

THE EFFECT OF FIRMS' SIZE ON BUSINESS CYCLES

BY PHILIPPE C. PEREIRA BORGES



UNIVERSITAT
JAUME I

Abstract: In this paper I analyse the cross-correlation between the effects of firms' size on the business cycle, a phenomenon called granularity. I follow Gabaix's research which argues that idiosyncratic firm-level shocks can explain an important part of aggregate movements and provide a microfoundation for aggregate shocks. The variable used in my research is the operating revenue. I consider this variable as a proxy of firms' size. Furthermore, the aggregate volatility of GDP is determined by the volatility of operating revenue of large firms. Hence, I can predict the GDP volatility via the operating revenue fluctuations of large firms. This paper shows that the forces of randomness at micro level create an inexorable amount of volatility at macro level.

Key Words: granularity, power-law, firm size, black swan and business cycle fluctuations.

Bachelor's Degree in Economics

Professor: Andrea Teglio

Green Card Number: X3630352-D / philippecpb@gmail.com

EC1049 Bachelor's Degree Final Project - Fourth Year

Index

1. Introduction	3
2. The description of the data set	4
3. Firm size distribution	6
3.1. Power-law	7
4. The GDP and the business cycle	8
4.1. Why Gretl?	9
4.2. Hodrick and Prescott	9
4.3. The trend and the business cycle	10
4.3.1 The trend	10
4.3.2 The business cycle	11
4.4. The GDP and growth of large firms	14
5. The impact of the firms size on the GDP	16
5.1. The importance of the black swan theory	18
5.2. Bankruptcy	19
6. Conclusion	21
7. Acknowledgement	21
8. Appendix	22
8.1. Power-Law	22
8.2. Power of each firm	27
9. References	28

1. Introduction

This paper analyzes how the behaviour of the biggest Spanish firms can influence the GDP, a phenomenon known as granularity. In the research done we can observe how the operating revenue fluctuations of the top firms explain an important part of the GDP fluctuations. Our research was carried out using Spanish data. Gabaix (2011) argues that the idiosyncratic level shocks of the firms can explain an important part of aggregate movements and provide a micro-foundation for aggregate shocks. Gabaix, by using American data, shows that the American GDP is driven by the sales of the top 50 and 100 companies. The sales of the top 50 firms form 24% of the American GDP. In this paper I will perform a similar analysis using the Spanish data. The goal, here, is to show a similar significant relationship between Spanish firms' size and GDP fluctuation. Other researchers have also investigated how large companies determine the GDP. Giovanni, Levchenko, Mejean (2014) added variables such as size or age of the company to try to improve the explanation of aggregate fluctuations. Existing research has focused on using aggregate shocks to explain business cycles, arguing that individual firm shocks average out in aggregate shocks. I show that this argument breaks down if the distribution of firm sizes is fat-tailed, as documented empirically. These interconnections provide a mechanism for transmission of shocks across sectors and over time (Duper 1999). The purpose of the model is, therefore, to build up the dependence among agents at the micro-level and to estimate their impact on the macro stability.

In Gabaix's granular hypothesis, (Gabaix 2011), idiosyncratic shocks of large firms have the potential to generate non trivial aggregate shocks that affect GDP. It may be worthwhile to contemplate the possible consequences of the hypothesis that idiosyncratic shocks originating from large firms are an important determinant of the volatility of aggregate quantities. Hence this mechanism might explain a large part of the volatility of many aggregate quantities such as inventories, inflation, short or long run movements in productivity and on the current account. In order to investigate the effect of firms' size on the GDP fluctuation, two ingredients are essential. First we need to prove that Spanish firms are distributed by a Pareto law (a well known distribution which shows that 20% of the population have 80% of the wealth). Second, we need to demonstrate the correlation of the biggest firms on the GDP business cycle.

Granular effects are likely to be even stronger outside the U.S.A, given that the United States is a very diverse country. One figure reported in the literature (Gabaix "The granular origins of aggregate fluctuations" figure 1 page 2; 2011) is the value of

assets controlled by the richest 10 families, divided by the GDP. This gives a new theoretical angle for the propagation of fluctuations. For instance, if Wal-Mart invests in innovative activities, its competitors may suffer in the short run, but then scramble to catch-up. This creates rich industry-level dynamics (that are already actively studied in IO) which could be very useful for studying macroeconomic fluctuations since they allow us to trace the dynamics of productivity shocks. All this helps us to understand better the origins of fluctuations: they do not come from mysterious “aggregate productivity shocks,” but from concretely observable shocks to the large players, such as Wal-Mart, Intel, and Nokia. This could explain the reason why people, in practice, do not know “the state of the economy” — i.e. the level of productivity, in the RBC language. In his view, this is because “the state of the economy” depends on the behaviour (productivity and investment behaviour, among others) of many large firms unknown to people (Gabaix 2009). The effect of firms' size on the aggregate shocks has been investigated in theoretical literature as well. For example, Grilli et al. (2014) have modelled a network economy where the agents' interconnectivity at micro-level can generate strong business fluctuations at macro-level without using external shock. Clearly, many other works deal with these important topics. It is worth noting that all these studies seem to find a similar result: of that the size and interaction among firms can create strong fluctuation at the macro level.

The rest of the paper is organized as follows: Section 2 describes the dataset. Section 3 analyzes the distribution of Spanish firms and the impact of firms on the GDP. The analysis focus the attention on the power - law which is investigated using the best-fit approach applied on the Decumulative-distribution function (DDF). Section 4 analyses the GDP by distinguishing between growth path and business cycle via the Hodrick-Prescott filter. Section 5 shows the impact of the firm's size on the GDP cycle. Specifically, I will analyze on one hand the size of the operating firms on the fluctuation and the impact of the bankruptcies on the GDP. Finally, Section 6 contains the conclusions of our study.

2. The description of the data set

This section will provide us with the explanation of the data set and its source. The panel data chosen was a period between the years 1989 to 2012 (24 years). I include not only a decreasing business cycle such as the crisis of the first years of 1990's

(Japan housing bubble crash and The First Gulf War), but also an increasing business cycle such as the previous years before the Spanish/ American real estate bubble crash. These increasing and decreasing business cycles will help us to show what the Spanish firms' sales behavior is in both periods and whether it has any relationship with GDP fluctuations. Moreover we will see this data seasonally adjusted. The source used for making this paper is SABI (Sistema de Análisis de Balances Ibéricos in its original Spanish name, that we could translate as System of Iberian Accounts Analysis). In this data set we can find the accounts of around 1.4 million of Spanish firm and 4 hundred thousand accounts of Portuguese firms.

The ones chosen to be analyzed in this paper were the companies that reached minimum of 6 million euros of operating revenue along the time series. Using this criterion I found a total of 9824 Spanish firms. This will be the number of my sample (9824). SABI, as most of the data bases, is not perfect. The main problem of SABI is that it does not always offer data for all years. "NA" is written when some annual data is lacking. Therefore, I assumed that firms had not started their operations when this missing value was on the first years (1989, 1990 and so on) and when the missing values were on the last years (2012, 2011 and so on) I understood that these firms had closed. When I found some "NA" among data I assumed that they did not submit their financial reports. In spite of this shortcoming SABI is reliable in term of accuracy of the information. In this data base we find the major 9824 Spanish firms in the time series selected. It is an important sample to show whether the biggest firms have any influence on the GDP in Spain. Each firm's tax code can be also found in the data base. It is important to point out that SABI includes oil firms that operate the Spanish market meanwhile Gabaix (2011) used non-oil firms in his work. In our database the biggest firm was an oil company with an annual average of €11.601 million/year in operating revenue in the 24 years and the smallest company was a renewable energy firm with an annual average of 6 million/year of operating revenue in 2012 (its first year). The average value of the operating revenue for our sample is € 61.933 million/year and the median operating revenue is €18.59 million/year. I can prove how the gap between average and medium values does not allow a normal distribution. Thus, in these cases the average value is well-defined but the variance is not. When this happens, we need to think of the black swan behavior. Also I will show how the black swan behavior affected the operating revenue of the firms in the dataset.

The data set does not consider the costs of each firm nor does it include any firms from the financial sector. The sample will be increased up to 49.106 firms in order to run the bankruptcy analysis (sample B).

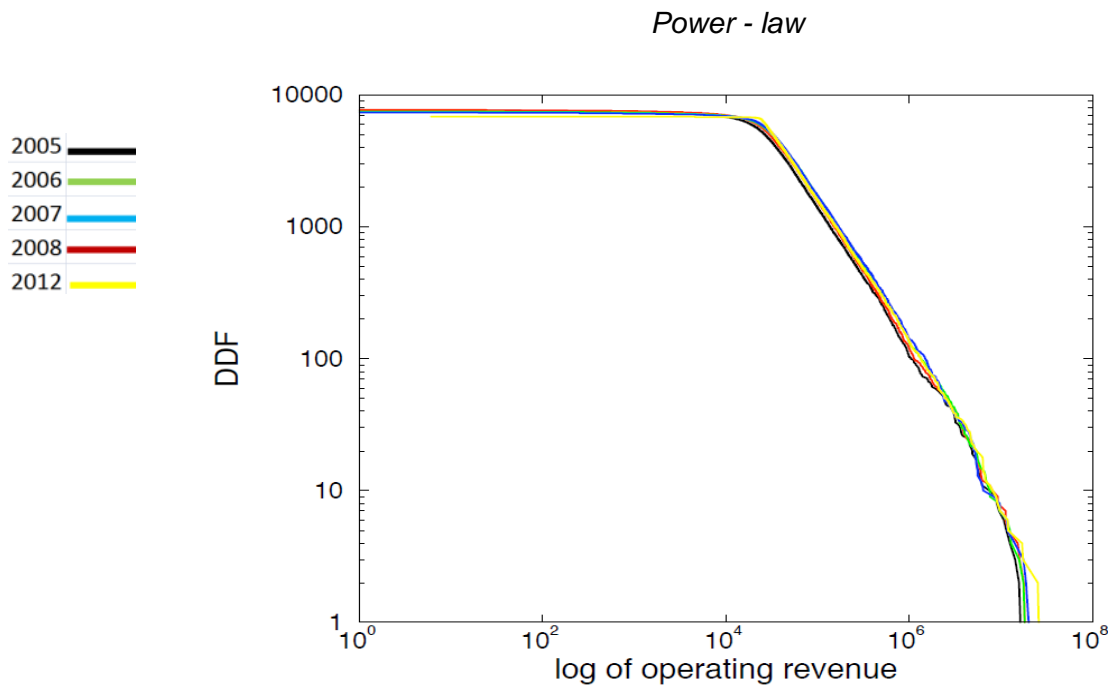
The Spanish GDP data base that I used in this paper was obtained by INE (Instituto Nacional de Estadística, which we could translate such as Statistical National Center, government - dependent institution). This data is showed in market prices and current prices.

3. Firm size distribution

The main goal of this paper is to prove whether the operating revenue fluctuations of large firms have any relation with the GDP fluctuation. Therefore, I decided to separate the 9.824 firms by size. I added all the operating revenues that each firm had earned in all the time series (1989 to 2012) to know what the size of each firm is and these results then determined the ranking of each one. I made four different groups of firms. The first top group has the top 20 firms. The second includes the top 50 firms. It needs to be pointed out that the first group of the top 20 firms will not be included in the other groups as their impact on the GDP will be repeated over time and it is better to analyze their effect separately. The third group includes the top 100 and the fourth the last remaining firms. All groups follow the criteria of non-repetition as we have mentioned before.

Following Gabaix's research (2011) on the economy of USA, one of the most diversified economies in the world, the percentage of their GDP determined by the top 100 firms can explain one-third of the GDP fluctuations. As Spanish market is less diversified than the American market this percentage is higher, 39.11% in the year 2012. It is important to know that a firm's position is determined by the addition of the operating revenue in the time series and it will remain in the same position even if the operating revenue increases or decreases at a certain point in time. I found that the top firms do not change their positions sharply. There are only two possibilities for a firm to decrease from a high position. It could be due to a bankruptcy or a case of a merger.

The Decumulative Distribution Function (DDF) is an empirical proof that shows that these firms follow a Power-law behavior. As it can be seen in the graph below, the result follows the same behavior of power-law regardless of the years. As we can see, each year is represented by a different color: 2005 (black), 2006 (green), 2007 (blue), 2008 (red) and 2012 (yellow).



I have chosen the years 2005, 2006, 2007 to understand the behavior of DDF in this specific moment during in the economic boom. I can prove that, in these years, the biggest firms were the ones that established the GDP fluctuations. In fact, in 2008, the conjecturable features changed but the DDF of that year kept the same behavior, showing that the contribution of these firms to the GDP was still high, more or less 40% of GDP. In 2012, at the height of the crisis, the features did not change the DDF's behavior.

As this picture shows, applying the best fit, this map of points becomes one curve. This is a non-cycle function. It keeps its relationship on both periods, the expansive and contractive one. Thus the best fit to link all these points among them will be using the power-law function. There every point means the operating revenue for each firm for each year.

3.1 Power-law

Firstly, it would be interesting to talk about what a power-law function is and its meaning in the study of the relationship between the operating revenue fluctuations of Spanish biggest firms and GDP fluctuations. In statistics a power-law is a function relationship between two values. One value functions as an exponential of the other one. A power-law distribution fits an approximate value or in some other cases it fits a limited range. The distributions of a wide variety of phenomena fit in a power-law. Some empirical distributions fit a power-law for all their values, but do not follow a power-law in the tail. One of the attributes of the power-law function is scale invariance.

However, it must be taken into account that the power-law usage is only applicable starting with a minimum value (x_{\min}). In such cases we say that the tail of the distribution follows a power-law. A power-law can be turned into a linear relationship if we plot the variables on logarithmic axes. Plotting two quantities against each other in this way is how we generally determine if they have a power-law relationship. Power-laws are very important because they reveal an underlying regularity in the properties of systems. Often highly complex systems have properties where the change between phenomena on different scales is independent of which particular scales we are looking at. The properties of the power-law function are applied on highly complex phenomena where only the scale changes. In many cases when the necessary conditions are fulfilled power-law is used as a best-fit. This self-similar property underlies power-law relationships.

The power-law function is the following:

$$Y = A * x^{-\alpha} \quad (1)$$

Where A is a constant.

x is the variable.

$$\alpha = \frac{1}{\gamma} \quad 2 < \alpha < 3$$

γ = Regression coefficient (SLOPE)

4. The GDP and the business cycle

In this section, we will explain the GDP evolution according to INE data to obtain a wider context. The Spanish GDP analyzed at time series had fluctuations depending on the business cycle. At these time series the GDP growth rate average was around 2.45% (INE's data). 2000 was the year with the highest growth rate, 5.3%. It must be taken into account that this year was within the period of the housing bubble. The next highest growth rate was in 1995 when the Spanish economy was recovering from the crisis of the 90's. Another year in which strong growth was detected was 1999 which was the first year of the housing bubble. Hereafter, we will talk about the periods with the highest negative growth rates. The worst year from these economic periods was 2009 where the GDP growth was -3.6%, followed by 2012 when the negative growth

rate was -2.10%. These years locate in the period where the housing bubble crashed. If we look at the political period, we can see that the best and the second worst year for growth both match with Popular Party (Tory) government. The absolute worst and the second best year occurred with PSOE (Labour) government. Down below, the reason for choosing Gretl software and the Hodrick and Prescott filter will be also commented. Eventually, we will separate the trend and the business cycle.

4.1 Why Gretl?

We have chosen to use Gretl since it is a cross-platform software package for econometric analysis and written in the C programming language. Also, it is free, open-source software and it accepts Excel format which was our data set format. The most important reason for using Gretl was that I was already familiar with the software and had previous knowledge about it which I learned on my third college year at Econometrics I.

4.2 Hodrick and Prescott

Cogley and Nason (1995) Macroeconomic time series often have an upward drift or trend which makes them not stationary. Since many statistical procedures assume stationarity, it is often necessary to transform data before starting an analysis. To analyze this time series (1989-2012) we had to filter GDP data and operating revenue data. In both cases we used the Hodrick and Prescott filter.

Hodrick and Prescott (1980) is a mathematical tool used to eliminate the cycle part. Using this tool we can differentiate between short run (business cycle) and long run (trend) components.

Let y_t for $t = 1, 2, \dots, T$ display the log for variable time series. The y_t is formed by a part that represents the trend, denoted by τ , and a cyclical portion, c so that $y_t = \tau_t + c_t$. We have appropriately chosen a positive value for λ . The trend component is calculated using the following equation:

$$\min_{\tau} \left(\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right). \quad (2)$$

Here $c_t = y_t - \tau_t$ represent the sums of deviations that affect the cycle part. The term λ is included to smoothen the cycle so that we can see clearly the trend.

Higher the λ , higher the negative penalization on the cycle. Hodrick and Prescott (1980) used $\lambda=1600$.

In this paper we have chosen the same value of λ in order to be able to compare the results with other papers and also because this is the standard value to work with Hodrick and Prescott filter. M. Pedersen (1999) shows that the Hodrick and Prescott filter with the standard value of $\lambda=1600$ is in many cases less distorting than other filters. Stock Watson (1999) said that business cycles do not hold out more than 32 quarters. M. Pedersen (1999) stated that if the cyclical component of seasonally adjusted time series is defined as cycles with duration less than 32 quarters, the use of the Hodrick and Prescott filter is recommended.

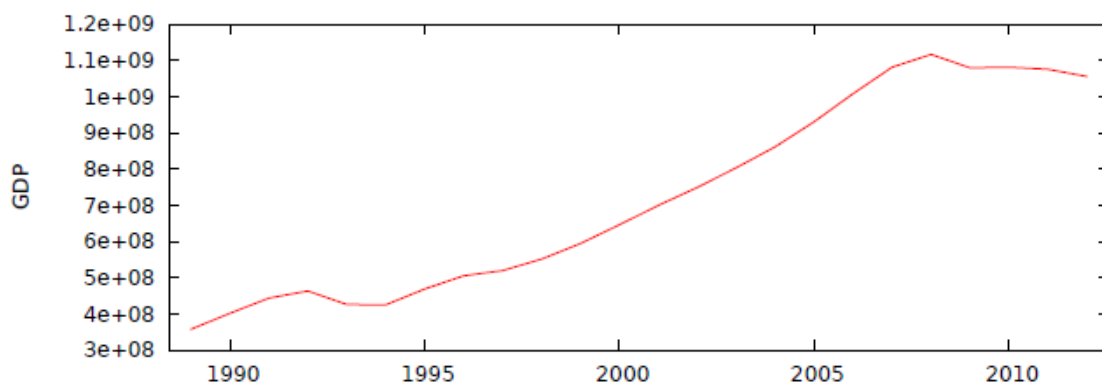
4.3 The trend and the business cycle

4.3.1 The trend

It must be taken into account that the following data was taken from the Spanish Statistical National Center. INE website could be checked to compare our analysis with the original data. We used the Hodrick and Prescott filter to work on this data. We got consistent results following the Hodrick and Prescott criteria and these results can be compared with the results obtained in other papers.

The trend is the gap between the potential GDP and the business cycle. We could say that the trend is the growth path that each economy follows. To separate the two in a clear way, the trend must be limited to a curve containing at most one extreme within the given data span (Wu, Huang, Long and Peng 2007).

The GDP growth path



The Spanish GDP trend was positive and followed the growth path as we can see on the graph 2. In this time series (1989-2012) its growth, thanks to the profits of the period of economic boom caused by the real estate bubble, was very strong in both real and nominal terms. Due to this constant increase Spanish economy has converged to the other European economies. Consequently, if this economy follows a granular behavior, we could conclude that each year the operation revenue of each analyzed firm would increase.

The volatility of sales of the biggest firms explains the volatility of the GDP. In this way, the biggest firms could boost Spanish GDP if being prioritized by the Spanish Government. The main problem is found on the maximum quantity of employees that these firms employ. This means that their operating revenue is a reflection of the aggregate consumption behavior. Top firms do not have the capacity to employ an important percentage of the labour force since they are capital intense. These firms represent approximately 40 percent of the Spanish GDP in 2012, but they only employ a minor part of the national labour force. If the government decided to boost only the top firms, the activity of smaller firms would suffer due to the fact that the majority of consumers do not work for these top firms. The top firms will then also suffer, as their operating revenue goes down due to the decrease in aggregate consumption. Therefore, governments must realize that favoring the activity of the top firms could harm the aggregate output. The Spanish GDP follows the above mentioned behaviour. As shown on the empirical evidences, it grows over time and follows its growth path, which is determined by its previous growth rates.

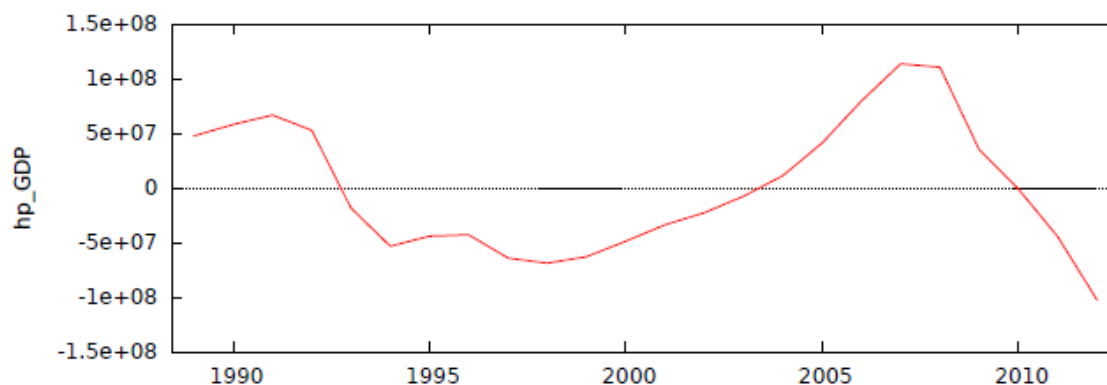
4.3.2 The business cycle

In this section we will explain the behaviour of Spanish business cycle in our time series. We determined the output below using the Gretl cyclic filter. The business cycle could be defined as the GDP's up and down movements in its own long-run growth path. These fluctuations on its own growth path just go towards one direction. On the one hand, when a business cycle growth is positive every year on a period of approximately seven years, it is called an economic boom. On the other hand, when the business cycle growth is negative every year on a period of approximately seven years, it is called a recession.

In our data, over one thousand startup firms emerged in 1991 and 1992. In 1993 this figure stayed slightly below one thousand (961 of the firms were established that year).

Below we can see the different business cycles which the Spanish economy has faced in our analyzed time series (1989-2014).

The Business Cycle



Prior to Keynes' General Theory, the study of these rapid fluctuations, the business cycle combined with the attempt to reconcile the observations with an equilibrium theory was regarded as the main outstanding challenge of economic research (Hodrick and Prescott 1997).

During the period of 1986-1992, a housing bubble increased the housing prices but it did not change the quantity of houses built in a dramatic way. Some experts state that the economic boom started in 1985, when the baby-boom generation started working and therefore increased the labour force. This also raised the demand on the housing market, pushing up the house prices. In real terms the house prices doubled during 1985-1991. However, other reasons can be found to explain this phenomenon. The main reason could be that Spain entered the European Union in 1986 which made the Community market accessible for the Spanish firms. The Spanish government then started a process of market liberalization and privatization of public enterprises.

Moreover, funds such as ERDF (European Regional Development Fund) and the fact that control over macroeconomic policies became stricter in order to fulfill the requirements set by the Euro Zone along with the stability plan, created the ideal conditions to boost the economic growth. All these helped the Spanish economic boom in the first years of this decade.

In 1993, turbulences of Global economy reached Spain. During the first years of the 90's the developed countries suffered from an economic and financial crisis. This crisis started in Japan when its housing market fell down. The consequences of the crisis became worse due to the "First Gulf War" and the increase in oil price, which affected positively the inflation. The effects reached Spain later than other economies.

One of the consequences of these turbulences was a decrease in the aggregate investment level. Investment was very high during 1990-1992 because the country was preparing itself to some great events such as The Universal Exposition in Seville in 1992 (this expenditure includes the AVE, the Spanish High Speed Rail), a new way to link Madrid to Seville. Another important economic impact was Barcelona Olympic Games (1992) and the entire infrastructure that was needed for hosting the games. The Hispasat project was another expensive investment funded by the Public Sector. Due to all this government spending, the public debt level increased immensely. In 1992, the Spanish unemployment level started to rise sharply. On the second Spanish quarter (1992) its GDP drops to 1.1% while the GDP growth rate continues negative (or zero) for the following five quarters until the third quarter in 1993, when it finally starts growing.

The government decided to take action and in 1997 Spanish government set up the Social Security Trust Fund (in its original name Fondo de Reserva de la Seguridad Social) by the Toledo Pact, the goal of which was to protect the future retirement benefits. The Social Security Trust Fund stopped paying the Health Care cost getting a budget surplus as the Health Care cost represented around the 15% of its spending. From 1996 the Spanish economy started an economic boom which lasted one decade until 2007. Its growth rate was even higher than the average growth rate of other European countries. The difference between this policy (Tory) and the previous one (Labour) was the privatization of many of the main public firms such as Argentaria, Telefónica, Endesa and Repsol. The employment created at that moment was mainly because of the new building law and auxiliary industries. Due to all this, the private housing debt increased. Lower real interest rates attracted households to get new credits and to apply for consumer credits and mortgages. The policy followed by Tories was spending less in the public sector and introducing the public firms to the financial markets. The main difference between the two political parties was the difference in their public spending policies.

Hence, the main conclusion is that the liberalization done by the Tories government helped the top firms which made these top firms to become more competitive. As a consequence, their operating revenues increased and this also raised the GDP. It must be taken into account that there is a cross-correlation between their operating revenue and the GDP. When we look at the names of the big firms in our list (data set) we will see that the top firms' positions are occupied by old public firms such as Telefónica, Endesa, Repsol, Iberia, Altadis, Tabacalera or Gas Natural. Thus, the top firms which fluctuate on the financial market and without the support of a

public budget are more competitive. Generally, when there is an increase of competitiveness, their mark-up is lower. Due to this, there should be a decrease in the operating revenue. However, this does not happen in our data set because there is not an increase of competitiveness in the markets. Once the top firms are privatized they keep on holding natural monopolies. The Spanish government does not make laws supporting competitiveness. Thus, these top firms' markup does not decrease. The main change that can be found is that the profits are now private. Consumers become the ones who lose and they cannot see that these top firms' privatization does not benefit them since there is no decrease in price. Following our model, it can be seen that this phenomenon helped the increase of the GDP because of the existing cross-correlation. Another important fact which should not be forgotten is the housing bubble in Spain.

However, just like every bubble, this housing bubble would not be any different and it crashed. During the economic boom around three million jobs were developed. An important part of these new jobs was carried out by foreign people and these immigrants added more growth to the GDP (Solow 1956).

But in 2007 the fall of Lehman Brothers made the Spanish crash even stronger. The international financial crisis blocked the financial markets from Spanish banks. As we can see on the graph, this started the last decrease part of the graph, where we are now. Spanish financial institutions had a strong dependence on foreign 's money flows. The housing market collapsed as a result of the end of these money inflows. This market was the principal part of that made Spanish economy. Since the real estate market received the main part of the investment during the boom cycle, hardly any productive investment was carried out.

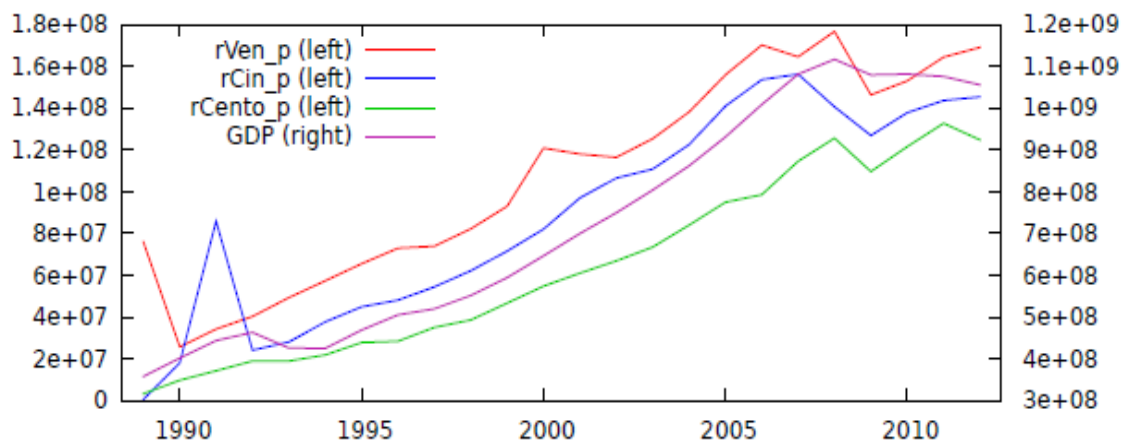
To finalize this business cycle analysis it is important to remember that aggregate economic variables in capitalist economies experience repeated fluctuations from their long-term growth paths (Lucas 1981).

4.4 The GDP and large firms growth

In this chapter I will analyze the evolution of the top firms and the GDP. Firstly I will explain the different lines that we can find in the graph. The main value is the GDP, the purple line, and its scale is on the right. As we can see, there are three different lines each one with their own meaning. Since our dataset was formed by 9824 firms, rVen_p is a line that represents the top 20 firms, rCin_p represent the top 50 firms and rCento_p represent the top 100 firms. It is important to know that each axis has a

different scale. The right axis, which measures the GDP, is ten times the size of the left axis which measures the top firms.

Growth path of GDP Top 20 Top50 and Top100



We can observe that the trends followed by all our firms are the same as the GDP trend. The volatility on the firms is higher than the GDP volatility, due to the fact that the GDP is the aggregation of all these firms and that it includes all firms in the country. In this paper we can affirm that the increase or decrease in GDP has a strong relationship with the top firms in this country (with their operating revenue increase or decrease). We can see that in the crisis of 2007 the movement on top firms' line are very strong, stronger than the GDP's line. This is explained because we are observing the changes from a different scale.

It is important to see that the top 20 companies represent an important fraction of Spanish GDP. Here we must add that in the dataset we worked with we treated each CIF as a different company. That might sound obvious but large firms often use different CIF's in order to decrease their taxation. For example, of the top twenty companies like Repsol and Telefónica each one was mentioned in the list three times, with three different CIFs. For the Spanish Government these are three different companies and hence we treated them as such as well. What we are faced with is that big firms underestimate their size (eg. in terms of operation revenue) to decrease their tax burden.

It is important to bear in mind that the line which represents the top 50 firms is very high in 1991 and 1992, due to the hosting of Barcelona's Olympic Games (1992). In those years we can find many public firms whose business was based on the public spending policies. However, since the work of these firms was finished after Olympic Games, the majority of the firms disappeared soon after.

5. The impact of the firms size on the GDP

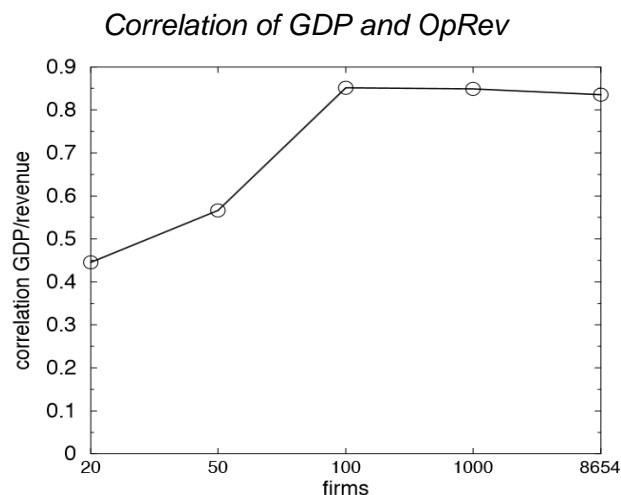
In this chapter we will demonstrate the degree of the correlation between firm size and the GDP. We decided to observe the results from the black swan point of view. The econometric model followed to get this result was the following:

Cross - correlation equation:

$$\rho_{XY}(\tau) = E[(X_t - \mu_X)(Y_{t+\tau} - \mu_Y)]/(\sigma_X\sigma_Y), \quad (3)$$

where μ_X and σ_X are the mean and standard deviation of the process (X_t) , which are constant over time due to stationarity; and similarly for (Y_t) , respectively. The fact that the cross-covariance and cross-correlation are independent in t is precisely the additional information (beyond being individually wide-sense stationary) conveyed by the requirement that (X_t, Y_t) are jointly wide-sense stationary. The cross-correlation of a pair of jointly wide sense stationary stochastic process can be estimated by averaging the product of samples measured from one process and samples measured from the other (and its time shifts). The samples included in the average can be an arbitrary subset of all the samples in the signal (e.g., samples within a finite time window or a sub-sampling of one of the signals). For a large number of samples, the average converges to the true cross-correlation.

Following the hypothesis that all of the operating revenue of each firm is reinvested in itself, we obtain a variable such as proxy of their stock (long run assets). We will use this assumption on bankruptcy analysis. On the cross-correlation equation (X_t) we have the operation revenue in period t and (Y_t) is the GDP in period t .



In our graph we can see the correlation between firms' size and GDP, and then I will separate them in five groups. In the first group I have formed of the top 20 firms, the twenty companies that determine 45% of the Spanish GDP. In other words, these firms shock fluctuations can explain the 0.45 GDP shock fluctuation. When we extend our sample to 50 firms (following the non-repetition criteria), 57% of the GDP stock fluctuations is explained only by these top 50 firms. Once we get to the top one hundred firms (non repetition criteria), these shock fluctuations explain more than the 0.8 of the GDP. All these features mean that we are talking about a granularity economy where a small number of companies can determine the behavior and the fluctuations of the GDP. The horizontal axis indicates the number of firms, and the vertical axis shows the percentage of the correlation between operating revenue and GDP. Many economic fluctuations are attributable to the incompressible "grains" of economic activity, the large firms (Gabaix 2009). Because of random growth on micro level, the distribution of firm sizes is very fat tailed (Simon 1955, Gabaix 1999, Luttmer 2007). That fat-tailedness makes the central limit theory break down, and idiosyncratic shocks to large firms (or, more generally, to large subunits in the economy), affect aggregate outcomes.

Using the next and second last group which consists of one thousand companies, we get a lower effect on GDP since we are applying our non repetition criteria and we exclude the top 20, 50 and 100 firms. The correlation between GDP and operating revenue decreases in spite of the higher quantity of firms. This is another feature that demonstrates that we are dealing with a granularity economy. It is good to bear in mind that in total we have analyzed 9824 firms and the main challenge is to identify the idiosyncratic shocks. Large firms can be volatile because of aggregate shocks, rather than the other way round (Gabaix 2009).

To sum up, in the graph we analyze the impact that each group has over the GDP fluctuations given their own operating revenue fluctuations. In this second analysis (right graph on the appendix 8.2) we have created one average firm that represents the group as one firm and we show how this firm can affect the GDP. Hence, we demonstrate that the top groups have a strong effect on the GDP; the stock fluctuations of top 20, top 50 and top 100 affect the GDP value. With these results we can conclude that we are dealing with a granularity economy. In the granular view, idiosyncratic shocks over large firms have the potential to generate small aggregate shocks that affect the GDP, and via general equilibrium, all firms (Gabaix 2011).

5.1 The importance of black swan theory

The inability to predict outliers implies the inability to predict the course of history, given the share of these events in the dynamics of events. What is surprising is not the magnitude of our forecast errors, but the absence of our awareness of it (Taleb 2007).

The black swan is an event or an occurrence that deviates beyond of what would normally be expected of a situation and that is thus extremely difficult to predict. This term was popularized by Nassim Nicholas Taleb's in his book "The Black Swan: The Impact of the Highly Improbable." He took his title from the shock that Europeans experienced when they discovered black swans in Australia. Until then, their data told them that all swans were white, so the discovery was unexpected. Black Swan logic makes what you do not know far more relevant than what you do know (Taleb 2007). This combination of low predictability and large impact makes the black swan a great puzzle. A black swan in markets is an event that has not occurred in the past, thus rendering useless risk management models based on historic data. Such a risk model would assume that all swans were white. Taleb told the CFA Institute in 2008 that the problem is not that black swans occur often but rather that they have truly catastrophic and unpredictable effects when they do happen, and so risk managers should concentrate on guarding against them.

A small number of black swans explain almost everything in our world, from the success of ideas and religions, to the dynamics of historical events, to elements of our own personal lives. Ever since we left the Pleistocene, some ten millennia ago, the effect of these black swans has been increasing (Taleb 2007). The 2008 crisis was a black swan phenomenon due to the low probability of it happening; the probability to suffer this long and deep financial crash was very low. If the different financial institutions around the world had covered their assets in a more secure way, the crisis would have never become as strong as it did. The point here is the existence of unpredictable facts- the normal facts do not surprise us and so we are protected from them but not from the unpredictable.

The black swan events change the behaviour in the period that it does occur and on the futures periods. A good example of a black swan event was the Lisbon earthquake or the 9/11 attack. Both of these episodes have changed the history mainly because no one was able to predict them.

In business cycle we usually ignore the black swan dilemma since the probability that a big company goes suddenly bankrupt is very low. This is important because in

the next point we will analyse the level of bankruptcies of firms for each year and if it is possible to find evidence of the black swan effect. We will see if one firm has enough power law to drag other firms down with it and whether there exists a wave of bankruptcies.

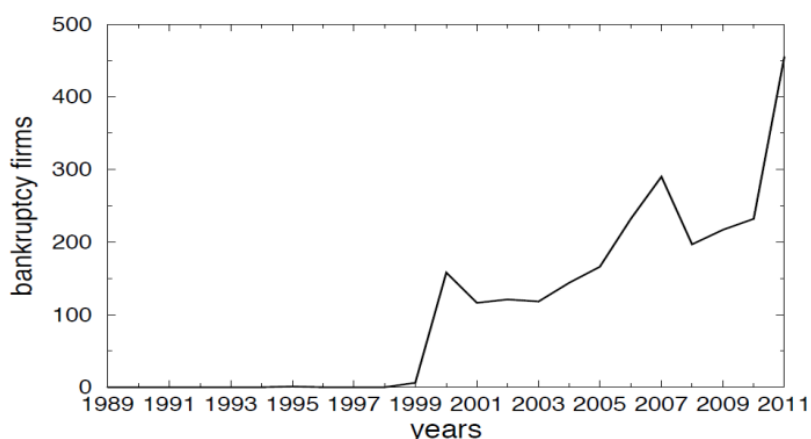
5.2 Bankruptcy

In this chapter we will analyse the bankruptcy behaviour of the firms in our sample. In order to do it we have used two different samples. The first sample that we will call sample A consists of the same 9.824 companies we have used in the previous analysis. The second sample, called sample B consists of 49.106 companies. We will restrict the use of sample B only in this chapter for bankruptcy analysis. The reason we include another sample in our work is that the first sample did not have enough firms that filed for bankruptcy in the period of 1989-2000. In this part we will show the correlation between bankruptcy of large and GDP fluctuations.

Firstly we want to explain the main reasons why sample A alone was simply not enough to create this analysis. The most important thing was that the number of observations was not representative being that the number of bankruptcies was very low. During the first five years of the data the majority of the firms had not yet started their economic activity and filing for bankruptcy was hence impossible. As a consequence the number of firms is not consistent and does not remain the same over the years and this is one of the complexities we face when working with dynamic data. In our database, the firms that maintained a certain position either continued in the same position or closed, one or the other.

In the next picture (sample A) we can see the bankruptcy behavior in all our data set as number of firms closed annually.

Number of firm defaults



The number of firms closed on moment t doesn't have a cross correlation with the GDP in moment t+1, the following period. There is no correlation between the firms that close and the GDP. However, we can detect a correlation between the size of the firms and the GDP: if the firms that go bankrupt are large, they might start a wave of bankruptcies. That's the core idea in this section.

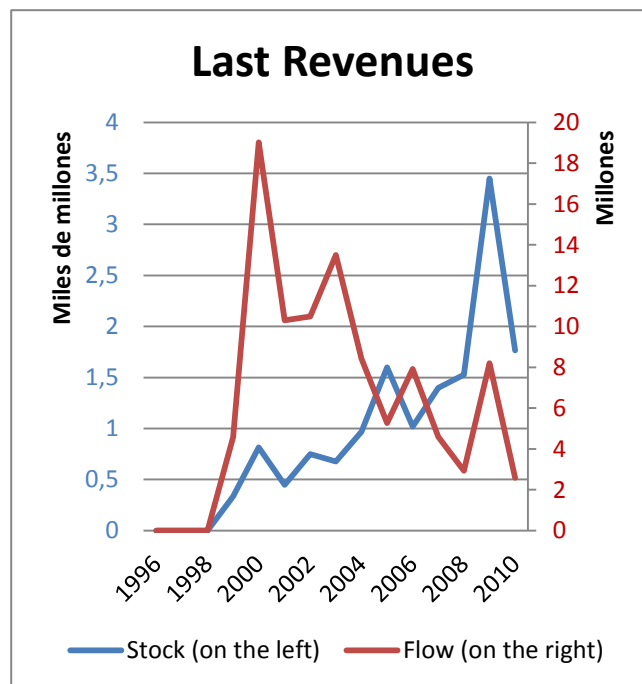
LAG	XCF	
-6	0.0005	
-5	0.0985	
-4	0.1944	
-3	0.3557	
-2	0.5198 **	
-1	0.5402 **	
0	0.2690	
1	0.1253	
2	0.1237	
3	0.1029	
4	-0.0113	
5	-0.0830	
6	-0.0861	

Moreover, we move on to show the cross correlation among GDP_t and $OpRevC_{t-1}$. The $OpRevC_{t-1}$ is the operating revenue of the top closed firms on t moment t-1 and GDP_t is the Gross Domestic Product in moment t (next period). Then, we are going to show whether the GDP is determined by the operating revenue of the year before. The next picture (on the left) called "Cross-correlation function for hp_GDP and hp_sizeMt" tells us that the size of the firms determines the GDP (their size is determined by their operating revenue). hp_GDP is the GDP filtered by Hodrick and Prescott and hp_sizeMt is the Bankruptcies, also filtered by Hodrick and Prescott.

We can see that the correlation between the Operating Revenue (defaulted firms) on period t has a positive cross correlation on the GDP on period t+1. In both cases we ran them using Hodrick and Prescott filter and we have removed the business cycle. The big firms' size on t determines the GDP on t+1. The sizes are determined by operating revenue level. This output means that the big firms have a power - law function, and consequently a small number of companies (by their stocks fluctuations) can determine the GDP stock fluctuations.

To follow with the analysis we have separated the closed firms into two groups according to their last operating revenue and their accumulated operating revenue. We assume that the firms invest in themselves the operating revenues earned in the previous years. We also work with a proxy. The sum of all operating revenues is a proxy that represents the total assets of the closed firms, their size. In the picture on the right we represent the stock accumulation and the stock flow. We can see that in the boom period, the main bankruptcies were filed by small firms, which we can detect from their stock size. However, when the financial crisis started the firms with a high stock started their bankruptcies. The data shows that in the year 2009 the GDP decreased by -3.60% to € 1.079.034 million (the highest decrease in the dataset). This empirical data demonstrates the granularity and the GDP's dependence on the fluctuations of large firms. These fluctuations affect positively on the GDP. This is a

black swan phenomenon because it has truly catastrophic and unpredictable effects when they do take place (Taleb 2007). Now let us analyze the operating revenue flow. We can see that the last operating revenue of these firms had different behavior during the boom and during the crisis. However, during the crisis we can see how a lot of small firms closed (look at the figure "Number of firm defaults"). They do not affect the GDP, but they are affected by the GDP cycle. Due to the correlation between GDP and small firms (GDP determines the behavior of small firms) in the decreasing cycles, we can find a lot of firms closed, but with low last operating revenue. With this other plot we can say that Spanish economy follows an idiosyncratic behavior and the big firms are like grains in the economy. In the granular hypothesis, idiosyncratic shocks from large firms have the potential to generate non trivial aggregate shocks that affect GDP (Gabaix 2004).



6. Conclusion

We can conclude that the sales of the top firms have a significant and positive cross-correlation with the GDP given the empirical evidence of our research. The GDP in t is determined by the operating revenue of the top firms in $t-1$. Idiosyncratic shocks from large firms have the potential to generate non trivial aggregate shocks that affect the GDP. We have investigated the explanatory power of “aggregate sales” to understand the swings in macroeconomic volatility. The number of firms that default in $t-1$ is not significant but the size of those filing for bankruptcy is, and it affects the GDP in t .

7. Acknowledgement

I would like to extend my sincerest thanks and appreciation to those patient souls who helped me to accomplish this paper.

I would first like to express my thanks to Mr. Teglio, the professor of my thesis, for his guidance, knowledge, patience and skills. I have wholeheartedly enjoyed the challenge of researching the granularity of Spanish economy. I also want to thank Marko Petrovic; the documents and advice he provided proved to be invaluable in my research. Special thanks is also extended to Mr. Tedeschi for his help, guidance and knowledge and for introducing me to Mr. Gallegati (one of the developers of an asymmetric information theory along with Joseph Stiglitz - Nobel Prize in Economics), who suggested me to follow the non repetition criteria. Gratitude is extended to Mr. Pereda for his technical assistance in obtaining research material. All possible mistakes found remain mine. I would also like to express my gratitude to my two reviewers of English language, Ms. Gual and Ms. Valtonen.

Finally, special recognition goes out to my family and my friends, for their support, encouragement and patience during my pursuit of Bachelor's Degree in Economics.

8. Appendix

8.1 Power-Law

Below we have the empirical results of the years represented in the DDF graph and the confirmation if these are or not power-law. Starting with 2005 and finishing with the last period 2012, among them we have the years 2006, 2007 and 2008.

Output of 2005

The contrast done was

H₀: power – law

H₁: no H₀

In this case (2005) $\alpha \approx 2$ $\alpha = 1 / -0.5269183$

Number of observations = 7655

Mean of independent variable = 10.50322

Mean of dependent variable = 7.943818
 Standard dev. of ind. variable = 1.536636
 Standard dev. of dep. variable = 0.9969476
 Correlation coefficient = -0.8121605
 Regression coefficient (SLOPE) = -0.5269183
 Standard error of coefficient = 0.004326906
 t - value for coefficient = -121.7772
 Regression constant (INTERCEPT) = 13.47816
 Standard error of constant = 0.04593018
 t - value for constant = 293.4488

Analysis of variance

Source	d.f	Sum of squares	Mean Square	F
Regression	1	5017.84	5017.84	14829.68
Residual	7653	2589.505	0.3383647	
Total	7654	7607.345		

$$y = 7.1366e+05 * x^{-0.52692}$$

Regression of set 0 results to set 1

1-pvalue≈0

Using the criteria of p-value we cannot reject H_0 = power-law. With this contrast we have demonstrated that the biggest firms had power-law on the year 2005.

Output of 2006

The contrast done was

H₀: power – law

H₁: no H₀

In this case (2006) $\alpha \approx 2$ $\alpha = 1 / -0.5506543$

Number of observations = 7725
 Mean of independent variable = 10.63041
 Mean of dependent variable = 7.952915
 Standard dev. of ind. variable = 1.484274
 Standard dev. of dep. variable = 0.9969695
 Correlation coefficient = -0.819806

Regression coefficient (SLOPE) = -0.5506543
 Standard error of coefficient = 0.004376809
 t - value for coefficient = -125.8118
 Regression constant (INTERCEPT) = 13.8066
 Standard error of constant = 0.04697856
 t - value for constant = 293.8915

Analysis of variance

Source	d.f	Sum of squares	Mean Square	F
Regression	1	5159.745	5159.745	15828.61
Residual	7723	2517.511	0.3259758	
Total	7724	7677.256		

$$y = 9.9113e+05 * x^{-0.55065}$$

Regression of set 0 results to set 1

1-pvalue≈0

Using the criteria of p-value we cannot reject H_0 = power-law. With this contrast we have demonstrated that the biggest firms had power-law on the year 2006.

Output of 2007

The contrast done was

H₀: power – law

H₁: no H₀

In this case (2007) $\alpha \approx 2$ $\alpha = 1 / -0.5628565$

Number of observations = 7479
 Mean of independent variable = 10.76961
 Mean of dependent variable = 7.920574
 Standard dev. of ind. variable = 1.454986
 Standard dev. of dep. variable = 0.9968909
 Correlation coefficient = -0.8215023
 Regression coefficient (SLOPE) = -0.5628565
 Standard error of coefficient = 0.004518105
 t - value for coefficient = -124.578
 Regression constant (INTERCEPT) = 13.98232

Standard error of constant = 0.04910023
 t - value for constant = 284.771

Analysis of variance

Source	d.f	Sum of squares	Mean Square	F
Regression	1	5015.316	5015.316	15519.68
Residual	7477	2416.256	0.3231585	
Total	7478	7431.572		

$$y = 1.1815e+06 * x^{-0.56286}$$

Regression of set 0 results to set 1
 1-pvalue≈0

Using the criteria of p-value we cannot reject $H_0 = \text{power-law}$. With this contrast we have demonstrated that the biggest firms had power-law on the year 2007.

Output of 2008

The contrast done was

H_0 : power – law

H_1 : no H_0

In this case (2008) $\alpha \approx 2$ $\alpha = 1 / -0.5914615$

Number of observations = 7368
 Mean of independent variable = 10.82662
 Mean of dependent variable = 7.905631
 Standard dev. of ind. variable = 1.405116
 Standard dev. of dep. variable = 0.9968539
 Correlation coefficient = -0.8336948
 Regression coefficient (SLOPE) = -0.5914615
 Standard error of coefficient = 0.004564782
 t - value for coefficient = -129.5706
 Regression constant (INTERCEPT) = 14.30916
 Standard error of constant = 0.04983559
 t - value for constant = 287.1274

Analysis of variance

Source	d.f	Sum of squares	Mean Square	F
Regression	1	5088.243	5088.243	16788.54
Residual	7366	2232.475	0.3030783	
Total	7367	7320.718		

$$y = 1.6383e+06 * x^{-0.59146}$$

Regression of set 0 results to set 1

1-pvalue≈0

Using the criteria of p-value we cannot reject H_0 = power-law. With this contrast we have demonstrated that the biggest firms had power-law on the year 2008.

Output of 2012

The contrast done was

H₀: power – law

H₁: no H₀

In this case (2012) $\alpha \approx 1$ $\alpha = 1 / -0.9885167$

Number of observations	= 6847
Mean of independent variable	= 11.05412
Mean of dependent variable	= 7.832345
Standard dev. of ind. variable	= 0.9871684
Standard dev. of dep. variable	= 0.9966662
Correlation coefficient	= -0.9790967
Regression coefficient (SLOPE)	= -0.9885167
Standard error of coefficient	= 0.002482065
t - value for coefficient	= -398.2639
Regression constant (INTERCEPT)	= 18.75953
Standard error of constant	= 0.02754622
t - value for constant	= 681.0201

Analysis of variance

Source	d.f	Sum of squares	Mean Square	F
Regression	1	6519.097	6519.097	158614.1

Residual	6845	281.3319	0.04110036
Total	6846	6800.429	

$$y = 1.4033e+08 * x^{-0.98852}$$

Regression of set 0 results to set 1

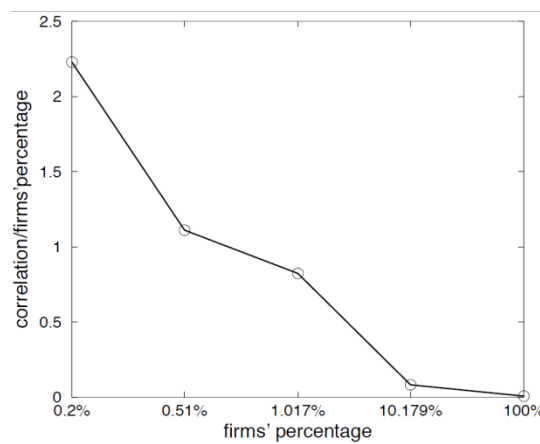
1-pvalue≈0

Using the criteria of p-value we cannot reject $H_0 =$ power-law. With this contrast we have demonstrated that the biggest firms had power-law on the year 2012.

8.2 Power of each Firm

Power of each Firm

In the second graph (on the right) we can find the influence of each firm on the GPD stock fluctuation. On the horizontal axis we can find the percentage that each firm of the group represents over the total of all firms in the sample, and on the vertical axis the correlation divided by the previously mentioned percentage. In other words, the correlation is divided by the percentage that each firm represents



within a particular group (group of 20/50/100/ 1000/9824 firms) over the total number of firms. Thus, each firm that is in the group of the top 20 firms represents 0.2% of the total number of firms. This was done for each of the five groups. With this tool we can then estimate the power of each firm in the sample, and consequently on the stock volatilities of the GDP. When $0.45/0.2 = 2.25$, this 2.25 means that on average each firm that is in this particular group determines the 0.025 of GDP variations. We are talking about power-law companies. If we follow the trend we can see that when we add in more firms, their average power over GDP decreases. Hence, if we consider all the firms in the total sample, an average firm in our sample has zero power to determine the GDP. The power to do so is only held by the big companies, as we can see on the graph (right).

9. References

- Antolin-Diaz, J.; Drechsel, T.; Petrella, I. *Following the Trend: Tracking GDP when Long-Run Growth is Uncertain*.
- Bernard, A. B. and Durlauf, S.N. (1995). *Convergence in International Output*. Journal of Applied Econometrics, Vol. 10, pp. 97-108.
- Carvalho, V. and Gabaix, X., (2013). The Great Diversification and its Undoing. American Economic Review Vol.103: 1697–1727
<http://dx.doi.org/10.1257/aer.103.5.1697>
- Cogley, T. and Nasonb, J. M., (1995). *Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research*. Journal of Economic Dynamics and Control 19, pp. 253-278.
- di Giovanni, J. and Levchenko A. A.,(2012), *Country Size, International Trade, and Aggregate Fluctuations in Granular Economies* Journal of Political Economy, Vol. 120, No. 6 pp. 1083-1132.
- di Giovanni, J; Levchenko A. A and Méjean I.,(2014). *Firms, Destinations, and Aggregate Fluctuations*.
- Dupor, B. (1999). *Aggregation and irrelevance in multi-sector models* Journal of Monetary Economics, Vol. 43, pp. 391-409.
- Gabaix, X. (2009). *Power Laws in Economics and Finance*. The Annual Review of Economics is online at econ.annualreviews.org. This article's doi: <http://10.1146/annurev.economics.050708.142940>
- Gabaix, X. (2011). *The Granular Origins of Aggregate Fluctuations*. National Bureau of Economic Research Working Paper Vol.15286
<http://www.nber.org/papers/w15286>
- Gabaix, X. (2014). *Power Laws in Economics: An Introduction*. Journal of Economic Perspectives. Working Paper.
- Hodrick, R.J. and Prescott, E. C., (1997). *Postwar U.S. Business Cycles: An Empirical Investigation*. Journal of Money, Credit and Banking, Vol. 29, pp. 1-16.

- Horvath, M., (2000). *Sectoral shocks and aggregate fluctuations*. Journal of Monetary Economics, Vol.45 69-106.
- King, R. G and Levine R.,(1993), *Finance and Growth: Schumpeter Might be Right* The Quarterly Journal of Economics, Vol. 108, No. 3 , pp. 717-737.
- King, R.G., Plosser, G.I., Rebelo, S.,(1987). *Production, growth and business cycles I. The basic neoclassical model*. Journal of Monetary Economics, Vol. 21.
- Lucas Robert E, Jr.,(1980). *Methods and Problems in Business Cycle Theory* Journal of Money, Credit and Banking, Vol. 12, No. 4, Part 2: Rational Expectations , pp. 696-715.
- Maravall, A., del Rio, A., (2001).*Time Aggregation and the Hodrick-Prescott Filter*. Banco de España — Servicio de Estudios Documento de Trabajo n.º 0108.
- Pareto, W. F.,(1896). *Cours d'Economie Politique*.
- Pedersen, T. M., (2001) *The Hodrick and Prescott filter, the Slutsky effect, and the distortionary effect of filters*. Journal of Economic Dynamics & Control Vol. 25: 1081-1101.
- Solow, R. M. (1956). *A Contribution to the Theory of Economic Growth*. The Quarterly Journal of Economics, Vol. 70, No. 1, pp. 65-94.
- Taleb, N. N., (2007). *The Black Swan: The Impact of the Highly Improbable*.
- Taylor, J. B. and Williams, J.C., (2008). *Further Results on a Black Swan in the Money Market*. Federal Reserve Bank of San Francisco. National Bureau of Economic Research Vol. 1050. Working Paper 13943 <http://www.nber.org/papers/w13943>
- Tedeschi, G., Grilli, R., Gallegati, M.,(2014). *Markets connectivity and financial contagion*. J Econ Interact Coord DOI 10.1007/s11403-014-0129-1.
- Wagner, J., (2012). *The German manufacturing sector is a granular economy*. Applied Economics Letters, 2012, Vol. 19, pp. 1663–1665.