynamics of large companies.

3 Data set

The data for this study are taken from *Sistemas de Balances Ibéricos* (SABI), a database of Bureau Van Dijk. The period under study is from 1999 to 2014, which is a lapse of time of 16 years. To analyze the characteristics of the economic cycle in this period

[4]Acemoglu et al. (2013) also show that the structure of the network, along with the nature of the idiosyncratic shocks, can have a significant effect on the frequency and depth of recessions.

I have extracted the cyclical component of real GDP.[5] In order to do this I applied the filter Hodrick-Prescott (Hodrick & Prescott 1997) to real GDP in logarithms. Such filter is defined as

$$\min_{\{g_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\}$$

where y is the GDP in logarithms, g is the GDP growth y λ is the smoothing parameter, which in this case it equals 100 because GDP is annual frequency period. As one can see in figure 2, in this period, Spain experienced a complete economic cycle. The expansion phase in the sample ranges from 1999 to 2008. However, the expansion phase began in 1995 with the entry of Spain in the Monetary Union. Because of low interest rates and the absence of exchange rate risk, Spain experienced a sharp increase in credit (Fernandez-Villaverde et al. 2013), which subsequently led to an increase in consumption and investment.

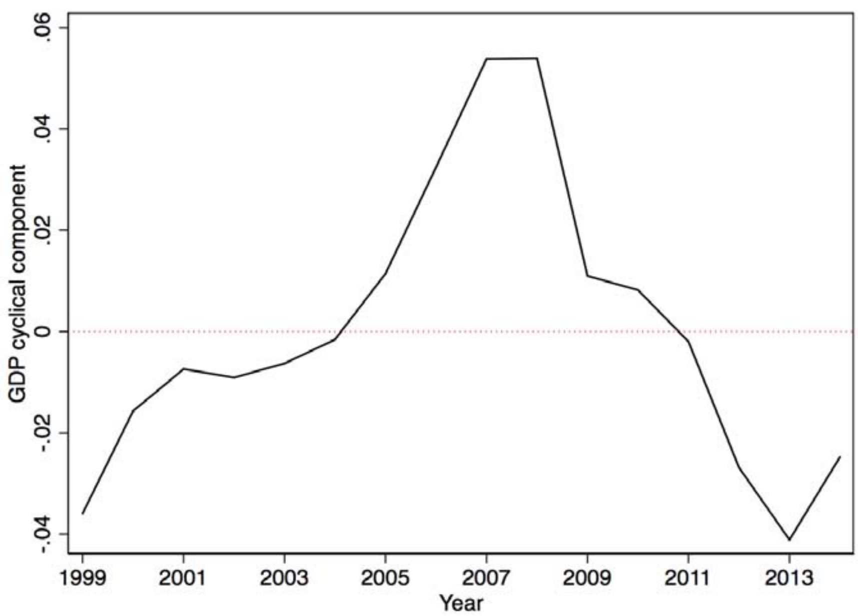

Fig. 2: Hodrick-Prescott cyclical component of the real GDP. *Source*: own elaboration.

The expansion phase was followed by a phase of deep recession, which began by the in uence of the global financial crisis and continued due to the important problems of the Spanish economy. In particular because of the bubble in the construction sector, which has a very important role in the Spanish economy, as shown in Figure 4, and the accumulated household debt during the boom years. In addition, another problem, which

[5]Real GDP has been obtained from the *Eurostat* database.

is structural, of the Spanish economy is low productivity. As shown in Figure 3, Total Factor Productivity (TFP) dropped continuously.[6] From 2009 to 2014 it has recovered almost half of the TFP lost during the period of expansion.

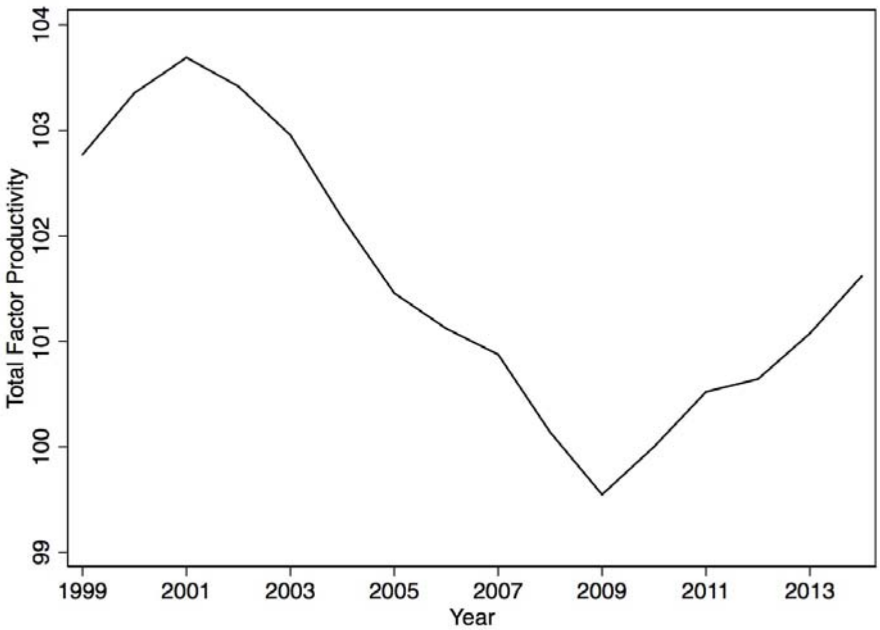

Fig. 3: Total Factor Productivity in Spain. Base 100 for the year 2010. *Source*: Banco de España.

The sample obtained from *SABI* database is made up 36474 firms. The variables taken are operating revenue, which for the purpose of this work I consider as sales from now on, number of employees and the SIC code for each firm.[7] However, to carry out this work there are certain sectors that should not be considered, since variations on their sales are not the result of variations in the current productivity but come from exogenous shocks. With the goal of not considering these types of companies in the sample I have carried out a filtering process which eliminated all belonging to sectors related to hydrocarbons (SIC codes 13, 28 and 29), energy (SIC code 49) and finance (SIC codes 60 and 69). In the case of oil and energy sectors, the changes in their sales are not consequence of changes in their productivity but they are mainly provoked by changes in the raw material prices, while in the case of financial companies, sales variations may be affected by non-real factors. Moreover, due to the limitations of having only two digits in the sample, I deleted the companies that its name contains the word “trading" or the

[6] For more information about the behavior of productivity in Spain since the entry into the Monetary Union until 2012 I suggest seeing the work of Hospido & Moreno-Galbis (2015).

[7] *Standard Industrial Classi cation* is a system for classifying industries through a four-digit code used by government agencies of different countries, including United States and United Kingdom. In this work I have used two digits.

word “petrol". Finally, all observations with missing values in one of the variables have not been considered, since in order to calculate the productivity for each firm I will need both sales and number of employees.

The resulting number of firms after applying all the filtering process is shown in the table 1. The total number of firms used in the following analysis is 31477 companies, of which about half come from four sectors, namely: (1) Wholesale Trade - Durable Goods, (2) Wholesale Trade - Nondurable Goods, (3) General Building Contractors y (4) Food & Kindred Products. Furthermore, these sectors have the larger number of firms inside de top 100 during the lapse of time studied. The top 100 companies is the classification of the largest firms by size, measured this as the amount of sales within a given year.

In order to find out whether the sample is representative of the Spanish economy, I have drawn the input-output table of 2010 as a network.[8] In Figure 4 are depicted the sectors by volume of inputs and outputs at current prices. Those sectors with the largest volume are at the center of the network, and on the periphery are those who have less volume. Additionally, the weighting of the nodes by volume allow to identify those sectors with the highest relative importance in the Spanish economy.

Both the summary table and the network agree in pointing out that the most important sectors in Spain are Wholesale, Construction and Food products. Other important sectors are Electricity and Real State. Yet these have not been considered in the sample, because of the reasons above mentioned.

[8]The year 2010 is the last year in which these tables tables were calculated, since INE calculates them every five years, and 2015 tables are not available yet. More recent input-output tables, year 2011, can be found at OCDE.stat. Nevertheless, these tables contain less sectors than the other ones and they are calculated in dollars.

Table 1: Sector definitions and number of firms in each sector. Firms operating in more than one sector are classified according to the first SIC code they have. The fourth columns refers to the firms which have been in the top 100 for at least one period. The fifth column refers to the top 100 firms that have observations in all periods. *Source*: own elaboration from SABI database.

Division	SIC/Sector	Total Nº. of rms	Top 100 rms	Top 100 long-lived
Agriculture, Forestry, & Fishing	01 Agricultural Production - Crops	112	0	0
	02 Agricultural Production - Livestock	258	0	0
	07 Agricultural Services	61	1	0
	08 Forestry	23	0	0
	09 Fishing, Hunting, & Trapping	49	0	0
Mining	10 Metal, Mining	17	0	0
	12 Coal Mining	23	0	0
	14 Nonmetallic Minerals, Except Fuels	159	0	0
Construction	15 General Building Contractors	4076	16	2
	16 Heavy Construction, Except Building	390	7	2
	17 Special Trade Contractors	943	3	0
Manufacturing	20 Food & Kindred Products	1918	19	2
	21 Tobacco Products	14	1	0
	22 Textile Mill Products	267	0	0
	23 Apparel & Other Textile Products	163	0	0
	24 Lumber & Wood Products	225	0	0
	25 Furniture & Fixtures	190	0	0
	26 Paper & Allied Products	422	4	0
	27 Printing & Publishing	472	0	0
	30 Rubber & Miscellaneous Plastics Products	475	2	1
	31 Leather & Leather Products	55	0	0
	32 Stone, Clay, & Glass Products	804	0	0
	33 Primary Metal Industries	484	11	1
	34 Fabricated Metal Products	932	0	0
	35 Industrial Machinery & Equipment	626	0	0
	36 Electronic & Other Electric Equipment	486	5	1
	37 Transportation Equipment	551	15	5
	38 Instruments & Related Products	106	0	0
	39 Miscellaneous Manufacturing Industries	101	0	0
Transportation & Public Utilities	40 Railroad Transportation	20	2	0
	41 Local & Interurban Passenger Transit	326	1	0
	42 Trucking & Warehousing	702	0	0
	43 U.S. Postal Service	16	1	0
	44 Water Transportation	273	0	0
	45 Transportation by Air	87	7	0
	47 Transportation Services	414	9	0
	48 Communications	251	16	2
Wholesale Trade	50 Wholesale Trade - Durable Goods	5169	35	5
	51 Wholesale Trade - Nondurable Goods	3757	24	2
Retail Trade	52 Building Materials & Gardening Supplies	67	1	0
	53 General Merchandise Stores	68	1	0
	54 Food Stores	296	16	5
	55 Automative Dealers & Service Stations	305	1	0
	56 Apparel & Accessory Stores	171	6	1
	57 Furniture & Homefurnishings Stores	175	1	0
	58 Eating & Drinking Places	170	0	0
	59 Miscellaneous Retail	321	2	0
Services	70 Hotels & Other Lodging Places	559	0	0
	72 Personal Services	74	0	0
	73 Business Services	1582	15	0
	75 Auto Repair, Services, & Parking	199	0	0
	76 Miscellaneous Repair Services	73	0	0
	78 Motion Pictures	143	1	0
	79 Amusement & Recreation Services	301	1	0
	80 Health Services	299	0	0
	81 Legal Services	48	0	0
	82 Educational Services	113	0	0
	83 Social Services	86	1	0
	84 Museums, Botanical, Zoological Gardens	18	0	0
	86 Membership Organizations	6	0	0
	87 Engineering & Management Services	986	6	0
Total No. Firms		**31477**	**231**	**29**

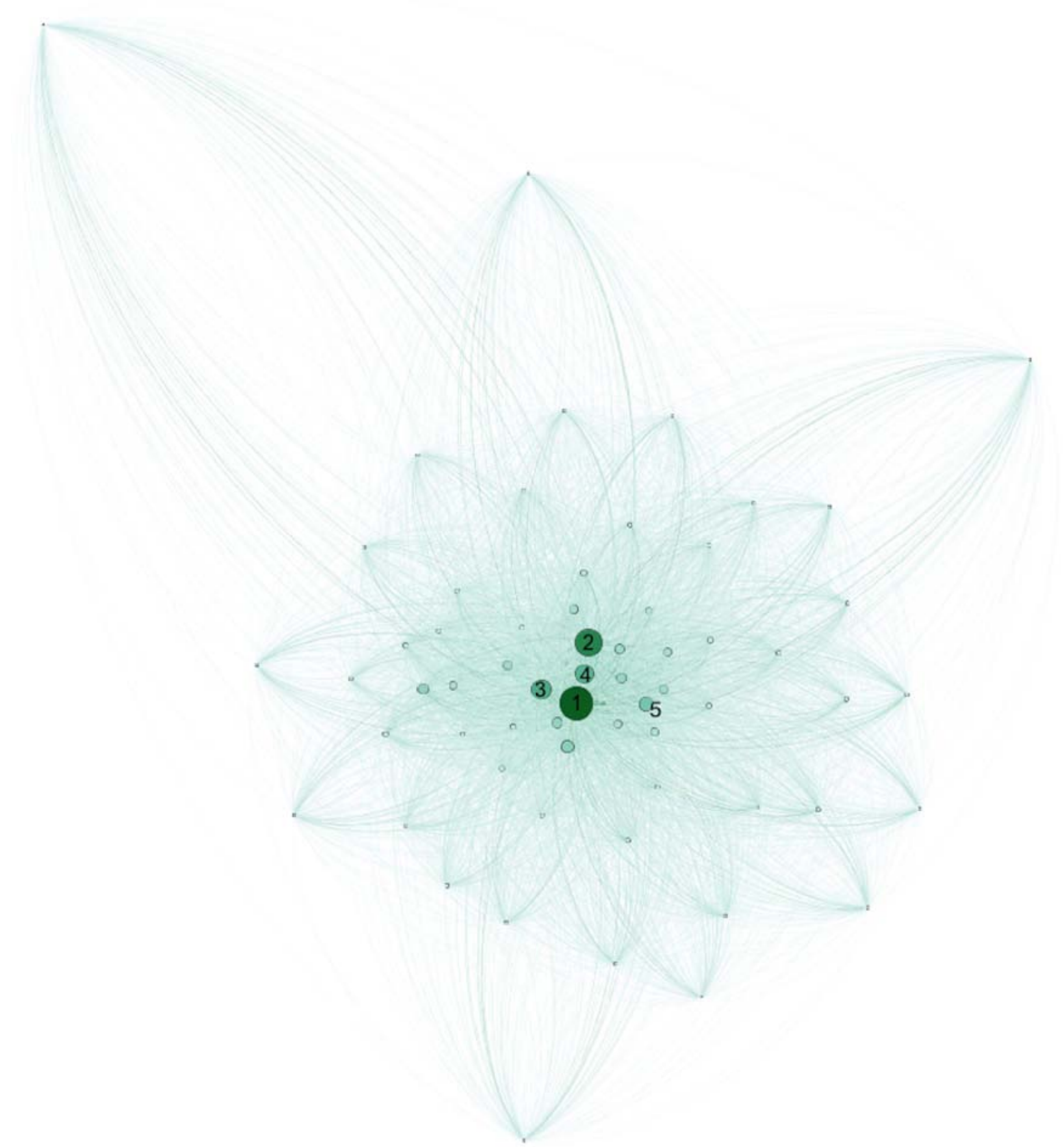

Fig. 4: The production network corresponding to Spain input-output table at basic prices in 2010. The sample is composed by 64 sector. Larger (green) nodes closer to the centre of the network represent sectors having more weight in the economy. The 1-5 labels give the top 5 in the ranking. Constructions(1), Food and beverages(2), Electricity(3), Whole sales(4) and Real state services(5). Although industries in the input-output tables are reported in the NACE code, the interpretation is not affected.[9] *Source*: Instituto Nacional de Estadística (INE). Done using GEPHI.

Having presented the process by which it was obtained the final sample of firms that will be used in this work, the next step is to present the main statistics that identify the sample.

[9]NACE acronym comes from the French term “nomenclature statistique des activités économiques dans la Communauté européenne". It is the classification system of economic activities used in the European Union. This code is used for the organization and recording of data in the framework of Eurostat.

3.1 Descriptive statistics

In order to introduce some of the characteristics of the sample, I used a box plot of the data. But because of the high dispersion of the observations, consequence of the existence of very small and very large companies, I have not shown the entire sample in a single graph but in six different parts.[10] To this end, I have ordered the companies in descending order by size, both for sales variable as employees, and then I have divided them.

For the sales variable, shown in Figure 5, the first notable feature in all panels is that the evolution of the median over the years is very similar to the evolution of the economic cycle, shown in Figure 2. One can also observe that most of the sales volume is concentrated on the right side of the distribution, this means that a relatively small part of the companies has a turnover much greater than most companies that make up the distribution.

Focusing on the first panel, the top 100 companies with the largest sales volume, the evolution of the largest firm in the sample stands out throughout the recession phase of the Spanish economy. This company is Mercadona SA, which since 2007 has the largest sales volume of the sample. This company, unlike others, has not suffered, at least in terms of sales, the impact of the recession. In particular, Mercadona has risen from 0.43% sales as a fraction of GDP in 2007 to 1.77% in 2014. This represents an increase of approximately 142%. In the hypothetical case that Mercadona suffered a shock it is hard to say that it would not have a significant impact on the Spanish economy. This case provides a first insight into the impact of large companies on the evolution of the economy.

Regarding the number of employees, shown in Figure 6, the figures are not far from those presented to the sales variables. There remains a right asymmetry in the distribution for all panels, so most of the number of employees presented in the distribution belongs to a small number of companies that are considerably larger. Furthermore, within the top 100 Largest Employers Mercadona SA has again highlighted by growth in times of crisis.

From Figures 5 and 6, we can say that there is a first evidence of the existence of granularity in the Spanish economy. What on the other hand is not a novel finding, since it is obvious that there are wide disparities between firms. However, which is not so obvious is whether this granularity is significant for the Spanish case.

[10]The criterion for dividing the sample is discretionary.

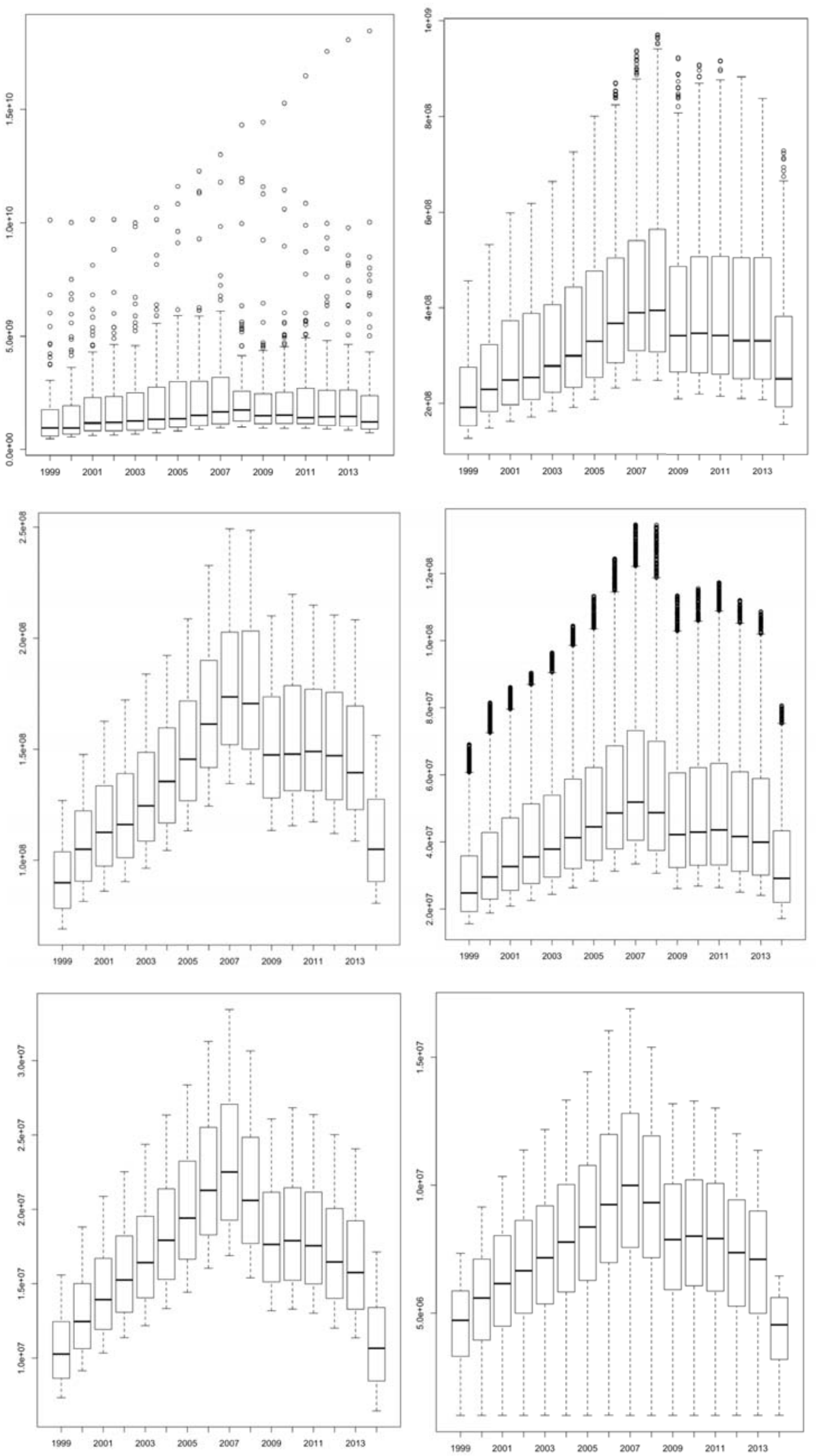

Fig. 5: Descriptive statistics for the variable Sales. Top left panel shows the firms ranked between 1 and 100 with more sales and top right panel shows firms between 100 and 500. At the middle, the left panel shows firms between 500 and 1000, and right panel between 1000 and 5000. Bottom right panel are firms between 5000 and 10000, and bottom left panel shows firms between 10000 and the end of the sample. *Source*: own elaboration from SABI database.

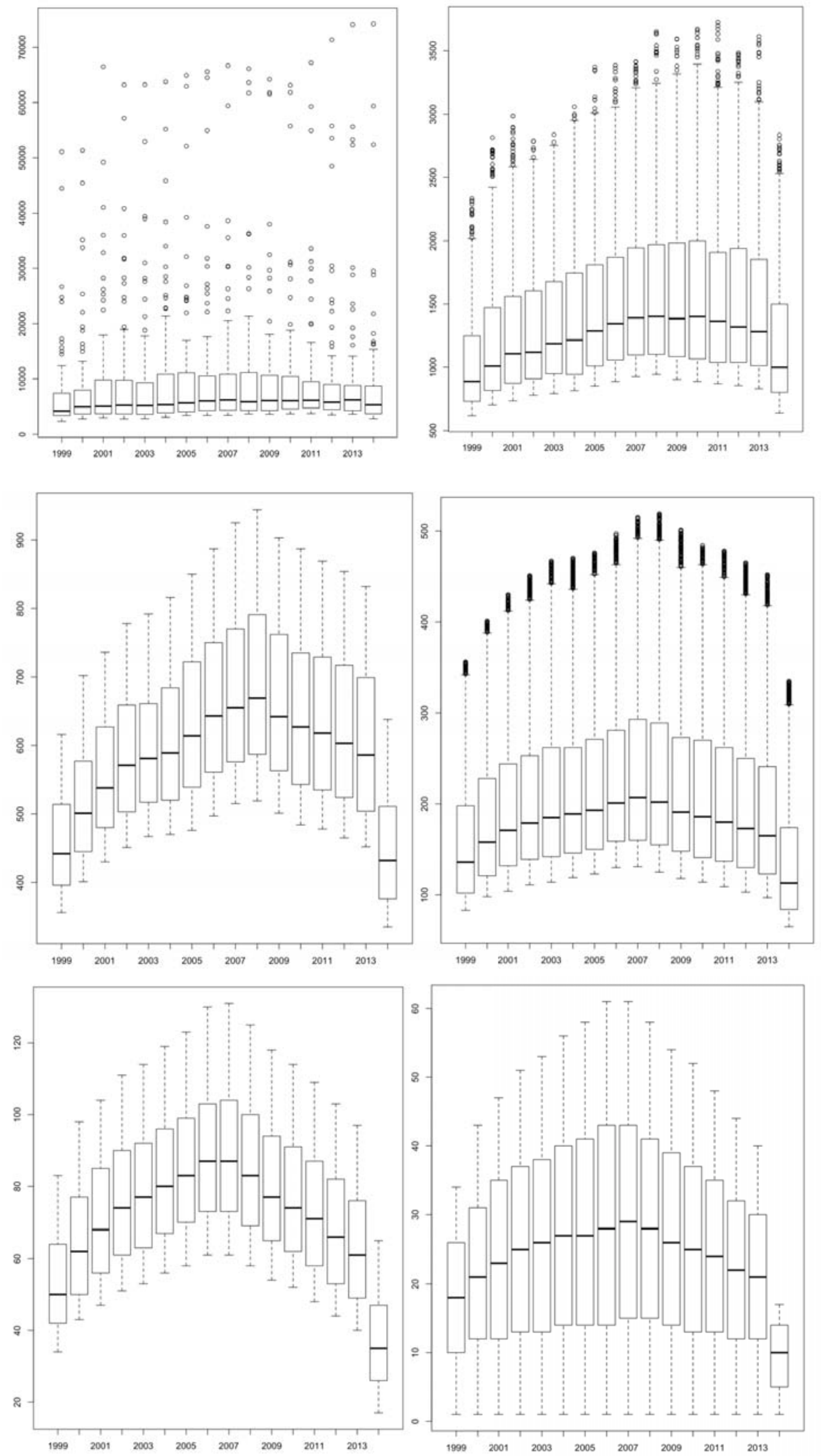

Fig. 6: Descriptive statistics for the variable Employees. The panels have been built following the same criteria than the ones showed in the figure 5. *Source*: own elaboration from SABI database.

3.2 Empirical distributions

To carry out the graphical representation of the distribution of company size, measured by sales volume, I have resorted to the kernel density estimation, a nonparametric method of estimating the probability density function. The kernel density estimator is defined as

$$\widehat{f}_h(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where $K(\cdot)$ is a kernel function and h is a smoothing parameter which must be greater than zero. For the representation of the distribution I have chosen a bandwidth equal to 0.5, following the analysis of Segarra & Teruel (2012). The resulting estimated density is presented in Figure 7 along with normal density.[11] As one can see, for both 2007 and 2014, the tails are wider at the estimated density than the normal one.[12]

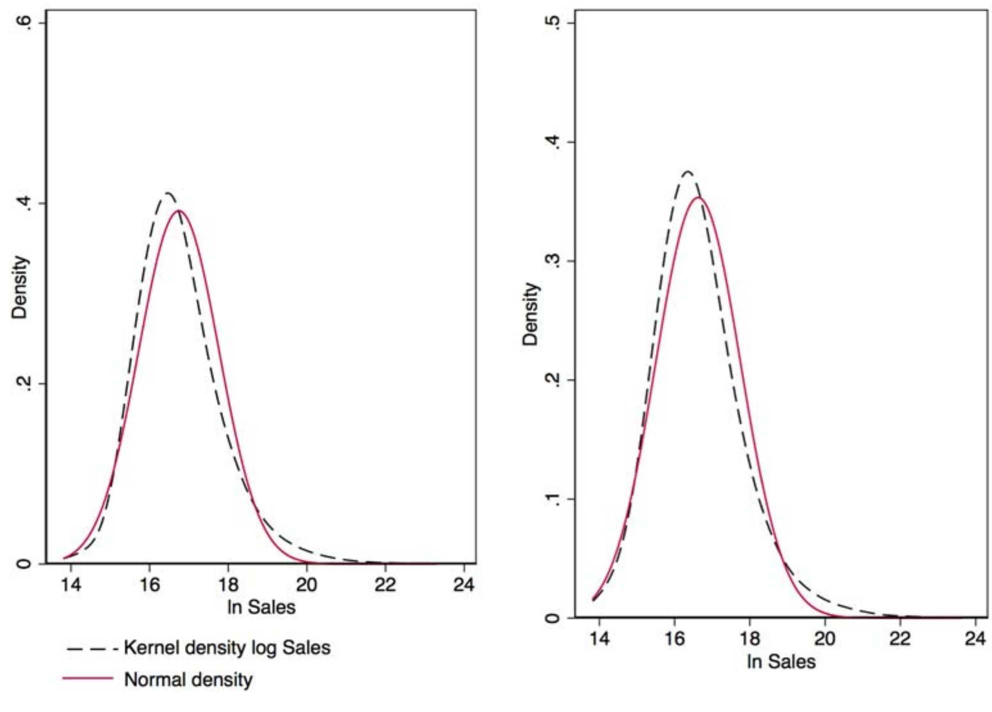

Fig. 7: Firm size distribution of $\ln Sales$ in 2007 in the left panel, and $\ln Sales$ in 2014 in the right panel. The curves are obtained using a normal kernel density smoother with a bandwidth of 0.5. *Source*: own elaboration.

With the aim to extend the analysis to all periods of the sample I made the normality

[11] In this case, I consider a log-normal distribution for sales variable, i.e $\ln S \sim \mathcal{N}(\mu, \sigma^2)$.

[12] The analysis could be made for any of the years under study, but I think that these years may be especially important, as they represent one the last year the growth cycle and another the last year of the sample, which seems to have a glimmer of recovery after the recession.

test based on the Jarque-Bera (JB) statistic. This test checks whether the skewness and kurtosis of the sample coincide with the normal distribution. The statistical JB is defined as

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

where n is the sample size, S is the skewness coefficient y K kurtosis coefficient.[13] The results are presented in Table 2, in which one can see how in the majority of the years under study there is a positive skewness and kurtosis are greater than three, which is the value of a normal distribution. The value of kurtosis is greater than three, this indicates that the distribution is leptokurtic, i.e. the distribution is fat-tailed. Finally, the test of normality is done. For this test, I specified the null hypothesis as existence of lognormality distribution. As it is shown in Table 2, there is significance, i.e. we can reject the null hypothesis of lognomality. In all the years under study a 1% significance level.

This find is in line with some studies, such as Ganugi et al. (2005) and Reichstein & Jensen (2005), who have have concluded that they can reject the lognormal hypothesis in firm's size distributions, measuring the size as the amount of sales.

Table 2: Skewness and kurtosis of ln(sales) by year. * Significant at 1%. *Sources*: own elaboration from SABI database.

Year	N	Skewness	Kurtosis	JB statistic
1999	16602	1.039	5.242	6461.8*
2000	18291	1.007	5.214	6829.2*
2001	19509	1.069	5.490	8754*
2002	20585	1.077	5.559	9596.8*
2003	21214	1.075	5.572	9938.7*
2004	21756	1.103	5.707	11051*
2005	22080	1.141	5.790	11953*
2006	22467	1.129	5.820	12217*
2007	21750	1.123	5.813	11747*
2008	20679	1.139	5.805	11254*
2009	20696	1.139	5.722	10865*
2010	20292	1.114	5.667	10211*
2011	19429	1.029	5.438	8242.4*
2012	18589	0.998	5.275	7093.9*
2013	17423	0.961	5.165	6086.3*
2014	11960	0.987	5.364	4727.7*

13 Pearson's skewness coefficient used is defined as $S = {\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^3\right\}}\Big/{\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{3/2}\right\}}$. And Pearson's kurtosis coefficient used is defined as $K = {\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4\right\}}\Big/{\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{2}\right\}}$

4 Power law

Axtell (2001) found evidence that the distribution of company size, measured in sales, can be adjusted by Zipf's law (Zipf 1949), which is within the category of power law distributions.

Therefore, based in the existing literature and evidence presented so far, I have decided to analyze the existence of a power law distribution to explain the firm's size distribution. This is important to quantify the level of heterogeneity in the size of firms, which is crucial for the granular effect.

4.1 Empirical speci cation

According to Gabaix (2008), a X variable describes a power law if its Complemetary Cumulative Distribution Function (CCDF), or $\mathbb{P}(X \geq x)$,[14] satisfies, at least in the upper tail, the following expression:

$$\mathbb{P}(X \geq x) \simeq kx^{-\zeta} \tag{4.1}$$

where $\zeta \geq 1$. This distribution is also known as Pareto distribution, since Pareto (1896) found that the upper tail of the distribution of a number of people with an income S greater than a large amount x is proportional to $1/x^{\zeta}$, for $\zeta > 0$, i.e.

$$\mathbb{P}(S \geq x) = \frac{k}{x^{\zeta}} \tag{4.2}$$

where k is a constant. A special case of Pareto's distribution is when the exponent is close to one. This case is known as Zipf's law. With $\zeta \simeq 1$ the firm's size is inversely proportional to the firm's rank (Segarra & Teruel 2012). This means that if we order from largest to smallest size a sample made up N observations for a variable x, $x_1 \geq \ldots \geq x_N$, the size of this variable is equal to $1/rank$ times the size of the greatest firm. Therefore, a Zipf's distribution can be expressed as:[15]

$$rank = N \cdot \mathbb{P}(X \geq x) = N\frac{k}{x^{\zeta}} \tag{4.3}$$

[14]Following Clauset (2011), a CCDF is defined as $1 - \mathbb{P}(X < x)$, where $\mathbb{P}(X < x)$ is the Cumulative Distribution Function (CDF), which it is defined as the density function that falls within a certain value x.

[15]The same approach is exposed by Stanley et al. (1995) in order to explain Zipf's plot.

Taking logarithms on both sides of the equation 4.3, the following expression is straightforward:

$$\ln\ rank = K - \zeta\ ln\ x \tag{4.4}$$

where K is a constant. Figure 8 shows in the left panel the CCDF, and the right panel the CCDF in logarithmic terms. As one can see, the CCDF has a characteristic shape, in “L". This indicates that a very small number of companies account for a large volume of sales, and a very large number of companies account for a very small volume of sales.

Figure 8 shows the result of applying the Zipf's law, or “rank-size rule" as it is also known. In the left pane is depicted empirical the CCDF, where “L" shape can be seen, which is characteristic of the "rank-size rule".

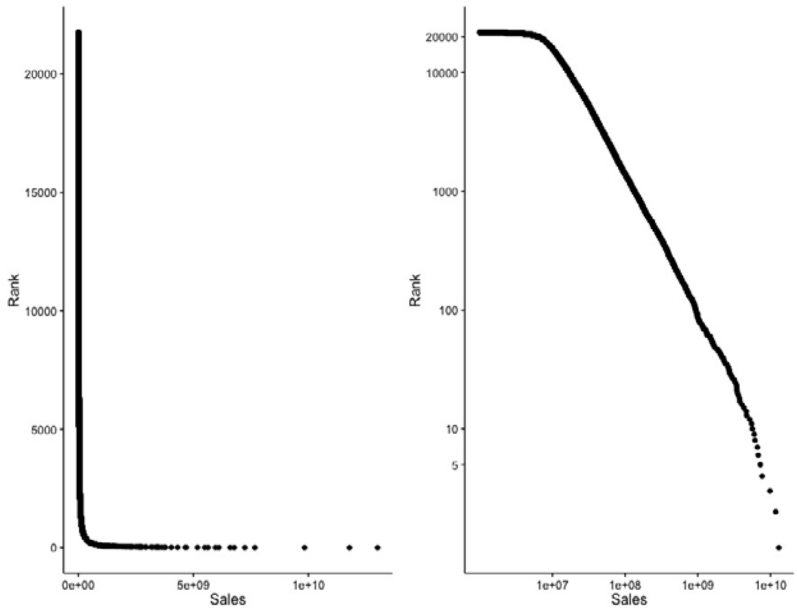

Fig. 8: Rank-Sales plot in the left panel and the same plot but in log-log scale, also known as Zipf's plot, in the right panel. Year 2007. *Source*: own elaboration.

In the equation 4.4., the exponent could be estimated using Ordinary Least Squares (OLS). However, Gabaix & Ioannides (2004) found that the exponent ζ estimated by OLS is biased. So I modify the equation 4.4 using the proposal made by Gabaix & Ibragimov (2011), which makes it possible to estimate by OLS, and results not only an unbiased but optimum coefficient. Such modification is as follows:

$$\ln\left(rank - \frac{1}{2}\right) = K - \zeta\ ln(x) + \varepsilon \tag{4.5}$$

for which the authors propose a standard error associated to $\widehat{\zeta}$ equal to $\widehat{\zeta}(2/n)^{1/2}$. Nevertheless, Clauset et al. (2009) claims that there are problems in using OLS to estimate the exponent of a power law distribution. In their work, these authors suggest that there are problems in using OLS in order to estimate the probability density function $p(x) \sim cx^{-\alpha}$, from the logarithmic transformation $\ln\ p(x) = c - \alpha\ \ln\ x$. Such probability density can be estimated by constructing a histogram of the data, and the resulting function can be adjusted by OLS. Specifically, these authors maintain that there are the following problems:[16]

1. Errors of OLS regression (ε in the equation 4.5) are difficult to estimate because they are based on assumptions that do not apply in this case.
2. An adjustment of a power-law function may represent a large fraction of the variance, and hence high R^2 cannot be taken as a evidence in favor of power law.
3. Settings obtained by regression methods usually do not meet basic requirements on probability distributions, such as normality.

Thus, Clauset et al. (2009), as well as Newman (2005) previously, suggest finding the estimator using Maximum Likelihood Estimation (MLE), both are based on the original work of Hill et al. (1975). In the case of continuous data, one begins with a probability density as specified in the following equation:

$$p(x)\ dx = P(x \leq X < x + dx) = C\ x^{-\alpha}\ dx \tag{4.6}$$

where X is the observed value, C es a constant, and $2 < \alpha < 3$. This equation needs a minimum in order to be able to specify whether it has a power law behaviour, since the density diverges when $x \to 0$, then the power law does not hold for all $x \geq 0$. Introducing that $\alpha \geq 1$ they reach the following equation:

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha} \tag{4.7}$$

And it is from this which is calculated $\widehat{\alpha}$ using MLE, which is presented in this equation

$$\widehat{\alpha} = 1 + n \left[\sum_{i=1}^{n} ln \frac{x_i}{x_{min}} \right]^{-1} \tag{4.8}$$

[16]For more detail on the drawbacks see the Appendix A of Clauset et al. (2009).

for this estimator, the authors propose a standard error associated equal to $(\widehat{\alpha}-1)/\sqrt{n}$.

Having exposed theoretically two alternatives to estimate the existence of power law, in the next section proceeds to present the results obtained.

4.2 Evidence of power law

The results obtained by applying the method proposed by Gabaix & Ibragimov (2011) are shown in Table 3. This exponents have been estimated for different sample sizes (N). The results show that as N increases the value of the exponent is smaller. Even for the total sample the exponent is less than 1 for all years. This does not imply the absence of power law, what happens is that this usually requires a minimum from which such behavior in the distribution can be seen (power law behaviour works usually asymptotically). This was mentioned earlier, when I discussed the criticisms made by Clauset et al. (2009) to using OLS for estimating the behavior of the power law distribution.

The problem that arises here is that when trying to estimate ζ using OLS there is no optimal procedure to establish the lower bound from which there may be a power law behaviour. A guide may be graphical representation, such as Zipf's plot, as discussed above in Figure 8, right panel. In this case one can see how from a certain volume of sales there is a truncation point in the distribution. From this point you could set the cutting to calculate regression.

However, despite the above mentioned results indicate the existence of power law, since all coefficients are statistically significant at 1%, and are above 1 for limits below one hundred thousand observations.

The alternative process proposed by Clauset et al. (2009) is more sophisticated than the previous one, and it provides a greater degree of accuracy. Following the work of these authors it has been calculated an optimal minimum (x_{min}) from which one can observe a power law behaviour in the distribution.[17] That threshold means, for each year, a significant reduction in the number of companies used. The year with which employs the greater number of observations is 1999, with 31%, while 2011 is the year with less observations employed, 11%. As it is shown, $\widehat{\alpha}$ is closer to a value of two, and it is between two and three, which is the usual range for the power law.

[17]The results shown in Table 4 were calculated using the package developed for RStudio by Gillespie (2014), who follows the specifications of Clauset et al. (2009). The R code used is shown in Appendix A.2, code 2.

Table 3: Results from *Ordinary Least Squares* (OLS) method. N refers to the threshold selected, $\widehat{\zeta}$ is the scaling parameter estimated by OLS, $Se(\widehat{\zeta})$ is the standard error associated to the estimated scaling parameter and R^2 refers to the coefficient of determination. * Significant at 1%. *Sources*: own elaboration from SABI database.

N	Parameters\Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
100	$\widehat{\zeta}$	1.306*	1.292*	1.355*	1.369*	1.387*	1.321*	1.366*	1.393*	1.441*	1.589*	1.560*	1.528*	1.463*	1.470*	1.402*	1.300*
	$Se(\widehat{\zeta})$	0.185	0.183	0.192	0.194	0.196	0.187	0.193	0.197	0.204	0.225	0.221	0.216	0.207	0.208	0.198	0.184
	R^2	0.954	0.952	0.936	0.936	0.937	0.929	0.938	0.942	0.942	0.970	0.976	0.972	0.967	0.975	0.967	0.966
500	$\widehat{\zeta}$	1.236*	1.268*	1.224*	1.249*	1.267*	1.237*	1.257*	1.278*	1.294*	1.302*	1.259*	1.258*	1.236*	1.232*	1.240*	1.148*
	$Se(\widehat{\zeta})$	0.078	0.080	0.077	0.079	0.080	0.078	0.080	0.081	0.082	0.082	0.080	0.080	0.078	0.078	0.078	0.073
	R^2	0.990	0.990	0.985	0.985	0.986	0.984	0.986	0.987	0.986	0.986	0.986	0.985	0.986	0.986	0.988	0.988
1000	$\widehat{\zeta}$	1.217*	1.234*	1.191*	1.200*	1.209*	1.200*	1.205*	1.224*	1.229*	1.217*	1.196*	1.188*	1.188*	1.185*	1.178*	1.106*
	$Se(\widehat{\zeta})$	0.054	0.055	0.053	0.054	0.054	0.054	0.054	0.055	0.055	0.054	0.053	0.053	0.053	0.053	0.053	0.049
	R^2	0.995	0.995	0.992	0.991	0.991	0.991	0.991	0.992	0.991	0.989	0.991	0.990	0.992	0.992	0.992	0.993
10000	$\widehat{\zeta}$	1.089*	1.107*	1.115*	1.134*	1.139*	1.142*	1.146*	1.148*	1.141*	1.106*	1.100*	1.093*	1.087*	1.072*	1.062*	0.995*
	$Se(\widehat{\zeta})$	0.015	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.014
	R^2	0.997	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.993
Total	$\widehat{\zeta}$	0.903*	0.904*	0.920*	0.935*	0.939*	0.947*	0.957*	0.960*	0.957*	0.938*	0.929*	0.920*	0.898*	0.879*	0.866*	0.858*
	$Se(\widehat{\zeta})$	0.010	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.011
	R^2	0.949	0.945	0.949	0.949	0.948	0.950	0.954	0.952	0.951	0.952	0.953	0.951	0.944	0.942	0.939	0.940

Furthermore, to ensure that the distribution is power law and not lognormal, a comparison was made between the two distributions. The hypotheses proposed for contrast are as follows:

H_0: power-law distribution

H_1: lognormal distribution

The p-value presented in the Table 4 shows that the null hypothesis cannot be rejected for any of the time periods considered. Hence we can say that there is a power law behavior in the distributions of each year, with a low probability of error. As an example, Figure 9 shows how power law distribution is adjusted by the process discussed above. One can see how the adjustment is made from the optimum minimum, this is why in most cases it is said that there is a power law behaviour in the upper tail of the distribution.

Table 4: Results from *Maximum Likelihood Estimation* (MLE) method. N indicates the total number of observations, n refers to the number of firms above the threshold, $\widehat{\alpha}$ is the scaling parameter estimated by MLE, and $Se(\widehat{\alpha})$ is the standard error associated to the estimated scaling parameter. The threshold is in E+07. *Sources*: own elaboration from SABI database.

Year	**N**	**n**	**Threshold**	$\widehat{\alpha}$	$Se(\widehat{\alpha})$	**p-value**
1999	16602	5249	1,49	2,102	0,0152	0,32
2000	18291	5579	1,69	2,111	0,0149	0,13
2001	19509	5136	2,03	2,130	0,0158	0,55
2002	20585	5175	2,18	2,141	0,0159	0,21
2003	21214	5048	2,42	2,158	0,0163	0,87
2004	21756	4289	3,03	2,166	0,0178	0,87
2005	22080	4495	3,13	2,166	0,0174	0,92
2006	22467	5070	3,09	2,161	0,0163	0,86
2007	21750	5539	3,05	2,158	0,0156	0,89
2008	20679	5536	2,79	2,108	0,0149	0,80
2009	20696	2625	4,77	2,122	0,0219	0,92
2010	20292	2777	4,65	2,121	0,0213	0,99
2011	19429	2141	5,96	2,141	0,0247	0,86
2012	18589	2793	4,47	2,122	0,0212	0,36
2013	17423	2474	4,81	2,112	0,0224	0,30
2014	11960	3242	2,72	2,076	0,0189	0,41

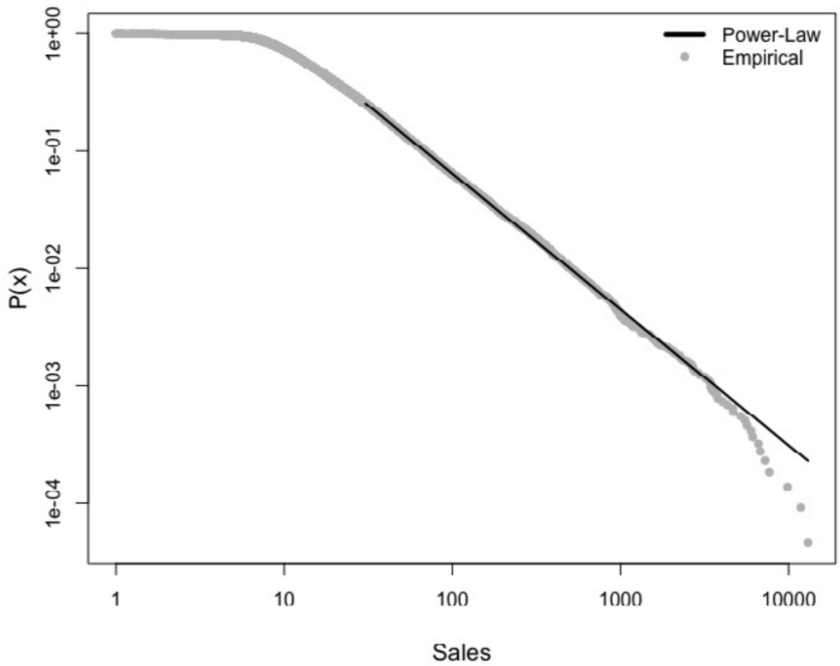

Fig. 9: Power law and its fit for the year 2007. Sales in million. *Source*: own elaboration.

Once submitted the result obtained by applying each of the processes, now I intend to do both comparable. As we have seen, the OLS regression given by Gabaix & Ibragimov (2011) is expressed in Pareto's law terms, following CCDF. Meanwhile, Clauset et al. (2009) establishes its approach from PDF. Adamic (2000) shows how coefficients estimated using OLS can be expressed in the same terms that the approach using PDF. The relationship, from a regression such as $\ln\ rank = K - \zeta\ ln\ x$, is $\alpha = 1 + \frac{1}{\zeta}$. As one can see in equation 4.5, the approach of this work is the inverse, hence the relationship ought to be the inverse, resulting in:

$$\alpha = 1 + \zeta \tag{4.9}$$

The resulting coefficients from applying this transformation to the OLS regression, with a sample size equal to the MLE method used, are presented in Figure 10. Where one can see how the estimated exponents of every year are lower than those resulting from applying MLE. However, they remain above two.

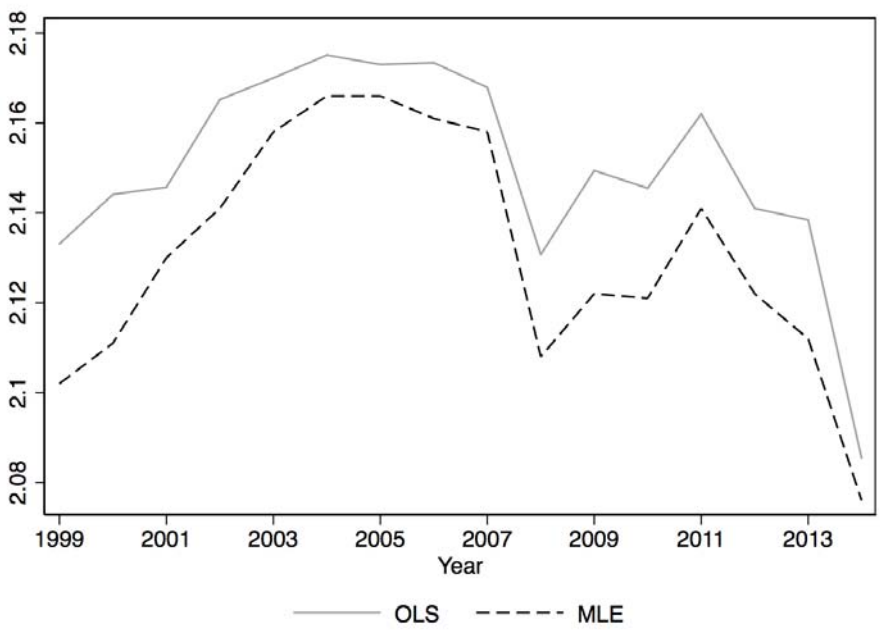

Fig. 10: Scaling parameter comparison between OLS ($\widehat{\alpha}_{OLS}$) and MLE ($\widehat{\alpha}_{MLE}$) methods. *Source*: own elaboration.

Presented the evidence for the existence of power law in the sample, in the following section I analyze the impact of this on the behavior of idiosyncratic shocks in the Spanish economy.

To sum up, showing that exist power law behavior in the firm's size distribution is essential, because it allows us to know that idiosyncratic shocks can resist in the aggregate, since they decay at a lower rate than the one stated by the Central Limit Theorem. This breaks the diversification argument, i.e. idiosyncratic shocks do not average out.

5 Idiosyncratic shocks with power law rms

5.1 The model and its calibration

The model used by Gabaix (2011) to expose its approach has the main characteristic that production is exogenous, like an endowment, which means that a company does not require input from any other to manufacture its products, implying that there is no kind of linkage between the N companies that make up the economy (this is Lucas' *Economy Islands* model). Therefore, if there is no link between companies, GDP (Y) is equal to the total amount of sales (S), in mathematical terms:

$$Y_t = \sum_{i=1}^{N} S_{i,t} \tag{5.1}$$

Then GDP growth will, as shown in equation 5.2, the sum of sales growth of each company multiplied by the relative weight of each company in the economy.

$$\frac{\Delta Y_{t+1}}{Y_t} = \sum_{i=1}^{N} \frac{\Delta S_{i,t+1}}{S_{i,t}} \cdot \frac{S_{i,t}}{Y_t} \tag{5.2}$$

where $\Delta Y_{t+1} = Y_{t+1} - Y_t$.

Gabaix defines sales growth of a representative firm as follows:

$$\frac{\Delta S_{i,t+1}}{S_{i,t}} = \sigma_i \varepsilon_{i,t+1} \tag{5.3}$$

where σ_i is the volatility of the firm i and $\varepsilon_{i,t+1}$ is an error, which is not correlated ($\mathbb{E}\left[\varepsilon_{i,t}, \varepsilon_{j,t}\right] = 0$), with mean equal to cero and variance equal to one. By replacing equation 5.3 into 5.2 one reach:

$$\frac{\Delta Y_{t+1}}{Y_t} = \sum_{i=1}^{N} \sigma_i \, \frac{S_{i,t}}{Y_t} \, \varepsilon_{i,t+1} \tag{5.4}$$

The volatility of GDP growth is defined as:

$$\sigma_{GDP} = \sqrt{\mathbb{E}\left(\frac{\Delta Y_{t+1}}{Y_t}\right)^2}$$

from the equation 5.4, I can derive the following:

$$\sigma_{GDP}^2 = \sum_{i=1}^{N} \sigma_i^2 \, \left(\frac{S_{i,t}}{Y_t}\right)^2$$

This equation indicates that the variance of GDP is the product of the sum of the variance of the idiosyncratic shocks weighted by their weight in the economy. If you continue solving the equation the result is:

$$\sigma_{GDP} = \sqrt{\sum_{i=1}^{N} \sigma_i^2 \, \left(\frac{S_{i,t}}{Y_t}\right)^2} \tag{5.5}$$

Finally, if all companies have the same standard deviation, $\sigma_i \equiv \sigma$, equation 5.5 shall become:

$$\sigma_{GDP} = \sigma \sqrt{\sum_{i=1}^{N} \left(\frac{S_{i,t}}{Y_t}\right)^2} \tag{5.6}$$

where the second term of the product is the square root of the Herfindahl index of the economy. Expressed in the same terms as Gabaix, it would be:

$$\sigma_{GDP} = \sigma\ h \tag{5.7}$$

where h is the square root of the Herfindahl index.

5.2 Testing the diversi cation argument

At this point, it is necessary to check whether the diversification argument is true, i.e. if all companies have the same weight in the economy. This fact implies that:

$$\frac{S_{i,t}}{Y_t} = \frac{1}{N}$$

therefore the equation 5.6 becomes:

$$\sigma_{GDP} = \frac{\sigma}{\sqrt{N}} \tag{5.8}$$

To check if this is the case of the Spanish economy I have obtained the number of Spanish companies, which is 3.168.878,[18] and I have calculated the average standard deviation of shocks in the productivity of companies in the sample for the period 1999-2014, which is about 33.13%, the shock is measured as the volatility $\Delta \ln(sales_{i,t})$. According to equation 5.7 Spanish GDP uctuations are around the 0.019%, which is not consistent with the observed uctuations, which would be around 4.36% between 1999 and 2014.

According to Gabaix (2011), when firm's distribution is a power-law distribution, as is the case of the sample that is being used, the diversification argument can only be

[18]Data for year 2015, extracted from *Directorio Central de Empresas*, calculated by INE. This number considers all companies regardless of whether they have a salaried employees or not, and its legal form. I have also calculated σ_{GDP} using the average number of companies, for the years between 1999 and 2015, which is 3.045.477, and the results do not vary significantly.

satisfied if $\zeta \geq 2$. And as one has been seen in Table 3, which follows the Gabaix's method, none of the coefficients is equal to or greater than two. Therefore, the result stated in the previous paragraph is not met because the power-law distribution (with $\zeta < 2$) makes idiosyncratic shocks do not vanish at a rate $1/\sqrt{N}$.

And then, how fast decay idiosyncratic shocks? With $\zeta \simeq 1$ Gabaix (2011) argues that GDP volatility follows

$$\sigma_{GDP} \sim \frac{a_\zeta}{\ln N}\,\sigma$$

where a_ζ is a random variable, whose distribution does not depend on N or σ. Then idiosyncratic shocks decay at a rate equal to $\ln N$ instead of $\sqrt{N}$, so even with a very large number of companies the idiosyncratic shocks can survive.

5.3 Approximation to the granular effects

To establish whether the effects of granularity are large enough to maintain the idiosyncratic shocks, Gabaix (2011), based on Hulten (1978) theorem, now uses the assumption that the economy is composed of companies that have linkages between them. These links are established because some companies buy to others intermediate input for use them in their own production.

Firstly, Hulten showed that aggregate TFP is:

$$\frac{\Delta TFP_{t+1}}{TFP_t} = \sum_{i=1}^{N} \frac{S_{i,t}}{Y_t}\,\frac{\Delta \pi_{i,t+1}}{\pi_{i,t}} \tag{5.9}$$

where $\pi_{i,t+1}$ is the productivity for firm i in the moment $t+1$. Hence the TFP growth volatility, assuming that all companies have the same standard deviation in productivity, is:

$$\sigma_{TFP} = \sigma_\pi \sqrt{\sum_{i=1}^{N} \left(\frac{S_{i,t}}{Y_t}\right)^2} \tag{5.10}$$

where $\sigma_\pi = \sqrt{\mathbb{E}\left(\frac{\Delta TFP_{t+1}}{TFP_t}\right)^2}$ and σ_π is the volatility in the productivity of each company. Equation 5.10 shows that the volatility of TFP is the result of idiosyncratic shocks in productivity weighted by the relative weight of the company in the economy. Expressed in the same terms as Gabaix, it would be:

$$\sigma_{GDP} = \sigma_{\pi}\ h \tag{5.11}$$

where h is the square root of the Herfindahl index.

Based on the equation 5.9, some authors have estimated that GDP growth is proportional to productivity growth multiplied by a factor of use (μ), this is shown in Equation 5.12.

$$\frac{\Delta Y_{t+1}}{Y_t} = \mu\ \frac{\Delta TFP_{t+1}}{TFP_t} \tag{5.12}$$

If the above procedure is applied to calculate the standard deviation, the equation would prove to be:

$$\sigma_{GDP} = \mu\ \sigma_{\pi} \sqrt{\sum_{i=1}^{N} \left(\frac{S_{i,t}}{Y_t}\right)^2} \tag{5.13}$$

which implies that $\sigma_{GDP} = \mu\ \sigma_{TFP}$.

In order to show whether the effects of granular can be important, I have taken the factor of use proposed by Gabaix (2011), $\mu = 2.6$, and I have calculated the square root of the Herfindahl index for companies in the sample during the years between 1999 and 2014, resulting in 4.81%. Also, the average volatility of the productivity of companies in the sample is around 37%, calculated as $\Delta\ \ln(sales_{i,t}/employees_{i,t})$. The result of $\mu \cdot \sigma_{TFP}$ is approximately 4.62 %, which approximates to the observed σ_{GDP}, which is 4.36%.

This section has shown how under a firm's size distribution characterized as a power law, the diversification argument does not apply, namely, that in an economy with a large number of companies shocks do not average out. In addition, I have provided evidence that the Spanish economy may have a significant component of granularity.

6 Granular Residual

Having shown evidence of the existence of granularity, now I intend to quantify how much of the variations of Spanish GDP is explained by productivity shocks in top 100 Spanish companies. To do this, I have calculated the Granular Residual in the same terms as

Gabaix (2011).

6.1 De nitions and methodology

Based on equation 5.9 Gabaix defines its granular residual as:

$$\Gamma_t = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} \left(g_{i,t} - \overline{g}_t\right) \tag{6.1}$$

where K will be the top 100 companies in Spain. The shock is defined as the difference between $g_{i,t}$ and $\overline{g}_t$, being $\overline{g}_t = Q^{-1} \sum_{i=1}^{Q} g_{i,t}$ and $Q = K$.[19] Finally, $g_{i,t}$ is the productivity growth rate, which is defined as the difference between $\ln\left(S_{i,t}/E_{i,t}\right)$ and $\ln\left(S_{i,t-1}/E_{i,t-1}\right)$. Where $S_{i,t}$ It represents the amount of sales of the company i en el period of time t, and $E_{i,t}$ is the number of employees in the company i en el period of time t. Then, $g_{i,t} - \overline{g}_t$ is the estimation of the idiosyncratic shock of firm i. And Γ_t is a sort of weighted sum of this idiosyncratic shocks.

The process for calculating the granular is based on obtaining the amount of sales and number of employees for the top 100 companies for the year $t-1$, and this has to be subtracted from the companies that were in the top 100 in the year t. Following this procedure, some observations have been lost in some years, since there are companies in the top 100 that do not repeat position from one year to another. Despite this, the number of missing observations is not significant for any of the years under study. In addition, to avoid possible distortions arising from excessively high productivity shocks I have made the winsorizing done by Gabaix (2011). This involves replacing the shocks higher than 20%, whether they are positive or negative. In mathematical terms it would be exposed as $T(\widehat{\varepsilon}_{it}) = \widehat{\varepsilon}_{it}$ if $|\widehat{\varepsilon}_{it}| \leq 20\%$, and $T(\widehat{\varepsilon}_{it}) = sign(\widehat{\varepsilon}_{it}) \cdot 20\%$ if $|\widehat{\varepsilon}_{it}| \geq 20\%$, where $\widehat{\varepsilon}_{it} = g_{i,t} - \overline{g}_t$.

In addition, to control shocks by industry I have also calculated the granular residual with industry demeaning, which instead of deducting the average productivity shock of the entire sample, as presented in equation 6.1, only takes into account the industry average productivity shocks. Gabaix shows that this method is a better estimation procedure for idiosyncratic shocks. This is re ected in Equation 6.2 through $\overline{g}_{I_it}$, where I_i refers to the company i belonging to the industry I. In this work I consider the industry as the sector

[19] In this case the number of companies used to calculate the average (Q) is the same as the number of companies that are in the top (K). But this does not have to be like this. I have done so to adjust the analysis to the one done by Gabaix.

associated with the SIC code which the company i belongs to.

$$\Gamma_t = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} \left(g_{i,t} - \overline{g}_{I_i t}\right) \tag{6.2}$$

The mean industry shock has been calculated as the average of shocks that companies belonging to a determined sector and inside the top 100 have had. This procedure has been done to avoid distortions, because if all firms belonging to the same sector (in the whole sample) were considered, then there would be distorted results.

6.2 Empirical Granular Residual

In order to quantify the granular residual explanatory power I have regressed GDP per capita growth on the granular residual and its lags.[20] Moreover, I have regressed TFP growth time serie on the granular residual and its lags.[21]

$$\frac{\Delta GDP_{t+1}}{GDP_t} = \beta_0 + \beta_1 \Gamma_t + \beta_2 \Gamma_{t-1} + \beta_2 \Gamma_{t-2} + \varepsilon_{t+1} \tag{6.3}$$

where β_0 is a constant and ε the error term.[22]

$$\frac{\Delta TFP_{t+1}}{TFP_t} = \gamma_0 + \gamma_1 \Gamma_t + \gamma_2 \Gamma_{t-1} + \gamma_2 \Gamma_{t-2} + \varepsilon_{t+1} \tag{6.4}$$

where γ_0 is a constant and ε the error term.

The results presented in Table 5 show how the explanatory power of granular residual on GDP growth is 36% taking into account the two lags, and 10% with a single lag. However, considering the adjusted coefficient of determination, which penalizes the inclusion of additional variables, the explanatory power is considerably reduced. As regard to TFP growth regressed on the granular residual, R^2 is equal to 30%, regardless of whether the regression includes one or two lags. This leads to the conclusion that the inclusion of the second lag is unnecessary, since it does not improve the explanatory power. Moreover, in this regression is important to point out the negative coefficients.

[20] GDP growth time serie has been obtained from Eurostat database.
[21] TFP time serie has been obtained from Banco de España web page.
[22] The number of lags that I have chosen is the same as Gabaix uses. He does not specify at any time the criteria by which he chooses this amount of lags. In this case, the results do not change if one includes more lags. Regression does not increase the explanatory power, and the coefficients do not increase their significance.

Table 5: Granular Residual Coefficients OLS Regression. Standard errors are given in parentheses. Significancy levels are indicated as: *Significant at 10%,**Significant at 5%,***Significant at 1%. *Sources*: own elaboration from SABI database.

	$GDP\ Growth_t$		$TFP\ Growth_t$	
(Intercept)	0.0270*	0.0147	0.0006	0.0017
	(0,0146)	(0,01242)	(0,0015)	(0,0021)
Γ_t	1.0567	2.2675**	-0.2464*	-0.3204
	(1.1780)	(0.7413)	(0.1137)	(0.1776)
Γ_{t-1}	-1.4912*	0.1546	-0.1820*	-0.1678*
	(0.8100)	(0.6402)	(0.0872)	(0.0879)
Γ_{t-2}		-1.6782*		-0.1111
		(0.8653)		(0.1148)
N	14	13	14	13
R^2	0.1060	0.3636	0.3002	0.3000
$R^2\ adj$	-0.0565	0.1514	0.1730	0.0662

If industry demeaning is applied, results improve in terms of explanatory power, this can be seen in Table 6. Regressing GDP growth on granular residual one can obtain that R^2 adjusted is greater than in Table 5. Particularly, twice times with the inclusion of one lag and a 20% greater with two lags. Nevertheless, regressing TFP growth on granular residual the explanatory power drops, it is reduced by half in the case of one lag and 7.5% for two lags.

Table 6: Granular Residual with industry demeaning coefficients OLS regression. Standard errors are given in parentheses. *Sources*: own elaboration from SABI database.

	$GDP\ Growth_t$		$TFP\ Growth_t$	
(Intercept)	0.0232*	0.0126	-0.0009	-0.0011
	(0.1153)	(0.0092)	(0.0015)	(0.0019)
Γ_t	2.6277	4.7397***	-0.1572	-0.0725
	(1.9663)	(1.0964)	(0.1647)	(0.2274)
Γ_{t-1}	-3.3398***	-1.0461	-0.1413	-0.2333
	(1.013)	(0.0884)	(0.3083)	(0.2288)
Γ_{t-2}		-2.6055***		0.1021
		(0.5819)		(0.2055)
N	14	13	14	13
R^2	0.2069	0.5633	0.1227	0.075
$R^2\ adj$	0.0627	0.4178	-0.0368	-0.2333

One possible explanation for the results of the TFP, so different to those presented by Gabaix (2011), is that the structural characteristics of the Spanish economy are quite peculiar. As shown in Section 3, during the period of expansion of the Spanish economy, the TFP decreased while the GDP increased. Indeed, the correlation coefficient between the GDP cycle, Figure 2, and TFP growth, Figure 3, is the -43%. This event is explained by the housing bubble during the expansion phase of the cycle that affected the main sector of the Spanish economy, construction sector. Besides this, the five most important sectors of the economy, shown in Figure 4, are characterized by low productivity.

6.3 Robustness

To conclude this section, I intend to establish a comparison between granular residual explanatory power, oil shocks explanatory power and spread explanatory power. This latter calculated as the difference between 3 month *Letras del Tesoro* and 10 years Bond.[23]

Table 7: Granular Residual with industry demeaning, oil shock and spread between short and long term debt coefficients OLS regression. Endogenous variable: $GDP\ Growth_t$. Standard errors are given in parentheses. *Sources*: own elaboration from SABI database.

	1	2	3	4	5	6
(Intercept)	0.0232*	0.0126	0.0312*	0.0307*	0.0552***	0.05208***
	(0.1153)	(0.0092)	(0.0152)	(0.0152)	(0.0157)	(0.0147)
Γ_t	2.6277	4.7393***				
	(1.9663)	(-1.964)				
Γ_{t-1}	-3.3398***	-1.0461				
	(1.013)	(0.0884)				
Γ_{t-2}		-2.6055***				
		(0.5819)				
Oil_t			7.5127e-05	0.0001**		
			(7.5221e-05)	(4.6755e-05)		
Oil_{t-1}			3.2128e-05	3.0502e-05		
			(6.3580e-05)	(7.8359e-05)		
Oil_{t-2}				6.8006e-05		
				(5.8565e-05)		
$Spread_t$					-0.0233***	-0.0233***
					(0.0055)	(0.0047)
$Spread_{t-1}$					0.0101	0.0139*
					(0.0073)	(0.0069)
$Spread_{t-2}$						-0.0039
						(0.0047)
N	14	13	14	13	14	13
R^2	0.2069	0.5633	0.1001	0.2285	0.5816	0.5830
$R^2\ adj$	0.0627	0.4178	-0.0628	-0.0287	0.5055	0.4441

I used the way suggested by Hamilton (2003) to measure oil shocks. This form is based on, from the quarterly series, find the amount that the current oil price exceeds the maximum price reached last year. Then the sum of shocks within each year is done in order to find the annual shock.

As shown in Table 7, the granular residual in the Spanish case has a greater explanatory power than oil shocks, and it almost has the same explanatory power than the spread

[23] Data on short and long term debt come from OCDE database. While data used for calculating oil shocks come from Federal Reserve St. Louis. Furthermore, as prices are in dollars I have switched them to euros, exchange rate has also been obtained from Federal Reserve St. Louis.

between short and long term debt. Therefore we can say, based on a fairly large sample of firms, that there is sufficient evidence to indicate that the Spanish economy is granular.

Finally, note that this analysis could not be performed by gathering all the variables and their lags in the same equation due to the insufficient number of available years. So in principle we do not know how much additional explanatory power we have with the granular residual.

7 Conclusion

This work has analyzed how the Spanish economy is granular in the sense of Gabaix (2011), i.e. large Spanish companies, in particular the 100 largest, represent a large GDP fraction and therefore the shocks that may have these companies can be trasnlated to the aggregate. The explanation of why the idiosyncratic shocks do not cancel out in the aggregate, as was stated in the diversification argument, is because the firm's size distribution does not have thin tails, but fat tails, which makes the law of large numbers does not apply. This leads shocks to decay at the rate $\ln\ N$, instead of $\sqrt{N}$, which is the rate resulting from the application of the Central Limit Theorem. Therefore shocks does not disappear so quickly when there is a power-law distribution (fat tails).

Using the granular residual as a tool to measure shocks, and regressing GDP on it, it has been found that the explanatory power is high, around 36% and 56%, considering the sector each company belongs to. This explanatory power is near, or it is even superior, to traditional measures of aggregate output variability. However, with regard to the explanatory power of granular on TFP, I have shown that the results are antagonistic to those presented by Gabaix (2011). As I mentioned, the cause of these results can be based on the bubble that suffered the main sector of the Spanish economy during the expansion phase of the cycle, which increased GDP but not the productivity of the Spanish economy.

This fact in part explains the limitations of this work, which is conditioned by a fairly limited period of time, where Spain has suffered a phase of large growth and then the worst recession in recent history.

But despite the limitations of the work, the granular residual has proved to be a useful tool, and should be considered in subsequent analyses. And therefore it can serve to draw some interesting implications, as most prominent example, this work shows that systemic risk is not an exclusive problem of the financial sector but it can also be present in other sectors as well. As we have seen, a negative shock in a big firm, such as

Mercadona, could significantly affect the Spanish economy.

Finally, I would like to suggest a line of future research. This could be based on the simulation of the behavior of the Spanish economy as a whole in the presence of granularity, and how it changes in the aggregate when there are potential idiosyncratic shocks.

References

Acemoglu, D., Akcigit, U. & Kerr, W. (2015), Networks and the macroeconomy: An empirical exploration, Technical report, National Bureau of Economic Research.

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

Acemoglu, D., Ozdaglar, A. & Tahbaz-Salehi, A. (2013), The network origins of large economic downturns, Technical report, National Bureau of Economic Research.

Adamic, L. A. (2000), Zipf, power-laws, and pareto-a ranking tutorial', *Xerox Palo Alto Research Center, Palo Alto, CA, http://ginger. hpl. hp. com/shl/papers/ranking/ranking. html* .

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

Bak, P., Chen, K., Scheinkman, J. & Woodford, M. (1993), Aggregate uctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics', *Ricerche Economiche* **47**(1), 3–30.

Blank, S., Buch, C. M. & Neugebauer, K. (2009), Shocks at large banks and banking sector distress: The banking granular residual', *Journal of Financial Stability* **5**(4), 353–373.

Carvalho, V. M. & Grassi, B. (2015), Large firm dynamics and the business cycle, Technical report, CEPR Discussion Paper No. DP10587.

Clauset, A. (2011), A brief primer on probability distributions', *Santa Fe Institute. http://tuvalu. santafe. edu/˜ aaronc/courses/7000/csci7000-001_2011_L0. pdf* .

Clauset, A., Shalizi, C. R. & Newman, M. E. (2009), Power-law distributions in empirical data', *SIAM review* **51**(4), 661–703.

del Rosal, I. (2013), The granular hypothesis in eu country exports', *Economics Letters* **120**(3), 433–436.

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

di Giovanni, J., Levchenko, A. A. & Mejean, I. (2014), Granular comovement, Technical report, Centre for Economic Policy Research.

Dupor, B. (1999), Aggregation and irrelevance in multi-sector models', *Journal of Monetary Economics* **43**(2), 391–409.

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

Fernandez-Villaverde, J., Garicano, L. & Santos, T. (2013), Political credit cycles: The case of the euro zone, Technical report, National Bureau of Economic Research.

Gabaix, X. (2008), Power laws in economics and finance, Technical report, National Bureau of Economic Research.

Gabaix, X. (2009), The granular origins of aggregate uctuations, Technical report, National Bureau of Economic Research.

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

Gabaix, X. & Ibragimov, R. (2011), Rank- 1/2: a simple way to improve the ols estimation of tail exponents', *Journal of Business & Economic Statistics* **29**(1), 24–39.

Gabaix, X. & Ioannides, Y. M. (2004), The evolution of city size distributions', *Handbook of regional and urban economics* **4**, 2341–2378.

Ganugi, P., Grossi, L. & Gozzi, G. (2005), Testing gibrat's law in italian macro-regions: analysis on a panel of mechanical companies', *Statistical Methods and Applications* **14**(1), 101–126.

Gillespie, C. S. (2014), Fitting heavy tailed distributions: the powerlaw package', *arXiv preprint arXiv:1407.3492* .

Hamilton, J. D. (2003), What is an oil shock?', *Journal of econometrics* **113**(2), 363–398.

Hill, B. M. et al. (1975), A simple general approach to inference about the tail of a distribution', *The annals of statistics* **3**(5), 1163–1174.

Hodrick, R. J. & Prescott, E. C. (1997), Postwar us business cycles: an empirical investigation', *Journal of Money, credit, and Banking* pp. 1–16.

Horvath, M. (1998), Cyclicality and sectoral linkages: Aggregate uctuations from independent sectoral shocks', *Review of Economic Dynamics* **1**(4), 781–808.

Horvath, M. (2000), Sectoral shocks and aggregate uctuations', *Journal of Monetary Economics* **45**(1), 69–106.

Hospido, L. & Moreno-Galbis, E. (2015), The spanish productivity puzzle in the great recession, Technical report, Banco de España Working Paper.

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

Kydland, F. E. & Prescott, E. C. (1982), Time to build and aggregate uctuations', *Econometrica: Journal of the Econometric Society* pp. 1345–1370.

Long Jr, J. B. & Plosser, C. I. (1983), Real business cycles', *The Journal of Political Economy* pp. 39–69.

Lucas, R. E. (1977), Understanding business cycles, *in* Carnegie-Rochester conference series on public policy', Vol. 5, North-Holland, pp. 7–29.

Newman, M. E. (2005), Power laws, pareto distributions and zipf's law', *Contemporary physics* **46**(5), 323–351.

Pareto, V. (1896), Cours d" economie politique reprinted as a volume of oeuvres compl etes', *Droz, Geneva* **1965**.

Reichstein, T. & Jensen, M. B. (2005), Firm size and firm growth rate distributions—the case of denmark', *Industrial and Corporate Change* **14**(6), 1145–1166.

Segarra, A. & Teruel, M. (2012), An appraisal of firm size distribution: Does sample size matter?', *Journal of Economic Behavior & Organization* **82**(1), 314–328.

Stanley, M. H., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A. & Stanley, H. E. (1995), Zipf plots and the size distribution of firms', *Economics letters* **49**(4), 453–457.

Summers, L. H. (1986), Some skeptical observations on real business cycle theory’, *Quarterly Review* (Fall), 23–27.

Wagner, J. (2012), The german manufacturing sector is a granular economy’, *Applied Economics Letters* **19**(17), 1663–1665.

Wooldridge, J. (2015), *Introductory econometrics: A modern approach*, Nelson Education.

Zipf, G. K. (1949), Human behavior and the principle of least effort., Technical report, addison-wesley press.

A R codes

In this work we have used the R Studio software for calculating all results presented.[24] I have chosen this software because it is a open software, allowing access to more tools. In addition, the lower complexity of the code and its intuitive interface make data manipulation and calculation faster and easier.

A.1 Power law R code

To show evidence that firm’s size distribution is a power law, I have used two approaches, the log-log regression and MLE estimation. The first, log-log regression, is quite simple in terms of code. As one can see in code 1, I have selected a year, in this case, as an example, it is 2007. Then I have ordered all the companies by their volume of sales, in decreasing order. After that I have assigned a number to each firm, this is the new rank variable. Then, as I am following the approach of Gabaix & Ibragimov (2011), to the rank variable I have subtracted $1/2$. And finally, with regard to data manipulation, I have selected the variables that will be part of the regression.

To make Table 3 I have used different sample sizes, for that it has been necessary to establish a cut off, the purpose of this line 5, which takes the n variable as cutting. And finally line 7 is the regression.

Code 1: log-log Regression

```
1 regresion = spanish_Data %>% filter(year==2007) %>% arrange(desc(Sales))
```

[24]The only exception is Figure 7, where the size distribution of the company is represented. This was made with Stata14.

```
regresion = spanish_Data %>% mutate(Rank=seq(from=1, to=nrow(xj))) %>%
    mutate(Rank_0=Rank-0.5) %>% select(Name,Sales,Rank)

n=10001 # For instance
regresion = regresion %>% filter(Rank < n)

lm(log(xj$rank)~log(xj$Operating.revenue)) #log-log Regression
```

MLE estimation is far more complicated, and it requires a package called powerRlaw. Code 2 is used to calculate the results presented in Table 4 and represent Figure 9. I will not comment on the functioning of each line that makes up the function "MLE_estimation", because the creator of this package has many articles where it is explained. What I think is worth to point out is that input function must be a vector, in this case sales in a given year.

Code 2: poweRlaw package for MLE estimation

```
library(poweRlaw)

MLE_estimation <- function(x,xlabel,ylabel,print=TRUE){
  m_bl = conpl$new(x)
  est = estimate_xmin(m_bl)
  m_bl$setXmin(est)
  p.plot <- plot(m_bl,main="",ylab="",xlab="",pch=16,col="grey")
  mtext(ylabel, side=2, line=2.4, cex=1.2)
  mtext(xlabel, side=1, line=3.3, cex=1.2)
  p.plot <- lines(m_bl, col="black", lwd=2)
  max=max(x)
  cutoff=round(est$xmin,digit=2)
  lmd=round(est$pars,digit=2)
  total=length(x)
  n=sum(x>=cutoff)
  print(p.plot)
  return(p.plot) # Function is configured to show the plot
}

MLE_estimation(Vector,"Sales","P(x)")
```

A.2 Granular R code

To calculate the granular I have created first what I called “input granular residual" (code 3), which it is a function that joins the dataframes top 100 large companies in the year $t-1$ and firms in the year t, regardless of whether they are in the top 100 or not. Each dataframe includes Sales and Employees variables, this is why in the code appear, sales.x $(t-1)$ y sales.y (t). In addition, I have included in the function the calculation of the firm's size relative to the GDP that each company has in the year $t-1$.

As an example the line 10 presents the calculation for the first year using data from the top 100 companies in 1999 and 2000, along with the nominal GDP of 1999.

Code 3: Input Granular Residual

```
1  input.granular=function(topxo,dx1o,gdp1="gdp"){
2    gx=left_join(topxo,dx1o,by="name")
3    gx=select(gx,name,sales.x,employees.x,sales.y,employees.y,sic.x)
4    gx=mutate(gx,gdp=rep(gdp1))
5    gx=mutate(gx,sgdp=sales.x/gdp)
6    gx=gx[complete.cases(gx),] #Leave only the observations that have no "missing values"
7  return(gx)
8  }
9
10 g00=input.granular(top99,p00,gdpn$X1999) #As an example (this is done for each year)
```

Code 4: Granular Residual

```
granular=function(gx){
  zt=log((gx$sales.x)/(gx$employees.x))
  ztt=log((gx$sales.y)/(gx$employees.y))
  g=(ztt-zt)
  gmed=mean(g)
  e=(g-gmed)
  e=ifelse(e< (-0.2),(-0.2),e)
  e=ifelse(e>0.2,0.2,e)
  gr=sum(gx$sgdp*e)
  return(gr)
}

```

```
a00=granular(g00) #A modo de ejemplo

gr=data.frame(a00,a01,a02,a03,a04,a05,a06,a07,a08,a09,a10,a11,a12,a13,a14) #This
    is the granular serie
```

Once I have the input, I created a function to calculate the granular from this (code 4). The granular function first calculates the change in productivity of each company between t (ztt) y $t-1$ (zt), which is g (g_i in the granular equation, 6.1). Then I calculated the average of g, which is $\overline{g}$, being this average subtracted from g. Possible "outliers" are softened, as I has been explained, and the sum of all the shocks of each company multiplied by the relative weight of the economy in $t-1$ is performed. Finally, the serie is created from the data obtained for the granular each year.

Code 5: Granular Residual Industry Demeaning

```
#Industry average
gmed_ind=function(gx){
  m_s= gx %>% group_by(sic.x) %>% summarise(mean(git))
  return(m_s)
}

gmed00_ind=gmed_ind(g00) #As an example

p00=left_join(g00,gmed00_ind, by="sic.x") #It is joined by sic

top.p=function(dx,b){
  dx=mutate(dx,rank=seq(from=1,to=nrow(dx)))
  dx=filter(dx,rank<b)
}

b=101
top00.ind=top.p(p00,b)#a modo de ejemplo

granular=function(gx){
  gmed= gx$`mean(git)`
  e=(gx$git-gmed)
  e=ifelse(e< (-0.2),(-0.2),e)
  e=ifelse(e>0.2,0.2,e)
  gr=sum(gx$sgdp*e)
```

```
25   return(gr)
26 }
27
28 a00=granular(top00.ind) #a modo de ejemplo
29
30 gr=data.frame(a00,a01,a02,a03,a04,a05,a06,a07,a08,a09,a10,a11,a12,a13,a14) #This
      is the granular serie
```

In the case of granular residual with industry demeaning, presented in the code 5, the difficulty increases because the average subtracted to the shock has to be the average of the top 100 firms that belong to the same sector. This is the purpose of the first function. When I have the average for the top 100 companies belonging to the same sector, I put together this average and the input granular, calculated in code 3, using the SIC code. This is the function of line 9. Once done this for every year I set the cut for the 100 largest companies. I create the granular function, under the same parameters as above, but this time I only take into account the companies that have the same SIC code inside the top 100.

B Granular Residual Graphs

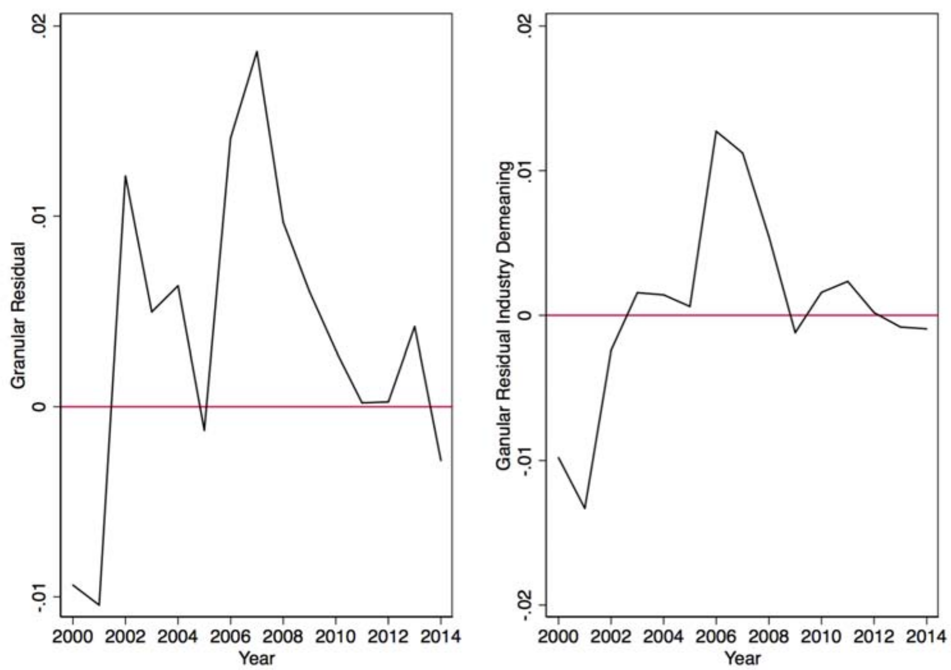

Fig. 11: Granular Residual and Granular Residual Industry Demeaned time series. *Source*: own elaboration.