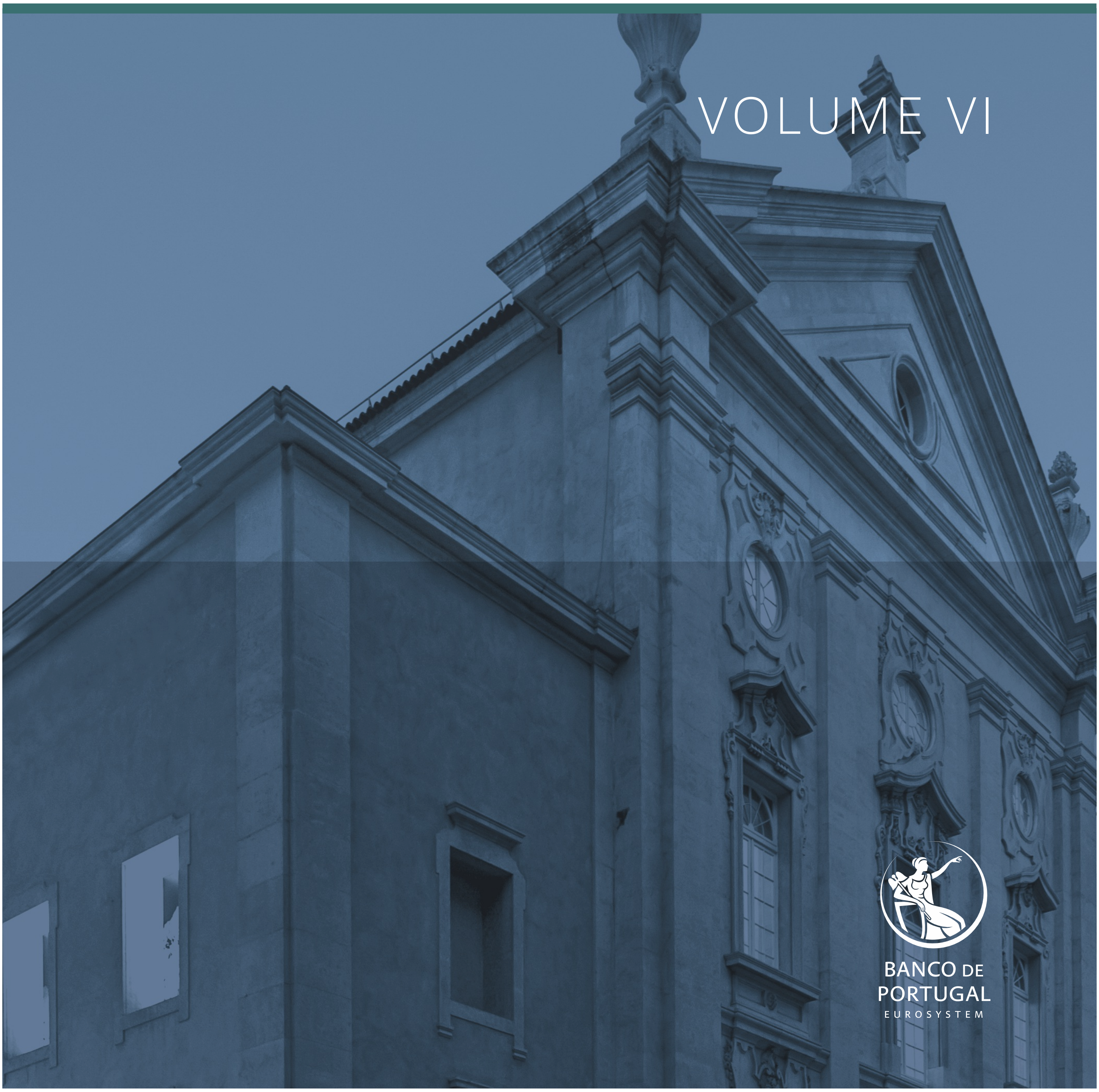
4

BANCO DE PORTUGAL ECONOMIC STUDIES



4

Banco de Portugal Economic Studies

Volume VI

Please address correspondence to
Banco de Portugal, Economics and Research Department
Av. Almirante Reis 71, 1150-012 Lisboa, Portugal
T +351 213 130 000 | estudos@bportugal.pt

BANCO DE PORTUGAL
EUROSYSTEM

Lisboa, 2020 • www.bportugal.pt

Banco de Portugal Economic Studies | Volume VI – no. 4 | Lisbon 2020 • Banco de Portugal Av. Almirante Reis, 71 | 1150-012 Lisboa • www.bportugal.pt • Edition Banco de Portugal • Design Communication and Museum Department | Design Unit • ISSN (*online*) 2183-5217

Content

Editor's note[1]

Pedro Duarte Neves

October 2020

---

1. This issue of the Banco de Portugal Economic Studies contains four studies. The first two use individual data from the Portuguese Household Finance and Consumption Survey and the new *Base de Dados das Maiores Empresas*(Database on the largest firms) for Portugal respectively. The latter, published and analysed for the first time in this issue of the Banco de Portugal Economic Studies, identifies – over a period of nearly forty years – the 200 largest Portuguese firms in each year. The other two studies use aggregate temporal data for the Portuguese economy to forecast housing prices in the first case, and to describe developments in aggregate productivity in the second case.

2. This issue starts with a study by Costa, Farinha, Martins and Mesquita, using data from the Portuguese Household Finance and Consumption Survey and the corresponding survey for the euro area, the Household Finance and Consumption Survey, to characterise self-employed households. The Portuguese Household Finance and Consumption Survey is the result of an already long and fruitful cooperation between Statistics Portugal and the Banco de Portugal, beginning in 1994 with the Household Wealth and Indebtedness Survey, and which today provides a wide range of information on household assets, liabilities, income, consumption and savings, with a very comprehensive description of socio-demographic variables. These surveys have been widely used by the Banco de Portugal's staff over the past 25 years to characterise the distribution of the real assets, financial assets and debt of Portuguese households, contributing to a more informed analysis of issues such as financial stability, monetary policy, and understanding and monitoring of the real economy.

This study focuses on two key aspects: a description of business-owning households in Portugal and in the other euro area countries as a whole, using 2017 (the last year of the survey) as the reference year. Very briefly, below are the main findings of the study:

i) 14% of Portuguese households contain a self-employed professional activity in which at least one member of the household is involved (11% in the euro area);
ii) a regression analysis simultaneously taking into account several socio-demographic characteristics of households concludes that self-employed activities tend to increase

E-mail: pneves@bportugal.pt

1. The analyses, opinions and conclusions expressed in this editorial are entirely those of the editor and do not necessarily coincide with those of Banco de Portugal or the Eurosystem.

with the level of net wealth both in Portugal and the euro area and are also positively related to the existence of debt and the fact that the business was inherited, varying in inverse proportion to the degree of risk aversion;

iii) in Portugal, around 37% of these households are highly dependent on their own businesses – in the sense that they are highly exposed in terms of income earned and exposure through wealth – even reaching approximately 55% for the 40% of households with lower income levels (figures which are close overall to those of the euro area).

Based on data from this survey, it is also possible to assess how different productive structures may react differently to the same adverse shock. The second aspect covered by this study, particularly important in the present situation, has precisely to do with the assessment of the degree of exposure of business-owning households to the COVID-19 pandemic crisis. The study – which uses a classification by the International Labour Organization of the sectors that are more sensitive to this pandemic crisis – shows empirical evidence that, in Portugal, the share of households affected by economic developments related to the pandemic is much higher than the euro area average, both during lockdown and the subsequent phase (56%, compared to 24% during lockdown; 10%, compared to around 5% in the subsequent phase). This helps to better understand the impact of the COVID-19 pandemic on the Portuguese economy both at the beginning and during the ongoing process of gradual recovery of economic activity.

3. The second study in this issue of the Banco de Portugal Economic Studies – by Amador, Lourenço, Magalhães and Pimenta – contains a very comprehensive descriptive analysis of the process of mobility in the largest Portuguese firms over the past four decades. The first aspect to be underlined is the wealth of information in the database developed by the authors, named *Base de Dados das Maiores Empresas* (Largest Portuguese Firms Database) by the authors, and which joins data from several statistical sources and defines data harmonisation procedures. The Database on the largest firms makes it possible to identify the 200 largest Portuguese firms in terms of turnover for each year during the 1981-2018 period, and subsequently analyse the dynamics in this ranking in terms of rises and falls and entries and exits.

The study is essentially descriptive in nature and does not attempt to obtain explanations or causal links for the mobility dynamics observed. In any case, among the results obtained are aspects that are particularly important for the characterisation of the largest Portuguese firms:

i) first, the share of the 200 largest firms in the total value added of the Portuguese economy has declined over the past 25 years – from approximately 15% to around 10% – mostly as a result of a decrease in the share of the 50 largest firms;

ii) second, this study estimates the probability of a firm remaining in the ranking (i.e. in the 200 largest firms) for several periods of time[2]:77% of the 25 largest firms after 10 years (60% after 20 years), 77% for the 26-50 class (51% after 20 years), 20% for the 176-200 class (12% after 20 years);

iii) finally, the authors also present evidence on several cases of direct entries into the ranking (e.g. for the 26-50, 101-125 and 126-150 classes), which is an interesting indicator of mobility.

Due to the lack of similar studies for other countries, it is relatively difficult to conclude how the mobility dynamics in the largest Portuguese firms compare to those of other European countries. A possibility that might be considered in the future is to perform an analysis of concentration indicators at specific sector level. The book *The Great Reversal* by Thomas Philippon, published in 2019, is possibly the most important recent study on this issue, concluding the following for the United States over the past 20 years: "US markets have become less competitive: concentration is high in many industries, leaders are entrenched, and their profit rates are excessive. (...) This lack of competition has hurt US consumers and workers: it has led to higher prices, lower investment, and lower productivity growth". In other words, Thomas Philippon presents an important finding that, over the past 20 years, both the degree of concentration and profit margins have increased in a considerable number of sectors of the United States economy. These developments have been different from those observed in the European Union, given that for the latter the same author shows evidence that overall both the degree of concentration and profit margins remained broadly stable over the past two decades. As a suggestion for a future analysis, it is possible to identify, for the same 20-year time period and for specific sectors, the main characteristics of developments in Portugal both in terms of profit margins and concentration indicators[3].

4. Household residential investment is a key element in the behaviour of business cycles. This point is made particularly clear in the work of the economist Edward Leamer *Housing IS the Business Cycle*, which shows the important empirical finding that, for the United States, eight of the ten recessions since World War II were preceded by substantial problems in residential investment and consumer durables[4].

Purchasing a home is surely a decision households consider very carefully, given its – present and future – importance to their net wealth. Household residential investment is sensitive to several factors, such as the levels and changes in interest rates on housing

2. Figures rounded to nearest percentage point.

3. The Banco de Portugal has published studies on this subject for the Portuguese economy for a shorter time period, 2000-09. See, for example, "Competition in the Portuguese economy: a view on tradables and non-tradables", João Amador and Ana Cristina Soares, Economic Bulletin of the Banco de Portugal, Spring 2012.

4. *NBER Working Paper* No 13428, September 2007.

loans, the level of household indebtedness and housing prices. Future developments in housing prices, in particular, are extremely important for households, given their large weight on household wealth[5]. Developments in housing prices are also key from a financial stability perspective, in particular to assess overvaluations (or undervaluations) in the market, corresponding to deviations from equilibrium taking into account developments in reference economic variables[6]. It is therefore useful for economists to develop econometric models that provide indications on future developments in these prices.

The third study included in this issue of the Banco de Portugal Economic Studies focuses precisely on an econometric modelling developed to forecast housing prices in Portugal and Spain. The study by Hill, Lourenço and Rodrigues attempts to assess how different types of uncertainty may help better forecast future developments in housing prices – in addition to the usual real and financial determinants. For this purpose, they incorporate three different types of uncertainty into the forecasting models: uncertainty surrounding the economic sentiment prevailing in the economy, reflecting consumer and business confidence; uncertainty surrounding domestic financial markets, translated by historical and implied volatility in stock market indices; and finally, uncertainty surrounding the forecasting model using dynamic model averaging.

The authors conclude that for Portugal, all blocks of information – real economy, financial indicators and uncertainty – have a predictive value for housing prices, which intensifies over the period under review. Interestingly, findings for Spain are qualitatively different: although most macroeconomic variables have predictive value, the importance of each predictor remains relatively stable over the period under review. Finally, the authors conclude that uncertainty measures are more important in predicting developments in housing prices for Portugal than Spain. It will certainly be interesting to assess how these forecasting models behaved during the current COVID-19 pandemic crisis and more specifically, how the different types of uncertainty mentioned above contributed (or not) to a better understanding of developments in housing prices.

5. The final study in this issue of the Banco de Portugal Economic Studies warrants a brief introduction. The book *The Rise and Fall of American Growth*, by Robert J. Gordon, published in 2016, is surely one of the most comprehensive analyses characterising economic growth in the United States from 1870 onwards. Robert Gordon has identified a wide range of structural developments – electric lighting, basic sanitation infrastructure, electric appliances, automobiles, air travel, air conditioning, the development of

5. For a recent summary of a number of stylised facts for the housing market (prices and volumes), see "Housing and Macroeconomics", Monika Piazzesi and Martin Schneider, *Handbook of Macroeconomics*, vol. 2, chapter 19, (2016).

6. The December 2019 issue of the Financial Stability Report of the Banco de Portugal includes, as a Special Issue, three alternative methodologies for assessing housing prices, based on statistical indicators, macroeconomic determinants and asset valuation models respectively.

chemicals and pharmaceuticals – which in the author's opinion, were a true economic revolution ("the Great Inventions"), resulting in unprecedented growth in the period 1920-70, leading to a previously unimaginable improvement in the quality of people's lives.

In his assessment, developments over the past four or five decades – especially the revolution in information and communications technology – did not have a comparable effect, explaining why the pace of economic growth and productivity declined during this period (when comparing the 1920-70 and 1970-2014 periods). These developments were observed in the United States as well as in most advanced economies.

The assessment expressed in this book is not particularly favourable in terms of expected developments. Robert Gordon has identified a varied set of recent trends – such as rising inequality in income distribution, stagnating average levels of education and schooling, an ageing population, slowdown in technological progress, increase in the overall debt levels of economic agents – which may result in the standard of living of today's younger generations not exceeding that of their parents.

6. The final study in this issue of the Banco de Portugal Economic Studies, by Amador and Santos, focuses precisely on the issue of economic growth and developments in total factor productivity, covering the 1990-2017 period both for Portugal and the European Union. For this purpose, the authors provide an estimation of an international production frontier, assuming that the various economies have access to the same world technology. On the basis of this estimation, the authors identify the contributions of the build-up of inputs and total factor productivity – corresponding to the share of economic growth that goes beyond the effect of the building-up of inputs (labour and capital) – to the rate of change in gross domestic product. The methodology adopted also breaks down developments in total factor productivity into two components: contribution from technological progress (changes in the frontier) and changes in efficiency (distance to the frontier).

The study covers 28 European Union countries over almost three decades (1990-2017). The authors have chosen the translogarithmic specification of the functional form for the production function, estimated using Bayesian methods. The authors have provided evidence that, because it is more flexible, this modelling of the productive structure has a higher explanatory value than the Cobb-Douglas function, a statistically significant result proven by the fact that posterior distributions of the parameters deviate from zero. However, the economic results obtained by using the two functional forms are qualitatively similar.

The results of the study are not far from the previous characterisations of growth in the Portuguese economy: developments in total factor productivity that are very close to zero over the past three decades – notwithstanding a few positive signals at the end of the period under review – relatively low capital intensity ratios (i.e. capital-labour ratios), low investment levels, and a low contribution from the labour input, owing to the high activity rate and ageing population. In addition, the authors show the following result: the change in efficiency – i.e. the change in the distance to the production function – made a residually negative contribution to total factor productivity over the past decade.

The main trends that might affect developments in the global economy over the next decade – demographic developments, automation, a transition to more sustainable growth, a continued evolution of demand with a stronger preference for digital services, and an increased share of emerging economies in the global economy – will tend to limit average output growth rates both in Portugal and the European Union, and might have very uneven effects on the various sectors of activity and labour skills[7]. Additional gains in terms of economic growth may be the result of the actions of economic agents in general, and economic policy-makers, to the extent that they are able to act on the determinants of economic growth and consequently bring about gains in productivity growth.

Identifying the determinants of economic growth is fundamental to understanding the opportunities for future growth in an economy[8]. The following is a possible classification for the determinants of economic growth: inputs (human capital, investment and financing, corporate management and organisation, innovation), market flexibility (for labour, product and financial markets, in addition to internationalisation) and other determinants (infrastructures, red tape costs, justice and property rights). In this context, the e-book published by the Banco de Portugal in October 2019 – *Portuguese economic growth: A view on structural features, blockages and reforms* – provides an important analysis of the results obtained for Portugal from several of the above-mentioned determinants of economic growth. Some of the results of this study may give economic policy-makers and society in general pause for thought.

This is precisely the challenge taken on by the Banco de Portugal Economic Studies: to contribute to a better understanding of the Portuguese economy and to a better informed debate on economic policy.

7. See Pedro D. Neves, *What do we learn with the e-book "Portuguese economic growth: A view on structural features, blockages and reforms?"*, Spillovers, Banco de Portugal, January 2020.

8. See, in this context, the conference "Portuguese Economic Development in the European Context: Determinants and Policies", organised by the Banco de Portugal in May 2002, which was a pioneer in this subject and included several applied studies on the determinants of economic growth.

Non-technical summary

October 2020

Business owners in Portugal and the euro area: characteristics and exposure to the pandemic

Sónia Costa, Luísa Farinha, Luís Martins and Renata Mesquita

This article characterizes the households owning businesses and their businesses, in Portugal and in the remaining euro area countries. It also aims to identify the households that are more dependent on their businesses and more exposed to the COVID-19 pandemic through business ownership. The analysis is based on data from Household Finance and Consumption Survey (ISFF for Portugal and HFCS for the other euro area countries).

The businesses analysed correspond to non-publicly traded companies or any other self-employed activity. The data for Portugal was collected in the last two waves of the ISFF, which took place in 2013 and 2017, and are comparable with the data collected from the households of the remaining euro area countries in the HFCS.

In Portugal, 14% of the households had at least one business in 2017. In the euro area, this percentage is around 11%.

The characteristics of the businesses and business owners are, in many ways, identical in Portugal and the remaining euro area countries. In a regression analysis having a higher level of net wealth, having debt and having inherited a business stand out as the main factors identifying households with businesses. Most households holding businesses have small businesses (71.2% of the businesses have less than three workers, according to ISFF 2017), in which they have unlimited liability and in which they are sole proprietors. There is however high heterogeneity in the type of business, namely its value, size (number of workers) and age, according to the income, wealth and other socio-economic characteristics of the business owners.

An adverse economic shock may affect income and wealth of the business owners. Businesses are a part of the households' assets. Thus, in case of bankruptcy or a loss of the business value, the value of households assets decline. Frequently, the dependence of the household financial situation on their businesses is not limited to the potential loss of income and assets, but also reflects potential liabilities taken by the household in the business debt. This is the case, for example, in unlimited liability businesses or in case of personal guarantees given for a business loan. Taking these aspects into consideration, the article presents household business dependency indicators. In Portugal, the percentage of households owning businesses while having a financial situation that is very dependent on their businesses is slightly lower than in the euro area (in 2017, 33% in Portugal, against 39.7% in the euro area).[1]

The crisis associated with the COVID-19 pandemic has a higher impact in some sectors of activity. Therefore, the business sector is a relevant factor to assess the exposure of the households' financial situation to the current crisis. In Portugal, the percentage of households with businesses in the sectors more affected by the pandemic crisis is considerably higher than in the remaining euro area countries (among households owning businesses, 56% against 24%, for the group of sectors more affected in the lock-down period). Under these circumstances, despite their slightly lower business financial dependency, the Portuguese households owning businesses are more exposed to the crisis through businesses (18%, against 10% in the euro area). This exposure is particularly high for the lower income households (30% in the two bottom income classes).

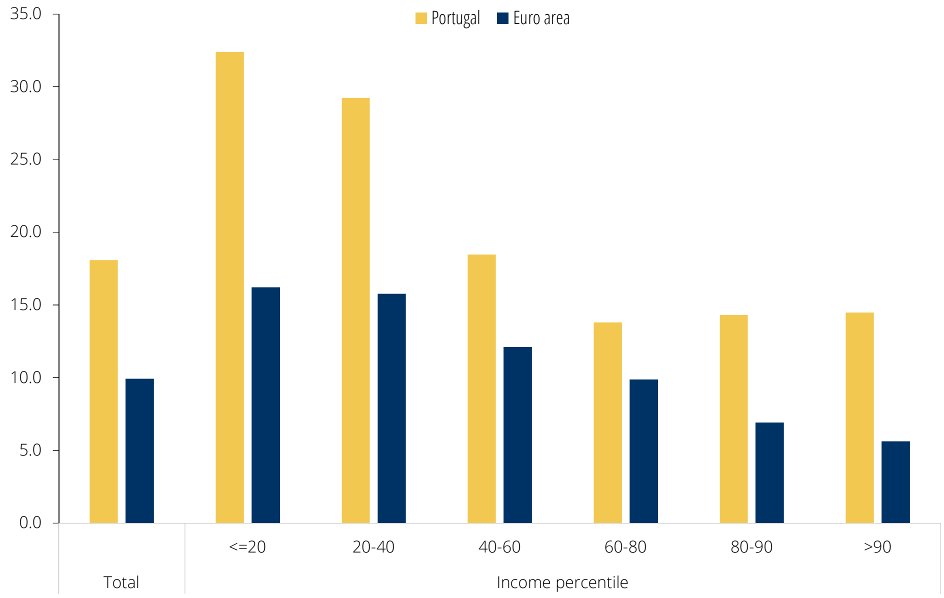


FIGURE 1: Percentage of households more exposed to the pandemic through business ownership, by gross income percentiles

Note: The euro area does not include Portugal and data do not include Spain.

1. A household is considered to have a financial situation very dependent on the business when at the same time it has a very high dependency through income (or labour market) and through wealth. The first is assumed to occur when all household members that work have their activity in the business. When comparing with the euro area, it is assumed that there is a very high dependency through wealth when one of the following situations occurs: the value of the business represents more than 50% of the household assets or the business is of unlimited liability. For Portugal, there is additional information that allows taking into account in the dependency indicator through wealth cases of personal guarantees given to business loans or loans granted by the household to the business. With this additional information, the percentage of households owing businesses that are highly dependent on the business is 36.7% in Portugal, in 2017.

Business owners in Portugal and the euro area: characteristics and exposure to the pandemic

Sónia Costa
Banco de Portugal

Luísa Farinha
Banco de Portugal

Luís Martins
Banco de Portugal

Renata Mesquita
Banco de Portugal

October 2020

Abstract
The interdependencies between households financial situation and the businesses they own are numerous and complex, involving labour market participation, income, assets and debts. These interconnections make business holders particularly exposed to economic shocks. This article, using data from the Household Finance and Consumption Survey, characterizes households' participation in businesses (non-publicly traded companies or other self-employment activities) and analyses households' exposure to the pandemic crisis through business ownership.
(JEL: D10,D31,G30)

1. Introduction

This article uses data from the Household Finance and Consumption Survey (ISFF for Portugal and HFCS for the other euro area countries) to analyse households' participation in non-publicly traded companies or any other self-employed activities of the household members. These companies/activities are referred to as businesses, according to the terminology used in these surveys. These businesses include the activity of sole proprietors and independent workers, as well as any non-publicly traded company that is actively managed by a household member. According to ISFF/HFCS data, in Portugal 14% of households had at least one business in 2017, which compares with around 11% on average in the other euro area countries.

The identification of the factors explaining firm creation and survival is widely present in the literature, what results in large part from the importance of firm creation and growth for innovation and job creation. Haltiwanger *et al.* (2013), for example, document that new firms contribute substantially to job creation in the United States. Despite the high number of firms that start up every year, new firms exit at a significant

Acknowledgements: The authors are thankful for the comments and suggestions of Pedro Duarte Neves, Fátima Cardoso, Nuno Alves, António Antunes and João Amador and also to the participants of the Financial Intermediation Reading Group of the Economics and Research Department of Banco de Portugal for the discussion and suggestions about this topic. The analyses, opinions and conclusions expressed herein are the sole responsibility of the authors and do not necessarily reflect the opinions of Banco de Portugal or the Eurosystem.

E-mail: smcosta@bportugal.pt; lfarinha@bportugal.pt; lpmartins@bportugal.pt; rmesquita@bportugal.pt

rate in their early age. Many studies, for example, Evans and Leighton (1989), Hurst and Lusardi (2004), Levine and Rubinstein (2017), Catherine (2019), Qi *et al.* (2018) or Humphries (2018) associate the decision to start a business and the success of new businesses with observable and unobservable characteristics of business owners, such as age, education, gender, access to finance, managerial experience, intellectual ability, and risk attitude.

In Portugal, little is known about the interactions between the characteristics of firms and their owners, partly because the data necessary to analyse these phenomena is scarce. The ISFF is the only statistical source in Portugal that allows to characterize in detail the financial situation of households both with and without businesses and, at the same time, to characterise those businesses. This article aims to contribute to the knowledge of the interactions between entrepreneurs and businesses, seeking to answer a set of questions. What are the characteristics of household businesses? How do households with and without businesses differ? What types of businesses do different households have? Which households are more financially dependent on their businesses? Which households are most exposed to the Covid-19 pandemic crisis through business ownership?

The characterization of households owning businesses, the type of businesses that they own and the interconnections between them are very relevant aspects in the Portuguese economy. On the one hand, given the small size of most businesses, the role of owners in their financing, through equity or debt, is crucial for their growth and survival. Liquidity constraints can hamper firm creation, survival and growth as shown, for example, by Evans and Jovanovic (1989), Schmalz *et al.* (2016), Fairlie and Krashinsky (2012) and Farinha *et al.* (2019). On the other hand, although businesses are concentrated in a relatively small group of households, they have a very relevant role in the financial situation of these households. The share of the businesses in total assets of business owners is around 43% in Portugal and 31% in the euro area (47% and 34% respectively in net wealth). The interdependence between the business and the financial situation of its owners is complex, comprising various aspects such as the labour status, labour income, distributed profits, ownership of assets and indebtedness. These aspects make households who own businesses particularly vulnerable under an economic crisis. In the current COVID-19 pandemic crisis, this vulnerability is particularly high for businesses in non-essential service sectors, where the contraction is expected to be stronger and more prolonged. The data used in this article makes it possible to identify the households that are most exposed to the current crisis through business ownership in Portugal and the euro area.

The next section describes the data and analyses the characteristics of businesses in Portugal and the euro area. Section 3 compares the demographic and socio-economic characteristics of households with and without businesses. Section 4 begins with an analysis of the type of business that households own in terms of its value, size and age, for different groups of households and then looks at the type of households and businesses in which the interdependencies are strongest. This section also identifies the groups of households most exposed to the pandemic crisis through business ownership. Section 5 presents the main conclusions.

2. What are the characteristics of household businesses?

This article uses data from the ISFF/HFCS to analyse households' participation in businesses. The ISFF is part of the HFCS project of the Eurosystem, which means that the data collected for Portugal is comparable to the remaining euro area countries. These surveys collect detailed information on wealth, debt and income of the households. The surveys also collect information about social and demographic aspects, consumption and savings and behavioural aspects, namely attitudes and expectations. In this article the data for Portugal correspond to the latest two waves of the ISFF, which took place in 2013 and 2017. For the remaining countries, the data collection periods of the last two waves of the HFCS occurred mostly in 2013/14 and in 2017. The ISFF and the HFCS are representative of the households living in each country.[1]

In the ISFF/HFCS the information about households' businesses is organized around three types of business participation: (i) self-employment businesses or non-publicly traded businesses in which a household member has an active role in running the business; (ii) non-publicly traded businesses in which the participation is only through investment; (iii) participation in publicly traded companies. In this article only the first type of participation is analysed since this is the case where the interdependence between the business and the household is higher.

In these surveys the information about businesses includes the sector of activity, the legal form, the number of workers, the identification of the household members that work in the business, the percentage of the participation in the business and the value of that participation at the moment of the interview. This is the value that the household considers it could to sell its business, thus it should reflect its financial situation, including the business' assets and debt. In the case of Portugal, the ISFF also includes the year in which the participation of the household in the business began and the turnover value in the year before the interview.

These surveys also include information about the involvement of the household in the financing of the business. The ISFF collects data about personal guarantees granted to the business, household loans granted to the business and loans granted to the household members to finance the business. The HFCS only collects information about the last type of involvement.

In Portugal, 14% of households had at least one business in 2017, which compares with 12.7% in 2013.[2] In the euro area, the percentage is around 11% in the last two waves of the HFCS, that is, lower but close to that of Portugal.[3] Although businesses are concentrated in a relatively small group of households, they have a very relevant

1. All of the presented results refer to extrapolated values for the whole population, that is, they correspond to the answers of each household in the sample, weighted by the number of households with similar characteristics in the population.

2. Although slightly higher than in 2013, the value of 2017 is not statically different from the previous wave (Costa *et al.*, 2020).

3. In this article, data relating to the euro area, excludes Portugal, and includes all other countries that are currently part of the euro area, with the exception of Lithuania in 2013/14 and Spain in 2017. In Lithuania the HFCS was not collected in 2013/14 and in Spain the microdata from the 2017 edition was not yet

role in the financial situation of their owners. The share of businesses in total assets of business owners in 2017 is around 43% in Portugal and 31% in the euro area (in 2013, 42% and 33% in Portugal and in the euro area, respectively). The share in net wealth is slightly higher (47% in Portugal and 34% in the euro area, in 2017).

Both in Portugal and the euro area, around 88% of households with businesses have only one business. In this article, the data refers to the main business of each household, which in general corresponds to the business with the highest value.[4]

Table 1 presents a description of the main business of households in Portugal and the euro area, based on the data from the last two waves of the ISFF/HFCS.

Both in Portugal and in the other euro area countries, most businesses have less than three workers and are held by a single household. Unlimited liability businesses are predominant but at a lesser extent in Portugal.[5] However, it should be noted that the legal form may not be fully comparable across jurisdictions given national legal specificities. In Portugal, unlimited liability businesses mainly cover independent workers and sole proprietors, while limited liability companies may have one or several partners.

Table 1 presents the distribution of businesses by the six NACE sections that are more frequent in Portugal. In 2017, these sections cover 79% of the businesses in Portugal and 63% in the euro area. Wholesale and retail trade and vehicle repair concentrates the largest number of businesses, that is, around 30% of the total in Portugal and around 15% in the other countries.

In 2017, the median value of the households' participation in the business was 20 thousand euros in Portugal and 30 thousand euros in the euro area. In both cases, this value is very unevenly distributed. In 2017, 25% of these participations were lower than 5 thousand euros, both in Portugal and in the euro area, and 10% worth more than 300 thousand euros in Portugal and 392 thousand euros in the euro area. Similarly, business turnover is also very heterogeneous. The duration of ownership, which like turnover is only available for Portugal, has a median of 11 years and a mean of 13.9 years.

---

available at the date this article was written. The results obtained with a sample of countries common to both periods of time are qualitatively identical to those presented throughout the article.

4. In ISFF/HFCS households are asked how many businesses of this type they own and then it is gathered detailed information about the characteristics of the three businesses with the highest value, while for any remaining businesses it is collected their aggregate value.

5. In this type of businesses the owner may have to use its personal assets when debts are not paid or when the business bankrupts, while in the limited liability case the responsibility of the owner is limited to the business' capital.

	Portugal		Euro area	
	2013	2017	2013/14	2017
Participation in the business (%)				
Median	100.0	100.0	100.0	100.0
Mean	83.2	85.2	87.7	88.9
Value of the participation in the business (EUR, thousands)				
Median	15.2	20.0	30.3	30.0
Mean	168.6	174.0	206.8	192.6
Turnover of previous year (EUR, thousands)				
Median	20.3	25.0	x	x
Mean	768.2	932.2	x	x
Number of workers (% of total business owners)				
1-2	75.9	71.2	75.5	68.8
3-9	17.7	20.6	19.0	26.9
>=10	6.4	8.2	5.5	4.4
Unlimited liability (% of total business owners)	61.1	54.3	78.8	79.4
Duration of ownership (years)				
Median	13.0	11.0	x	x
Mean	15.6	13.9	x	x
Sectors of activity (% of total business owners)				
Agriculture, forestry and fishing	9.0	12.0	13.7	13.8
Manufacturing	9.7	7.9	7.7	5.9
Construction	9.4	9.7	10.0	11.8
Wholesale and retail trade; repair of motor vehicles and motorcycles	26.1	30.1	17.2	14.5
Accommodation and food service activities	13.1	9.5	5.7	5.4
Professional, scientific and technical activities	9.7	10.0	10.5	10.9
Other activities	23.0	20.8	35.3	37.7
Personal guarantees granted to the business (% of total business owners)	13.1	10.8	x	x
Loans granted to the business by the household (% of total business owners)	7.9	6.8	x	x
Loans granted to the household to finance the business (% of total business owners)	6.2	5.1	17.8	10.9
Memo:				
Business owners (%)	12.7	14.1	10.8	10.6
Value of the businesses (% of gross wealth)	42.3	42.8	32.6	31.1
Value of the businesses (% of net wealth)	48.5	47.1	36.1	34.2

TABLE 1. Characteristics of the businesses of the households in Portugal and in the remaining euro area countries

Notes: The euro amounts of 2013 were adjusted for inflation. The euro area refers to the euro area excluding Portugal and the data do not inlcude Lithuania in 2013/2014 and Spain in 2017.

Regarding the involvement of households in the financing of the business, in Portugal, the most common type is to provide guarantees for loans granted to the business, followed by loans granted by the household to the business (in 2017, approximately 11% and 7% of households with businesses had these types of involvement). In 2017 around 5% of Portuguese business owners had taken out loans with the aim of financing their business, which is about half the percentage of households with this type of loans in the euro area.

Between 2013 and 2017, in Portugal, there was an increase in the median and mean values of both the business and turnover, as well as in the percentage of business with 3 or more workers. Additionally, the percentage of households involved in the financing of the business declined. This is in line with the recovery of the economic activity and the improving of the financing conditions of businesses during this period. However, in general terms the characteristics of the businesses do not present substantial differences between the two waves neither in Portugal nor in the euro area.

Finally, it is interesting to note that the characteristics of the businesses described in this section, namely their small size, the predominance of unlimited liability businesses and the involvement of the household in their financing, are very similar to those reported for the United States based on the Survey of Consumer Finances (SCF) (Kennickell *et al.*, 2017).

3. How do households with and without businesses differ?

The purpose of this section is to analyse the relationship between the characteristics of households and their participation in businesses in Portugal and the euro area. This analysis is only possible because the ISFF/HFCS includes data for both households that have businesses and those that do not. However, the data do not include the characteristics of the households in the period before the establishment of the business or in the moment before its closure. Thus, the analysis carried out does not allow to conclude whether the characteristics of the households, namely in financial terms, determine the decision to participate in the business or if they were determined by this decision. It is also not possible to draw conclusions about the role of these characteristics in the success of the business. The analysis is thus essentially descriptive, not intending to establish causal relationships.

The households are characterized in terms of their financial situation, sociodemographics and other aspects that may condition the participation in businesses.

The financial situation is measured by income, net wealth and a variable that identifies the existence or not of debt. Income includes all types of income received by any household member and is the gross amount (i.e., before taxes and social security contributions). Net wealth corresponds to the sum of real and financial assets deducted from all debts.[6] Income, assets and debts of business owners have some components that are directly related to their active participation in businesses. Household income includes income from self-employment and distributed profits. Assets include the value of all businesses owned and actively managed by the household (which is considered a component of real assets in this type of data) as well as any loans the household has made to the business (as a component of financial assets). Debts include any loan granted to the household members whose purpose is to finance the business (a component of debt).

The analysis also considers information on inheritances received by the household any time in the past. Inheritances are relevant to the probability of having a business, not only because the business itself can have been inherited, but also because inheritances can provide individuals the capital needed to start a business particularly in cases of credit constraints.

Sociodemographic variables used in the analysis are age, education level, gender and marital status (married or other). These variables refer to the reference person, who was selected among household members according to the definition of Canberra (United Nations, 2011) that corresponds in most cases to the person with the highest income in the household. Additionally, a variable related to the number of household members aged 16 years or over is included in the analysis. This variable makes it possible to take into account the positive effect of the number of individuals of working age on income, wealth and on the probability of having a business.

The analysis also considers a measure of the degree of risk aversion. This variable aims to capture the fact that investing in a business is risky. This variable includes two categories (lower risk aversion and higher risk aversion), which were built from the answers to a question about the level of risk that the household is willing to take in investment decisions.[7]

In order to take into account potential non-linear relationships between the probability of having a business and income, net wealth, age and number of working age household members, these continuous variables were categorized by classes. Income and net wealth classes were based on the percentiles of the respective variables, considering a narrower breakdown at the top of the distributions, since the empirical evidence suggests higher concentration of businesses in high levels of wealth and income.

Table 2 presents the percentage of households with businesses for the different classes defined according to the variables described above. The first two columns of the table

6. In most countries, including Portugal, net wealth refers to the time of the interview and income to the previous calendar year. The reference periods for variables in the different countries as well as other aspects related to HFCS can be found in HFCN (2016a), HFCN (2016b), HFCN (2020a) and HFCN (2020b).

7. This question includes four possible response options: substantial financial risks, above average financial risks, average financial risks and are not willing to take any financial risk. Households that selected one of the first two categories were classified as having a lower risk aversion and those that selected one of the last two categories were classified as having a higher risk aversion.

include the data for Portugal and the subsequent two columns the data for the euro area (excluding Portugal) for the last two waves of the ISFF and the HFCS.

	Percentage of business owners in each class				Memo: percentage of households in each class			
	Portugal		Euro area		Portugal		Euro area	
	2013	2017	2013/14	2017	2013	2017	2013/14	2017
Total	12.7	14.1	10.8	10.6	100.0	100.0	100.0	100.0
Income percentile								
<=20	3.2	7.1	4.5	4.7	20.0	20.0	20.0	20.0
20-40	7.0	6.1	6.8	6.5	20.0	20.0	20.0	20.0
40-60	11.8	12.6	8.2	8.3	20.0	20.0	20.0	20.0
60-80	13.4	17.5	11.9	11.4	20.0	20.0	20.0	20.0
80-90	25.2	24.3	16.8	15.6	10.0	10.0	10.0	10.0
>90	30.7	30.4	28.0	28.5	10.0	10.0	10.0	10.0
Net wealth percentile								
<=20	3.6	3.6	2.7	2.7	20.0	20.0	20.0	20.0
20-40	7.1	6.9	6.5	5.9	20.0	20.0	20.0	20.0
40-60	9.3	10.9	8.5	8.3	20.0	20.0	20.0	20.0
60-80	12.3	16.0	10.4	11.3	20.0	20.0	20.0	20.0
80-90	22.8	23.7	17.6	15.7	10.0	10.0	10.0	10.0
>90	39.2	43.0	33.5	33.9	10.0	10.0	10.0	10.0
Has debt								
Yes	17.7	19.6	14.8	14.6	45.9	45.7	42.6	40.2
No	8.4	9.6	7.8	7.9	54.1	54.3	57.4	59.8
Received an inheritance (other than a business)								
Yes	15.0	17.1	12.4	11.4	26.7	28.7	25.1	26.7
No	11.8	13.0	10.2	8.6	73.3	71.3	74.9	73.3
Inherited a business								
Yes	40.8	77.1	65.4	46.4	0.3	0.4	0.9	0.3
No	12.6	13.9	10.3	10.5	99.7	99.6	99.1	99.7
Risk aversion								
Lower	22.9	54.1	21.4	23.0	1.2	1.3	4.5	5.1
Higher	12.5	13.6	10.3	10.0	98.8	98.7	95.5	94.9
Number of members ≥ 16 years old								
1	4.4	5.7	4.9	6.2	22.0	25.2	36.7	39.3
2	12.6	15.4	11.9	11.5	48.8	47.3	45.7	44.7
3	17.1	18.4	18.7	16.7	19.6	19.1	11.2	10.5
>3	23.0	22.5	22.1	22.4	9.5	8.4	6.4	5.5
Age								
<35	11.0	14.3	8.1	7.6	11.2	9.8	14.5	15.1
35-44	16.8	21.0	14.3	14.3	20.8	19.3	17.8	16.4
45-54	17.0	20.2	17.4	17.3	20.1	20.3	20.2	20.0
55-64	14.9	15.6	13.7	13.3	18.0	18.4	18.0	18.3
>=65	6.1	5.4	3.6	4.0	29.9	32.2	29.6	30.3
Education								
Less than secondary	10.8	11.1	7.8	7.1	69.4	64.9	31.3	26.5
Secondary	14.2	18.2	10.2	10.5	13.7	15.6	41.5	44.9
Tertiary	18.9	21.1	15.2	14.0	16.9	19.5	27.2	28.6
Gender								
Male	13.9	15.8	12.6	12.5	59.0	58.2	62.8	62.0
Female	10.8	11.8	7.7	7.5	41.0	41.8	37.2	38.0
Married								
Yes	15.7	18.8	14.6	13.7	61.1	55.4	49.8	48.1
No	7.9	8.3	6.9	7.7	38.9	44.6	50.2	51.9

TABLE 2. Percentage of business owners in Portugal and euro area countries, by household characteristics

Notes: The euro area refers to the euro area excluding Portugal and the data do not inlcude Lithuania in 2013/2014 and Spain in 2017.

The participation in businesses across groups of households has a similar pattern in Portugal and in the euro area, as well as in the two periods considered. The percentage of households with businesses increases with the level of income, with the level of net wealth, as well as with the education of the reference person and is higher for households that have debt. By age group, the percentage of households with businesses reaches maximum values for households where the reference person is between 35 and 54 years old and minimum values in those where he is 65 years old or more. The prevalence of business owners increases with the number of working age members and is higher when the reference person is married or when he is male. Having received inheritances, in particular having inherited a business, is also positively correlated with business participation. Finally, the level of risk aversion is negatively correlated with business ownership.

The variables used to characterize the households, shown in Table 2, are correlated with each other, what limits the interpretation of the results of the univariate analysis. In order to better evaluate which are the main household characteristics that distinguish business owners from other households, a multiple regression approach was taken. The results of this approach, which does not solve the endogeneity issues, can be interpreted as correlations between the dependent variable and each explanatory variable conditional on the other explanatory variables.

Logit models were estimated in which the dependent variable takes the value 1 when the household owns a business, and the value zero otherwise. The characteristics of the households were included as explanatory variables. To facilitate the interpretation of the results, the explanatory variables take the form of dummy variables that classify households into different classes, according to each of the characteristics. These variables take the values 1 or zero, depending on whether the household belongs to a certain class or not. Thus, the estimated coefficients must be interpreted as differences relative to the omitted classes.[8] In estimating these models, the sample was restricted to households in which the reference person is under 65 years old in order to focus the analysis on ages that typically participate in the labour market.

The results of the estimation confirm that households that own a business have similar characteristics in Portugal and the euro area, both in 2013/14 and 2017 (Table 3). The results suggest that the likelihood of owning a business increases with the level of net wealth and is higher for indebted households than for households that do not have debt, for households that inherited a business than for those who have not and for households where the reference person is less than 35 years old than for those where he is between 55 and 64 years old. For example, in 2017 in Portugal, a household with net wealth above the 90^{th} percentile was 0.42 percentage points more likely to own a business than a household with net wealth below the 20^{th} percentile. In turn a household in which the reference person is between 55 and 64 years old was 0.12 percentage points less likely to own a business than a household in which the reference person is under 35 years old.

When the value of the business is subtracted from wealth, the marginal effect of wealth continues to show an increasing profile along wealth classes, suggesting that households with businesses are also distinguished from the rest by having higher levels of the remaining assets.[9] Without information on the level of the household wealth prior to the creation of the business, it is impossible to know if business owners are wealthier

8. The omitted categories correspond to the following household groups: first income class (up to the 20^{th} percentile); first net wealth class (up to the 20^{th} percentile); just one member aged 16 years old or older; lower risk aversion, no debt; did not receive inheritances; did not inherited a business; the reference person is less than 35 years old; the reference person has lower than secondary education; the reference person is female and; the reference person is not married.

9. The results of these regressions are not presented in the article for space reasons, but can be made available by the authors upon request. The same applies to other results referred to in this section and which are not part of the basic regression included in Table 3.

because they run a business or if they own a business because they were wealthier beforehand.[10] However, it is interesting to note that the variables that probably capture more exogenously the initial financial situation of the household, such as the level of education and inheritances, are not statistically significant or have an unexpected sign.[11] In Portugal, households whose reference person has tertiary education are less likely to own a business than households whose reference person has a lower than secondary level of education. Thus, although the percentage of households with businesses increases with the level of education, when controlling for the remaining characteristics, more educated households are less likely to own a business.

In the case of debt, the positive correlation with the ownership of a business may reflect the recourse to credit to finance the business. The regression results, however, remain practically unchanged when households with loans taken out for the purpose of financing the business are excluded from the sample. Another possible explanation for the positive relationship between the existence of debt and businesses is the fact that some of the households that do not have debt face credit constraints that hamper the set-up of a business.

For the remaining variables included in the regressions, the results for Portugal and the euro area show some differences. In the univariate analysis, participation in businesses increases with the level of income, in any of the regions and periods. However, when the remaining household characteristics are taken into account, this relationship is no longer evident. In the regression for the euro area, the coefficients on income are significant with a negative sign for middle income classes and not significant for the other classes. The marginal effects in the euro area have a U-shape suggesting that the likelihood of owning a business for households at both extremes of the income distribution is higher than for those at intermediate levels. This might be the result of individuals with bad experiences as employees starting an activity on their own, in which they obtain lower income than employees with similar observable characteristics (Evans and Leighton, 1989). In the other extreme of the income distribution, the result might be associated with the fact that individuals with higher unobservable abilities have a higher probability of earning high income in their own business than as employees (Levine and Rubinstein, 2017).

10. This positive relationship may also reflect a sample selection effect, because when households have a more fragile financial situation, businesses may have a lower probability of survival.

11. In the case of inheritances, the result remains when, as an alternative to the dummy variable, having no inheritance, a variable with the inheritance value updated for the current moment is included in the regression.

	PT 2013		PT 2017		EA 2013		EA 2017		PT 2017-2013	EA 2017-2013	EA-PT 2013	EA-PT 2017
	marginal effect	*t-ratio*	marginal effect	*t-ratio*	marginal effect	*t-ratio*	marginal effect	*t-ratio*	*t-ratio*	*t-ratio*	*t-ratio*	*t-ratio*
Income percentile												
20-40	0.0326	0.83	-0.1189	-2.87	0.0040	0.28	0.0008	0.05	-2.63	-0.15	-0.67	2.62
40-60	0.0568	1.62	-0.0535	-1.48	-0.0240	-1.72	-0.0217	-1.39	-2.22	0.12	-2.22	0.65
60-80	0.0386	1.01	-0.0667	-1.88	-0.0298	-2.16	-0.0281	-1.88	-2.03	0.10	-1.79	0.82
80-90	0.1008	2.26	-0.0139	-0.35	-0.0277	-1.87	-0.0285	-1.72	-1.98	-0.03	-2.85	-0.46
>90	0.0844	1.82	-0.0707	-1.60	-0.0146	-0.98	-0.0043	-0.25	-2.51	0.46	-2.05	1.34
Net wealth percentile												
20-40	0.0479	1.10	0.0884	1.89	0.0991	6.99	0.1000	5.75	0.50	0.01	1.42	0.67
40-60	0.1050	2.39	0.1492	3.47	0.1324	8.92	0.1444	8.05	0.53	0.49	0.97	0.57
60-80	0.1333	2.89	0.2292	5.91	0.1611	11.32	0.1842	9.95	1.28	0.96	1.00	-0.06
80-90	0.2203	4.97	0.2945	7.08	0.2233	14.53	0.2285	12.38	0.78	0.15	0.68	-0.28
>90	0.3171	6.98	0.4184	10.58	0.3186	21.14	0.3349	18.07	1.07	0.57	0.87	-0.14
Has debt												
Yes	0.0500	2.82	0.0331	1.72	0.0489	7.64	0.0470	5.92	-0.78	-0.21	0.28	1.12
Received an inheritance												
Yes	0.0082	0.46	0.0149	0.81	-0.0157	-2.08	-0.0121	-1.48	0.22	0.33	-1.32	-1.43
Inherited a business												
Yes	0.1320	2.40	0.2685	3.53	0.3260	14.90	0.2414	8.85	1.28	-2.49	3.94	0.33
Risk aversion												
Higher	-0.0326	-1.04	-0.1681	-2.84	-0.0397	-3.29	-0.0538	-4.34	-1.93	-0.80	-0.36	1.66
Number of members ≥ 16 years old												
2	0.0314	0.98	0.0153	0.45	0.0309	2.81	0.0004	0.03	-0.39	-1.92	0.10	-0.40
3	0.0486	1.46	0.0027	0.07	0.0429	3.65	-0.0006	-0.04	-0.97	-2.42	0.00	-0.08
>3	0.0553	1.51	-0.0146	-0.35	0.0434	3.45	0.0075	0.53	-1.32	-1.92	-0.16	0.53
Age												
35-44	-0.0024	-0.09	-0.0225	-0.65	0.0078	0.70	0.0093	0.75	-0.44	0.08	0.37	0.90
45-54	-0.0315	-1.08	-0.0623	-1.88	0.0006	0.05	0.0037	0.28	-0.61	0.18	1.01	1.80
55-64	-0.0607	-2.01	-0.1180	-3.37	-0.0383	-3.20	-0.0361	-2.54	-1.09	0.13	0.53	1.88
Education												
Secondary	-0.0376	-1.71	-0.0170	-0.75	-0.0088	-1.21	-0.0092	-1.05	0.72	-0.03	1.18	0.23
Tertiary	-0.0535	-2.12	-0.0644	-2.35	0.0071	0.79	-0.0115	-1.09	-0.15	-1.33	2.22	1.63
Gender												
Male	-0.0093	-0.50	-0.0058	-0.30	0.0114	1.64	0.0227	3.11	-0.50	1.64	1.09	1.59
Married												
Yes	0.0056	0.27	0.0724	3.22	0.0020	0.25	-0.0027	-0.29	2.08	-0.38	-0.15	-2.97
Year 2017 dummy	x	x	x	x	x	x	x	x	2.49	1.36	x	x
Euro area dummy	x	x	x	x	x	x	x	x	x	x	-1.84	-3.77
N	4,371		3,948		47,854		40,395		8,321	88,254	52,227	44,343

TABLE 3. Logit regressions for the probability of owning a business in Portugal and euro area countries

Notes: The euro area refers to the euro area excluding Portugal and the data do not inlcude Lithuania in 2013/2014 and Spain in 2017. The marginal effects correspond to the average marginal effects, calculated in comparison with the ommited categories. The last four columns in the table include the t-ratios of the difference in the estimated coefficients for 2017 relative to 2013 and for the euro area relative to Portugal.

In Portugal the results for income are more similar to the euro area in 2017 than in 2013. The same is true for the effect of the degree of risk aversion. Households with greater risk aversion are less likely to have businesses than those who are less risk averse, in any period, in the euro area, and in 2017 in Portugal. In 2013, in Portugal, the degree of risk aversion does not seem to be related to the ownership of a business. The fact that in 2013 the differences in the results for Portugal and the euro area are larger may be associated with the 2011-2013 strong recession in Portugal that may have encouraged participation in business by households who would not otherwise participate.

Finally, the descriptive statistics in Table 2 show that, regardless of the period or countries considered, participation in businesses increases with the number of working age individuals in the household and is higher when the reference person is married or when he is male. However, in the multiple regression analysis, these characteristics are not consistently identified as relevant to the ownership of businesses.

The above conclusions about the relationship between household characteristics and business participation may differ according to the business type. A relevant feature that distinguishes businesses is the legal form in terms of households' liability: limited or unlimited. In the latter, which are typically smaller businesses, there is no clear distinction between the finances of the household and the business. When the regressions are estimated separately for limited liability and unlimited liability businesses, the positive relation of business ownership with net wealth, business inheritances and with the existence of debt, still holds.[12] The same stands for the negative relationship with risk aversion. However, for education, the negative relationship in Portugal with the ownership of businesses is closely linked to the unlimited liability case, not holding for the limited liability businesses.

4. Relationship between the household and business characteristics

This section explores one of the advantages of the data which is the possibility to link household and business characteristics. The analysis focus household groups defined in terms of income, net wealth, age and education. The first subsection describes the type of businesses these groups have in terms of their value, size and age. The second subsection focuses on the characteristics of the relationship between the household and the business associated with a greater potential business dependency of the household. Finally, the last subsection takes into account the potential dependency as well as the business sector of activity, in order to identify the groups of households that might be more exposed through businesses to the current crisis resulting from the COVID-19 pandemic.

12. These results hold for Portugal and the remaining euro area countries, for the two periods, except for the case of debt in 2013 for Portugal, which was not significant for the unlimited liability businesses.

4.1. *What types of businesses do different households hold?*

As mentioned in section 2, although most households' businesses are small and have low value, they are very heterogeneous. Thus, it is relevant to analyse, for different household classes, what type of businesses they own.

Table 4 includes the value of the business for different household classes as well as indicators related with two important business characteristics: size and age. In each group of households, the size of the businesses is measured by the share of businesses with 10 or more workers. The age of the business is measured by the duration of ownership. Data refers to 2017 and includes Portugal and the remaining euro area countries, with the exception of the age of the business, which is only available for Portugal.

	Business value (EUR, thousands) Median		Businesses with 10+ workers (% of businesses in each class)		Duration of ownership (years) Median
	Portugal	Euro area	Portugal	Euro area	Portugal
Total	20.0	30.0	8.2	4.4	11.0
Income percentile					
<=20	5.0	20.0	1.7	2.7	8.0
20-40	10.0	20.4	5.7	3.6	11.0
40-60	13.0	20.0	7.5	1.1	11.0
60-80	29.1	22.8	3.1	1.9	11.0
80-90	20.0	25.0	10.1	4.1	10.0
>90	54.7	50.0	17.4	9.3	12.0
Net wealth percentile					
<=20	1.0	0.5	3.3	1.3	7.0
20-40	5.0	8.0	7.0	1.3	4.0
40-60	6.9	15.0	4.5	2.7	8.0
60-80	12.0	24.0	1.9	2.6	11.0
80-90	50.0	50.0	9.8	1.4	16.0
>90	105.0	150.0	15.2	9.3	14.0
Age					
<35	30.0	20.0	11.7	3.6	5.0
35-44	15.0	21.5	4.8	2.3	7.0
45-54	24.2	30.0	11.2	7.4	12.0
55-64	37.6	30.0	8.3	3.0	21.0
>=65	11.5	21.0	6.3	3.1	21.0
Education					
Less than secondary	13.9	25.0	6.4	1.1	16.0
Secondary	23.5	30.0	10.8	4.9	7.0
Tertiary	30.0	27.9	9.7	5.3	8.0

TABLE 4. Characteristics of the businesses by household classes in Portugal and euro area countries

Notes: Data for 2017. The euro area refers to the euro area excluding Portugal and the data do not inlcude Spain.

The previous section has shown that households' participation in businesses increases with income and net wealth. The value of the main business of each household also has a positive correlation with these variables. This determines a high concentration of total business wealth in the wealthiest households (Costa *et al.*, 2020). The median value of the businesses is particularly high in the upper classes of income and net wealth. In the top 10% net wealth group, the median value of the business is 105 thousand euros in Portugal and 150 thousand euros in the remaining euro area countries, which contrasts with business values below or equal to 1000 euros in the bottom 20% net wealth group. The households with the highest income and net wealth are also those in which the largest businesses are most common. Additionally, in Portugal the youngest businesses are concentrated in the lowest income class and the oldest businesses in the upper classes of net wealth. It is likely that the percentage of successful and growing businesses is higher among the oldest businesses, that is, among those that have survived the longest. This might contribute for the positive relationship between the financial situation of the households, namely their net wealth, and the value, size and age of the business.

By education level of the reference person, the business value varies more in Portugal than in the remaining euro area countries. In any case, the business value is lower for households with education levels below the secondary. For these households, businesses also have a smaller size and, according to the data for Portugal, are older. The predominance of older businesses in the lower education levels reflects, in part, the sharp negative relationship between age and education existing in Portugal. By age groups, in general terms, the value of the business is higher when the reference person is between 45 and 64 years old, than in the remaining age classes. Additionally, the share of businesses with more than 10 workers is higher for household groups below 55 years old, which, as expected, also have more recent businesses.

4.2. *Which households are more financially dependent on their businesses?*

An economic shock can substantially affect the financial situation of business owners because it can cause a loss in their income, but also in their wealth. Business is part of households' assets. Thus, in case of bankruptcy or a loss in the value of the business, the value of households assets decline. This loss is more significant for the household the greater the share of the business in households assets. In addition, some households grant loans to the business, which means that the value of the business is not the only asset of the household related to the business. Often the dependency on the business is not only a potential loss in the value of the assets but also reflects the potential responsibilities assumed by the household. In the case of unlimited liability businesses, the household is potentially responsible for all business debt. In addition, even in limited liability cases the owners can be responsible for the business debts due to personal guarantees.

Households are defined as highly business dependent through income (or labour market) when all household members that work have their activity in the business. Around half of the households owning businesses, both in Portugal and in the euro area, are in this situation (Table 5). The percentage globally decreases with income, slightly increases with age and does not have a very clear relationship with wealth. In the two lowest income classes, high income dependency is found on almost 70% of Portuguese households owning businesses (around 60% in the euro area).

In the case of dependency through wealth, a household is defined as highly business dependent in one of the following situations: the value of the business represents more than 50% of total assets, the business is of unlimited liability or the household gave a personal guarantee in a loan of the business or granted a loan to the business.[13]

As mentioned in section 2, in Portugal, and even more in the euro area, unlimited liability businesses predominate. In Portugal, the median value of unlimited liability businesses is around 8 thousand euros, less than half the median value of all businesses. This type of businesses is particularly frequent for households in the lower wealth and income classes (more than 75% of households) and for those whose reference person is older and less educated (about 66%). In the euro area the incidence of unlimited liability businesses changes more across types of households.

For about 18% of households with businesses, the value of the business represents at least 50% of their assets, a percentage very similar to the euro area. In Portugal, this percentage is relatively stable across income classes, being slightly higher in the higher income class, and to a lesser extent in the two lower income classes. In the remaining euro area countries, the share of the business value in total assets has a globally decreasing profile with income, with the share in the two lowest income classes being much higher than the average. This indicator of business dependency presents a globally decreasing pattern with the age of the reference person. In Portugal, the business dependency through wealth is very high for the youngest (less than 35 years old), who typically have not accumulated other wealth. For almost 40% of households owning businesses in this age group, the value of the business represents more than 50% of the total assets. By wealth classes there is a higher dependency at the extremes of the distribution than in the intermediate net wealth levels. At the lower end, this finding appears to be mainly due to an effect of the denominator. In turn, in the highest class of wealth businesses have the largest median value.

Among the Portuguese business owners, about 12% have a limited liability business but have given guarantees for obtaining credit for the business or have provided loans to the business. This percentage is higher than 20% for the households with the highest income and wealth, for those whose reference person is less than 35 years old or has tertiary education. These results suggest that it is the households with a better financial situation that can finance the business or grant loan guarantees. Additionally, younger households typically have more recent businesses that do not have a credit history and so require a greater engagement of the owner in its financing.

13. The information about personal guarantees granted to the business and household loans granted to the business is only available for Portugal.

	Dependency through labour market		Dependency through wealth					Aggregate business dependency		
	All workers work for the business		Unlimited liability business		Business value>50% of assets		Limited liability business and personal guarantees/loan to the business	Without guarantees/ granted loans		With guarantees/ granted loans
	Portugal	Euro area	Portugal	Euro area	Portugal	Euro area	Portugal	Portugal	Euro area	Portugal
Total	48.3	49.2	54.3	79.4	17.9	16.7	12.2	33.0	39.7	36.7
Income percentile										
<=20	67.1	63.9	76.5	69.1	17.4	34.9	12.2	57.3	51.3	57.6
20-40	66.9	57.8	73.8	76.4	19.4	30.1	2.2	55.4	48.9	55.9
40-60	41.5	51.5	55.4	77.2	16.1	15.7	5.4	28.5	41.9	29.1
60-80	46.1	51.0	54.2	77.4	16.8	11.0	11.9	30.2	42.4	32.5
80-90	36.4	39.9	50.7	72.6	14.1	14.3	9.1	22.8	30.6	26.9
>90	49.7	42.6	37.9	69.6	23.4	11.2	24.7	27.6	33.1	38.2
Net wealth percentile										
<=20	46.7	44.6	80.0	69.2	16.9	24.5	0.0	38.4	32.0	38.4
20-40	47.3	53.1	72.9	77.3	22.1	27.3	2.3	45.1	43.9	45.3
40-60	45.1	46.3	68.8	81.2	9.8	19.8	3.8	36.2	38.4	36.3
60-80	47.5	46.6	60.7	76.7	9.4	7.6	9.4	33.3	37.5	37.5
80-90	41.4	49.5	56.5	80.0	8.4	9.4	17.0	23.4	43.3	30.2
>90	55.0	51.5	30.5	64.5	32.6	19.8	21.1	31.6	39.8	36.8
Age										
<35	47.4	40.2	28.2	65.1	38.9	24.9	21.5	26.9	28.4	32.9
35-44	41.3	51.4	50.6	71.5	15.8	19.2	15.3	25.2	41.9	29.8
45-54	46.7	44.7	59.1	74.0	16.1	17.3	10.2	33.4	36.1	36.9
55-64	55.8	53.0	57.7	77.6	17.3	12.4	10.8	41.4	43.1	44.5
>=65	56.8	58.4	66.9	77.4	11.1	11.4	4.3	40.9	49.3	42.4
Education										
Less than secondary	48.7	60.3	66.2	80.3	13.7	18.4	5.2	36.0	52.6	37.7
Secondary	42.6	48.6	56.5	77.8	24.2	18.3	13.1	33.2	40.2	34.9
Tertiary	51.6	44.7	31.9	65.8	21.0	14.1	23.8	27.5	32.9	36.2
Memo: business value (median, EUR, thousands)	28.3	30.0	7.9	24.0	155.0	151.6	89.6	20.0	30.0	22.3

TABLE 5. Business dependency indicators for households in Portugal and euro area countries: dependent households in percentage of business owners in each class

Notes: Data for 2017. The euro area refers to the euro area excluding Portugal and the data do not inlcude Spain.

The last column of Table 5 shows an aggregate indicator of business dependency in Portugal that includes all the previous dimensions of dependency. Dependency is very high when both dimensions (in terms of income and wealth) occur simultaneously. According to this indicator, in Portugal 37% of households that own businesses have a financial situation that is highly dependent on the business. This percentage reaches maximum values, above 55%, for the lowest income households. This reflects the high dependence on income, and simultaneously a high dependency on wealth resulting from the existence of unlimited liability businesses. These factors also determine that the percentage of highly dependent households is higher for the older age groups. By level of education or classes of net wealth, the percentage of dependent households is less heterogeneous.

In order to compare data from Portugal with the remaining euro area countries, Table 5 also includes an alternative dependency indicator that does not take into account dependency through guarantees or loans granted by the owner to the business (that are only available in Portugal). According to this indicator, the percentage of very dependent households is 33% in Portugal and 40% in the euro area. It is interesting to note that in Portugal the percentage of dependent households continues to hit maximum values, above 50%, in the lowest income classes, and that the same happens in the remaining countries.

4.3. *Which households are most exposed to the COVID-19 pandemic crisis through business ownership?*

The COVID-19 pandemic forced authorities to adopt containment measures that have disrupted a significant part of the economic activity in many countries. The progressive easing of these measures allowed a gradual restart of the economy. However, in some sectors, due to their specificities, activity remained far below pre-pandemic levels. For the purposes of the analysis in this section, two groups of sectors were considered. The first, broader group, includes the sectors that have been most affected during the lock-down period. This group includes the sectors in which the impact was considered medium-high and high by the ILO in April 2020 (International Labour Organization, 2020). These sectors are Manufacturing, Trade and repair of motor vehicles and motorcycles, Accommodation and food services activities, Real estate activities, Administrative activities, Arts, entertainment and recreation and Transport and storage. The second group, more restricted, covers only Accommodation and food services activities and Arts, entertainment and recreation. These sectors are expected to maintain a very low level of activity given that they require greater physical proximity or are very limited by the constraints created by pandemic to tourism.

In Portugal 56% of business owners have their business in a sector highly affected during the lock-down period. This percentage is more than double that observed, on average, in the euro area (24%) (Table 6). The share of households whose business belongs to the sectors with a slower recovery is 10% in Portugal, twice that of the euro area.

	Households with businesses in more affected sectors			
	Sectors more affected in the lock-down period		Slower recovery sectors	
	Portugal	Euro area	Portugal	Euro area
Total	55.6	23.0	10.5	5.0
Income percentile				
<=20	57.5	28.1	13.8	7.1
20-40	49.5	24.1	11.5	6.5
40-60	61.4	31.1	12.3	5.7
60-80	51.6	23.2	8.2	5.6
80-90	62.0	22.8	11.6	3.8
>90	52.0	21.3	8.8	3.3
Net wealth percentile				
<=20	41.9	22.3	11.2	3.5
20-40	49.6	20.8	13.9	5.3
40-60	60.2	27.0	16.1	9.1
60-80	53.6	28.7	12.5	5.6
80-90	55.7	22.4	8.9	3.6
>90	59.0	22.7	5.8	4.2
Age				
<35	38.4	25.6	4.1	5.4
35-44	52.3	26.3	9.2	6.9
45-54	57.5	25.4	10.2	5.2
55-64	60.3	25.0	13.4	4.7
>=65	65.2	15.3	14.4	3.2
Education				
Less than secondary	60.1	15.3	13.3	3.5
Secondary	62.9	26.3	9.8	5.9
Tertiary	42.7	26.5	6.1	5.3

TABLE 6. Percentage of households with businesses in sectors more affected by the COVID-19 pandemic

Notes: Data for 2017. The euro area refers to the euro area excluding Portugal and the data do not inlcude Spain. The "Sectors more affected in the lock-down period" are: Mining and quarrying, Manufacturing, Wholesale and retail trade, repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities, Real estate activities, Administrative and support service activities and Arts, entertainment and recreation. The "Slower recovery sectors" are: Accommodation and food service activities and Arts, entertainment and recreation.

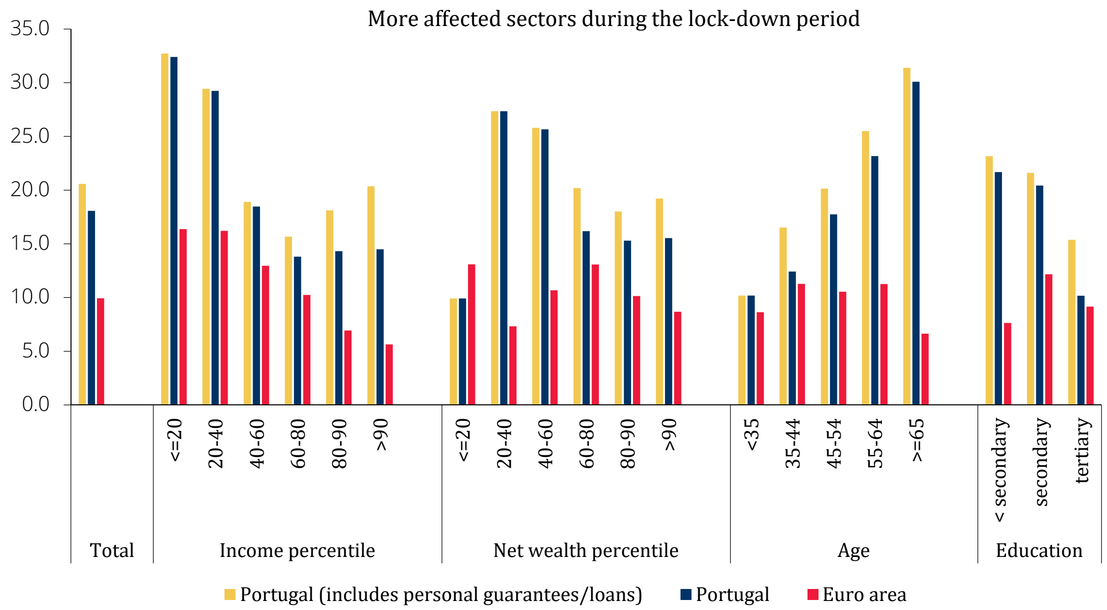
In Portugal, the share of business owners whose business belongs to the group of sectors more affected during the lock-down period increases with net wealth up to the class between the 40^{th} and 60^{th} percentiles. In the euro area, the share of these sectors is higher in the two intermediate net wealth classes. When considering the group of sectors with the slower recovery, the share of households with businesses in the more affected sectors has a more marked profile with income and wealth. In this case, the share of the more affected sectors is decreasing globally with income, both in Portugal and in the other countries of the euro area. By net wealth classes, the share of households with businesses in these sectors increases until the class between the 40th and 60th percentiles and decreases in the subsequent classes.

To evaluate the exposure of business owners to the pandemic crisis, the information on the share of households with businesses in affected sectors is combined with the indicators of business dependency analysed in the previous subsection. A household is considered very exposed to the pandemic crisis if the business belongs to one of the more affected sectors and the household financial situation is very dependent on the business, that is, if both its income and its assets have the risk of reducing substantially with the crisis.

In Portugal, about 18% of households owning businesses were very exposed to the pandemic during the lock-down period, which compares with a percentage of around 10% in the remaining euro area countries (Figure 1). Considering the sectors with a slower recovery (Accommodation and food services activities and Artistic, entertainment and recreation), very exposed business owners are 4.5% and 3% of households with businesses, in Portugal and in the euro area, respectively.[14]

The share of households very exposed to the crisis decreases globally with income, both in Portugal and in the euro area. In Portugal, this share is close to 30% in the two lower income classes, when considering the set of sectors most affected in the lock-down period (almost 10%, when considering the sectors where the crisis is likely to be prolonged). In Portugal, exposure also decreases with net wealth (from the class between the 20th and 40th percentiles) and with the education level, and increases with age. In the remaining countries, the pattern with these variables is less defined. In the euro area, the highest share of households highly exposed to the crisis occurs in the 35-44 age group and in households whose reference person has the secondary education.

14. When the dependency measure also considers household guarantees and loans for the business, information that is only available for Portugal, the share of Portuguese business owner households very exposed to the pandemic is 21% and 4.8%, respectively, when considering the sectors most affected in the lock-down period or the sectors with the slowest recovery.



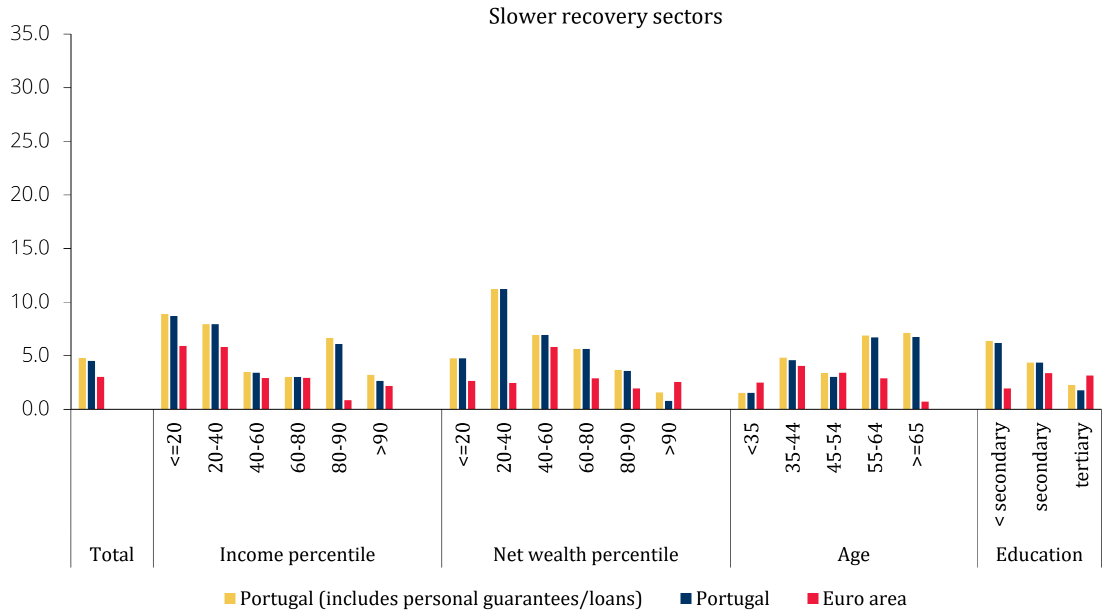


FIGURE 1: Business owners very exposed to the pandemic in Portugal and euro area countries

Note: Data for 2017. The euro area refers to the euro area excluding Portugal and the data do not inlcude Spain. The "Sectors more affected in the lock-down period" are: Mining and quarrying, Manufacturing, Wholesale and retail trade, repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities, Real estate activities, Administrative and support service activities and Arts, entertainment and recreation. The "Slower recovery sectors" are: Accommodation and food service activities and Arts, entertainment and recreation. A household is considered to be very exposed to the pandemic if the household business belongs to one of the more affected sectors and if the financial situation is very dependent on the business.

5. Conclusion

According to data from the Household Finance and Consumption Survey (ISFF for Portugal and HFCS for the other euro area countries), in Portugal 14% of households had at least one business in 2017, which compares with around 11% in the rest of the euro area. Businesses, i.e., non-publicly traded companies or other self-employed activities that are owned by households, have very similar characteristics in Portugal and in the rest of the euro area countries. Most households holding businesses have small businesses, in which they have unlimited liability and in which they are sole proprietors. In Portugal, in 2017, the value of the main business of each household was less than 20 thousand euros for half of the business owners and less than 5 thousand euros for 25% of these households (30 thousand euros and 5 thousand euros, in the other countries, respectively). However, businesses are highly heterogeneous according to the level of wealth, income, age and other characteristics of the owners. In Portugal, households with a higher level of income, a higher level of net wealth or belonging to the age group in which the reference person is between 45 and 64 years old have businesses with a higher value, larger size (in terms of the number of workers) and older.

But what distinguishes households with businesses from other households? In the regression analysis having a higher level of net wealth, having debt and having inherited a business stand out as the main factors identifying households with businesses. Households with a lower degree of risk aversion or those that belong to the extremes of the income distribution are also more likely to own a business.

An economic shock can substantially affect the financial situation of households owning businesses because it can affect their income and also their wealth. In the current context, given the heterogeneity of the effects of the pandemic crisis across different sectors, to assess household exposure to the crisis it is important to take into account not only the household financial dependency on the business but also the sector of the business. In Portugal, the dependency of business owners' financial situation on the business is slightly lower than in the euro area. However, the share of households very exposed to the pandemic crisis through business ownership is larger due to the higher concentration of businesses in sectors very affected by the crisis.

References

Catherine, Sylvain (2019). "Keeping Options Open: What Motivates Entrepreneurs?" available in *SSRN 3274879*.

Costa, Sónia, Luísa Farinha, Luís Martins, and Renata Mesquita (2020). "Portuguese Household Finance and Consumption Survey: results for 2017 and comparison with the previous waves." *Banco de Portugal, Banco de Portugal Economic Studies*, 6(1).

Evans, David and Boyan Jovanovic (1989). "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy*, 97, 808–27.

Evans, David and Linda Leighton (1989). "Some Empirical Aspects of Entrepreneurship." *American Economic Review*, 79, 519–35.

Fairlie, Robert W. and Harry A. Krashinsky (2012). "Liquidity Constraints, Household Wealth, And Entrepreneurship Revisited." *Review of Income and Wealth*, 58(2), 279–306.

Farinha, Luísa, Sónia Félix, and João Santos (2019). "Bank Funding and the Survival of Start-ups." *Banco de Portugal, Working Paper*, (19).

Haltiwanger, John, Ron Jarmin, and Javier Miranda (2013). "Who Creates Jobs? Small Versus Large Versus Young." *Review of Economics and Statistics*, 95, 347–361.

HFCN (2016a). "The Household Finance and Consumption Survey: Methodological Report for the Second Wave." *ECB, Statistics Paper Series 17*, december.

HFCN (2016b). "The Household Finance and Consumption Survey: Results from the Second Wave." *ECB, Statistics Paper Series 18*, december.

HFCN (2020a). "The Household Finance and Consumption Survey: Methodological Report for the 2017 Wave." *ECB, Statistics Paper Series 35*, march.

HFCN (2020b). "The Household Finance and Consumption Survey: Results from the 2017 Wave." *ECB, Statistics Paper Series 36*, march.

Humphries, John Eric (2018). "The Causes and Consequences of Self-Employment over the Life Cycle." Mimeo, *Yale University*.

Hurst, Erik and Annamaria Lusardi (2004). "Liquidity Constraints, Household Wealth, and Entrepreneurship." *Journal of Political Economy*, 112, 319–347.

International Labour Organization (2020). "ILO Monitor: COVID-19 and the world of work." Second edition.

Kennickell, Arthur B., Myron L. Kwast, and Jonathan Pogach (2017). "Small Businesses and Small Business Finance during the Financial Crisis and the Great Recession: New Evidence from the Survey of Consumer Finances." In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, pp. 291–349, University of Chicago Press.

Levine, Ross and Yona Rubinstein (2017). "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *Quarterly Journal of Economics*, 132, 963–1018.

Qi, Shusen, Steven Ongena, and Hua Cheng (2018). "When They Work with Women, Do Men Get All the Credit?" Swiss Finance Institute Research Paper Series 18-01, Swiss Finance Institute.

Schmalz, Martin, David Sraer, and David Thesmar (2016). "Housing Collateral and Entrepreneurship." *The Journal of Finance*, 72.

United Nations (2011). "Canberra Group Handbook on Household Income Statistics." *United Nations Economic Commission for Europe*, second edition.

Non-technical summary

October 2020

Rise and fall of the largest firms in Portugal

João Amador, Mário Lourenço, Cloé Magalhães and Ana Catarina Pimenta

Several studies have been analysing firms' entry and exit of the market. However, there is still limited evidence on the dynamics of top firms throughout very long periods of time. This article builds on a new database that records the largest 200 firms operating in Portugal, according to their annual turnover, in each year of the period 1981-2018 to assess the rise-fall dynamics and the probability of exiting the top 200.

The top 200 firms represented about 10% of Portugal's total gross value added in 2018 and belonged mainly to the sectors of trade, accommodation and food services and industry.

On average, each firm belongs to the top 200 for nine years and there is a large number of firms that take part of the ranking for a few years. The median stay in the top 200 is six years and only about 6% of the firms were part of the top 200 for at least 30 years. The top 50 firms show a different behaviour than the others, staying longer in the top 200.

The dynamics in the ranking of the 200 largest firms is also analysed along two complementary approaches. Firstly, transition matrices for different time spans inform on the probability of firms moving across pre-established ranking intervals. Secondly, considering only firms whose entrance in the ranking is identified in a given moment in time, estimated survival functions inform on the likelihood of them remaining in the ranking within different time spans.

Considering eight classes of positions in the ranking – [1-25]; [26-50]; [51-75]; [76-100]; [101-125]; [126-150]; [151-175]; [176-200] –, to which is added the status "out" of the ranking, the median number of changes between classes is 4 for firms that belong to the ranking between 6 and 10 years and increases to 11 and 9 for those that belong to the ranking between 26 and 30 years and between 31 and 38 years, respectively. This points towards more stability when firms are a part of the ranking for a long time.

The results suggest the relative resilience of the top firms in the ranking. For example, the probability of firms in the class [1-25] to remain in this position in the future decreases over time: 87% and 81% after one and two years, respectively, and around 29% after 20 years (Figure 1). For the firms placed in the class [26-50] the probabilities of keeping their position also decrease as the time span widens, reaching approximately 17% after 20 years. In the lower classes of the ranking the probabilities decay faster and stand below 5% after 20 years.

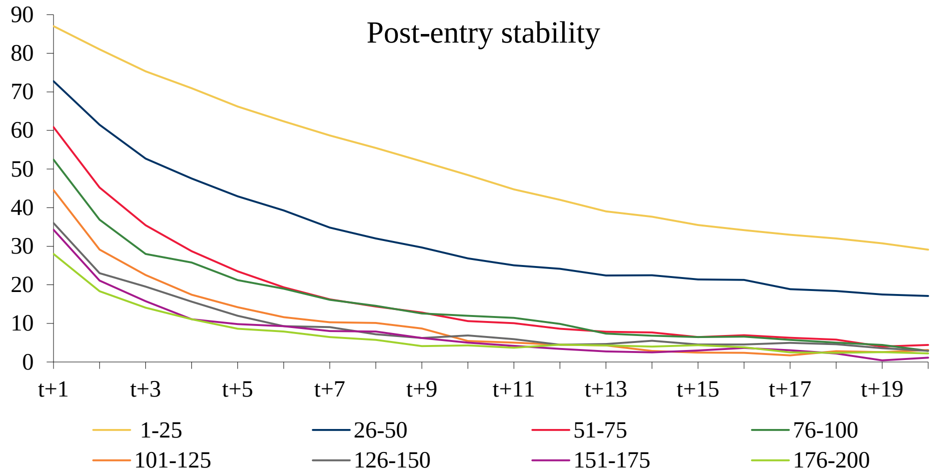


FIGURE 1: Probability of remaining in the same class of positions in the ranking between year t and year t+x, with x varying between 1 and 20 years

In addition, for all classes, it is more likely to fall or exit the ranking than to rise in the ranking. The fact that, on average, the rise in the ranking is harder than the fall is not contradictory with stability in top positions.

These results are corroborated by survival estimates. Approximately 74.6% of the firms remain in the ranking one year after entry and the estimated median duration is four years, meaning that 50% of the firms are expected to remain in the ranking for four or less years. In addition, as expected, the smallest firms (in the 1st quartile of the distribution) are less likely to remain in the ranking. By contrast, the largest firms (in the 4th quartile of the distribution) clearly have the highest survival probabilities after two years and up to the 35th year in the ranking.

Rise and fall of the largest firms in Portugal

João Amador
Banco de Portugal
Nova School of Business and Economics

Mário Lourenço
Banco de Portugal

Cloé Magalhães
Banco de Portugal

Ana Catarina Pimenta
Banco de Portugal

October 2020

Abstract
Several studies have been establishing stylized facts regarding firm entry and exit of the market. However, there is still limited evidence on the dynamics of top firms throughout very long periods of time. This article builds on a new database that records the largest firms operating in Portugal in terms of annual turnover. We consider the set of the top 200 firms in the period 1981-2018 and assess their dynamics across different classes in the ranking and the probability of exiting this group. The article concludes that there is more stability in terms of the firms placed in the top classes of the ranking. Moreover, on average, for different ranking classes and time horizons, the probability of rising in the rank is smaller than the probability to fall. The survival in the ranking differs according to the sector in which firms operate. Firms in the electricity and water supply sector survive for longer periods, while the median duration is lower in industry and construction. (JEL: L11, L20, L25)

"All live to die, and rise to fall."
Christopher Marlowe

1. Introduction

Economic literature refers firms' demography as a driver of economic growth. One relevant dimension is the entry and exit of firms in the market – the designated extensive margin. The relationship with economic growth is linked with the idea of "creative destruction" by Schumpeter (1911, 1942). According to this view, firms that enter the market bring new goods and services that, if successful, will replace outdated ones. This process makes firms that produce outdated goods or services exit the market, generating short-term losses in activity that will be more than offset in the medium and long term, thus bringing net gains in value added. For new firms whose goods

Acknowledgements: The authors thank Beatriz Pires and Bernardo Sarrasqueiro for the excellent work on the compilation of the dataset. The authors are also thankful to Nuno Alves, António Antunes, Carlos Gouveia, Filipa Lima and Pedro Duarte Neves for their useful comments and suggestions. The analyses, opinions and conclusions expressed herein are the sole responsibility of the authors and do not necessarily reflect the opinions of Banco de Portugal or the Eurosystem.

E-mail: jamador@bportugal.pt; mfllourenco@bportugal.pt; clmagalhaes@bportugal.pt; apimenta@bportugal.pt

or services do not pass the test of market, there should be a swift exit process, freeing up resources for new start-ups or for the growth of incumbents.

Another important dimension of firms' demography is their growth – the designated intensive margin – measured in terms of turnover, gross value added or employment. Firms' growth dynamics depend on multiple factors. Some of them relate to the specific characteristics of the firm, such as its orientation towards foreign markets, the ability to innovate and differentiate its goods or services or the quality of management. Other factors are external to the firm and relate with the regulatory environment, financing conditions, inefficiencies in the functioning of labour and product markets and overall macroeconomic developments. These elements also have a bearing on the extensive margin. For example, the underlying market competition conditions influence both firms' entry and exit, as well as changes in their market shares.

The empirical literature on firms' demography is too large to be fully mentioned here. Given their importance for long-term economic growth, much of the literature focuses on startups' probability of survival. Some contributions are those of Wagner (1994), Audretsch *et al.* (1999), and Mata *et al.* (1995), which analysed the post-entry performance of new manufacturing firms in Italy, Germany and Portugal, respectively. In a similar strand of work, several studies have pointed out that startups' survival depends on initial size (e.g. Mata and Portugal (1994) and Mata *et al.* (1995)), age (e.g. Dunne *et al.* (1989)), bank funding (e.g. Farinha *et al.* (2019)), human capital in the firm (e.g. Mata and Portugal (2002)), among others. A totally different strand of literature focuses on the role of very large firms and their granular impacts on aggregate outcomes. This literature started with the seminal work by Gabaix (2011) and a related application using data for Portuguese exporters is that of Cabral *et al.* (2020). A third strand of literature focuses on the behaviour of very large firms and the rise in product market concentration. Examples of this literature are Autor *et al.* (2017), which associates superstar firms to the fall in labour share in the US, and OECD (2018), which surveys information on market concentration in the US, Japan and Europe. Nevertheless, our study does not directly relate to these strands of analysis.

Possibly due to lack of information, the literature has not been covering the evolution of firms over long periods of time, which limits the ability to fully assess their rise-fall dynamics. Information about the year of creation of a firm allows for the estimation of survival functions that convey information on the probability of survival in each moment, but this is different from assessing the individual path of each firm regarding its relative size in the economy. Nevertheless, although existing studies cover relatively short time spans, they consider large sets of firms, which increases the representativeness of results. These aspects determine the unavailability of analyses that could be deemed comparable to our work and, hence, used to benchmark its results.

In this article we contribute to the literature by considering firms' rise-fall dynamics in wider time horizons. We use a database that identifies the largest Portuguese firms, which spans over four decades and was built for this purpose from previously scattered business information. More specifically, the sample used in this analysis covers the top 200 firms in terms of turnover in each year between 1981 and 2018. The methodological approach is mostly descriptive, i.e., we are not suggesting explanatory factors for the identified firms' rise and fall dynamics. Results point to a higher stability of the firms in the top classes of the ranking when compared with those in lower classes, thus indicating that the largest firms in the ranking generally maintain their positions. Moreover, on average, the probability of rising in the ranking tends to be smaller than the probability of falling, which is explained by the effect of firms entering directly to intermediate positions in the ranking.

The article is organized as follows. The second section presents the details of the database. Additionally, descriptive statistics regarding the sectoral dimension and the distribution of the largest 200 firms are briefly analysed. The third section presents the results of the rise and fall dynamics of firms in the top 200 during the period considered. Firstly, we present the transition matrices between ranking classes for different time horizons. Secondly, survival functions are estimated, relating the number of years in the ranking and the probability of staying in. The fourth section includes some concluding remarks.

2. Database and descriptive statistics

This section presents the data sources and procedures underlying the construction of the *Largest Portuguese Firms Database*. This dataset was constructed with a view to support this article but it will be made available for further research. In addition, this section provides a brief set of descriptive statistics regarding the referred dataset.

2.1. Data sources, treatment and harmonization

The *Largest Portuguese Firms Database* contains the top Portuguese firms in terms of turnover, mainly combining data from the Simplified Corporate Information (IES, the Portuguese acronym for "*Informação Empresarial Simplificada*"), which contains information on balance sheets and income statements for almost all non-financial Portuguese firms from 2006 onwards, and for more remote periods, data directly collected from hard copies of specialized business publications that summarize publicly available information. The *Largest Portuguese Firms Database* covers the 1976-2018 period, including each firm's yearly turnover (ranking variable).

For the years 1992 onwards, the database uses the reference population of active Portuguese firms estimated by the Statistics Department of Banco de Portugal. This dataset contains, for each firm, variables such as the identification number, headquarters location and the main sector of economic activity according to NACE Rev.2, as well as the number of employees, turnover, total assets and capital. This reference population collects information from several sources. Besides the already mentioned major contribution from the IES database, the Central Registry of Legal Entities, a database managed by the Institute of Registries and Notary of the Ministry of Justice, the Statistical Units File of Statistics Portugal, the Quarterly Survey on Non-Financial Corporations (ITENF, the Portuguese acronym for "*Inquérito Trimestral às Empresas Não Financeiras*"), the Integrated System of Securities Statistics, Banco de Portugal's Central

Credit Register and information obtained for the compilation of the Portuguese Balance of Payments and International Investment Position statistics are all taken into account.

Up to 1992, the yearly rankings of Portugal's largest firms published by specialized business publications were used to feed the database. Several publications were combined in order to obtain the largest information set possible. For the years from 1976 to 1978, data was collected from the *SEMAP* ranking of Portuguese top firms; for the years between 1979 and 1990, such ranking was collected from the *Negócios* magazine and, for 1991, the *Exame* magazine was considered. For each of these sources, all data deemed relevant was manually inserted into the database: company name, number of employees in each year, turnover, among other elements. Nevertheless, the number of top firms was not consistent across publications: the rankings ranged from the top 100 companies (for the most remote years) to rankings containing the top 500 companies in each year.

As the *Largest Portuguese Firms Database* combines several sources of information, some harmonization procedures were required. Given the need to uniquely identify each company throughout the relevant time span of the database, the most recent firm identification number was used. The link between the firm's most recent identifier and previous ones was manually established whenever possible. When this procedure was not possible (for example, in the cases of firms included in remote years that have ceased activity or that have merged with new firms) specific codes where attributed to identify the firms throughout the entire database. Firms' classification by economic activity, institutional sector and location of their headquarters were considered the same throughout the time span of the database, corresponding to the most recent information available.

Most firms included in the *Largest Portuguese Firms Database* belong to the Non-Financial Corporations sector, as defined by the European System of Accounts (ESA 2010). Given their importance in some activities, state owned firms are also included, belonging to the Non-Financial Corporations or the General Government sectors. Non-financial holdings which are classified as Financial Corporations are also considered in the database in order to cover the activity of economic groups.

A database comprising only a small set of large firms is necessarily affected by undesired attrition. This can occur due to the creation or exit of large special purpose entities that may not have a connection with actual economic activity or due to mergers and acquisitions. These problems were addressed in several ways. Firstly, firms established in the Madeira Free Trade Zone, whose participation in the ranking was relevant for several years, were removed.[1] Secondly, specific events involving the top ranked firms (e.g. mergers, spin-offs, etc.) were also addressed. These events may have resulted solely from the restructuring of groups, hence leading to artificial cases of entry

1. The Madeira Free Trade Zone (*Zona Franca da Madeira*) is in activity since 1986. Several large special purpose entities established headquarters in this location with a view to benefit from special tax conditions. The original database included 82 entities located in the Madeira Free Trade Zone, most of which recorded in the 1995-2003 period and in the "Transportation and communication" and "Other activities" sectors. These entities were removed from the final version of the database.

and exit from the database. In such cases, whenever possible, the entities involved were aggregated and a single identification number was considered throughout the time span of the *Largest Portuguese Firms Database*. These adjustments made it possible to take into consideration the restructuring of large economic groups, even if the dissolution of some holding companies could be corrected.

For the purpose of this article, the top 200 firms listed in the *Largest Portuguese Firms Database* were considered from 1981 to 2018. Data from 1976 to 1980 were excluded because the database lists less than 200 companies in those years. Overall, our final selection contains 7600 observations, comprising 835 distinct firms and 38 years of information.

2.2. *Descriptive statistics*

The share of the top 200 firms on total Gross Value Added (GVA) of the Portuguese economy is quite high. These firms represented about 10% of Portugal's total GVA in 2018 (Figure 1). This share has decreased since 1995, when it represented approximately 15% of total GVA. This evolution was mostly determined by the decrease of the share of the GVA generated by the top 50 firms of the ranking, in particular by firms in the "Electricity and water" and "Transportation and communication" sectors. The share of these firms relatively to the respective sectoral GVA is higher than in the remaining sectors, having decreased from close to 96% and 83%, respectively in 1995 to values close to 40% in 2018. As for the remaining activities, firms in "Industry", "Construction" and "Trade, accommodation and food services" in the ranking present a stable share of the respective sectoral GVA, fluctuating between 5% and 20% in the horizon under analysis.

Considering the set of firms included in the rankings at least once during the 1981-2018 period, 35% belong to the sector "Trade, accommodation and food services" and 31% to "Industry". "Construction" and "Transportation and communication" represent 10% and 7% of firms in the ranking, respectively, while "Electricity and water" represents 2% of firms. Those operating in other activities correspond to 15% of the firms in the ranking.

The distribution of the top 200 firms by sector in each year is shown in Figure 2. "Industry" represented 37% of this set of firms in 1981, reaching a maximum of 47% in 1985. In the two decades that followed, the share of "Industry" in the set of companies analysed decreased to a minimum of 19% in 2005. After the global economic and financial crisis of 2008, the relevance of this sector steadily increased, standing at 32% of the set of largest companies analysed in 2018.

The sector "Trade, accommodation and food services" has shown an inverse trend. The share of firms in the top 200 operating in this sector increased from 27% in 1985 to 48% in 2005. Later, this proportion dropped to around 40%. As for the remaining activities, "Construction" represented around 5% of the firms in the ranking until 1994. In the two decades that followed, the share of this sector in the ranking stood at around 8%. After reaching a maximum of 11% of the firms in the ranking in 2011, the share of "Construction" decreased to a minimum of 3% in 2018.

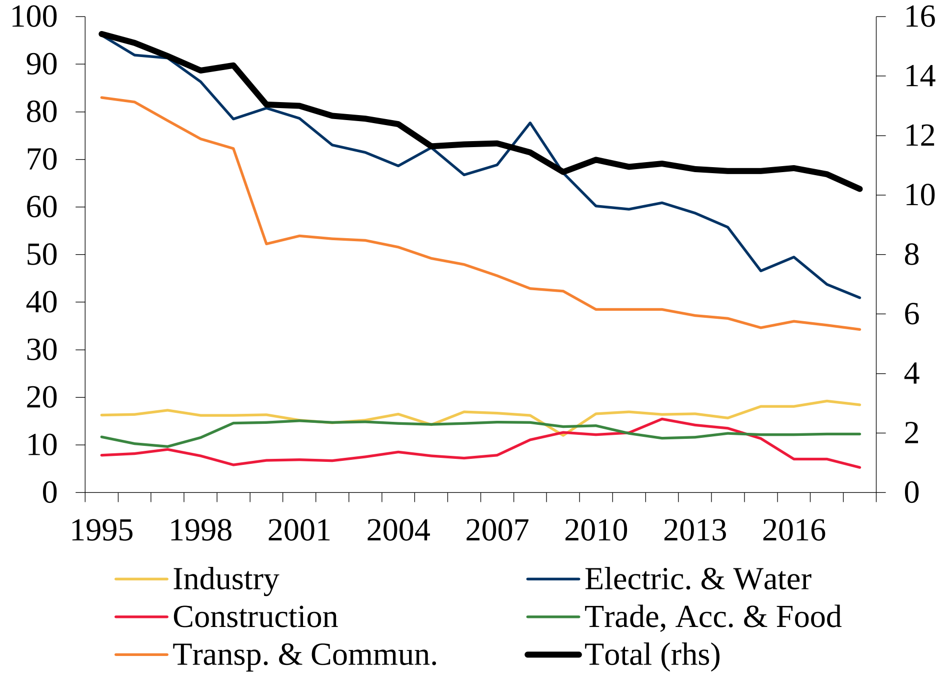


FIGURE 1: Share of top 200 Portuguese firms on Gross Value Added (GVA), percentage

Source: Banco de Portugal and Statistics Portugal.
Notes: Gross Value Added of the economy available as of 1995. The sectors presented in this figure correspond to aggregations of NACE Rev.2 sections: "Industry" (sections B and C), "Electricity and water" (sections D and E), "Construction" (section F), "Trade, accommodation and food services" (sections G and I), "Transportation and communication" (sections H and J). The GVA generated by the "Transportation and communication" firms in the top 200 is affected by the restructuring of a "Communications" group in 2000.

Finally, "Electricity and water" increased its relevance in the rankings, from around 1% in the early 1980s to 6% in 2018. This evolution reflects the developments in the electricity market during the last decades, notably the privatization of *Energias de Portugal* in the late 1990s and the segmentation of production, distribution and retail activities imposed by the implementation of the Iberian Energy Market as of 2006. The restructuring of these companies led to the establishment of new operators that have become a part of the list of the top 200 largest Portuguese firms.

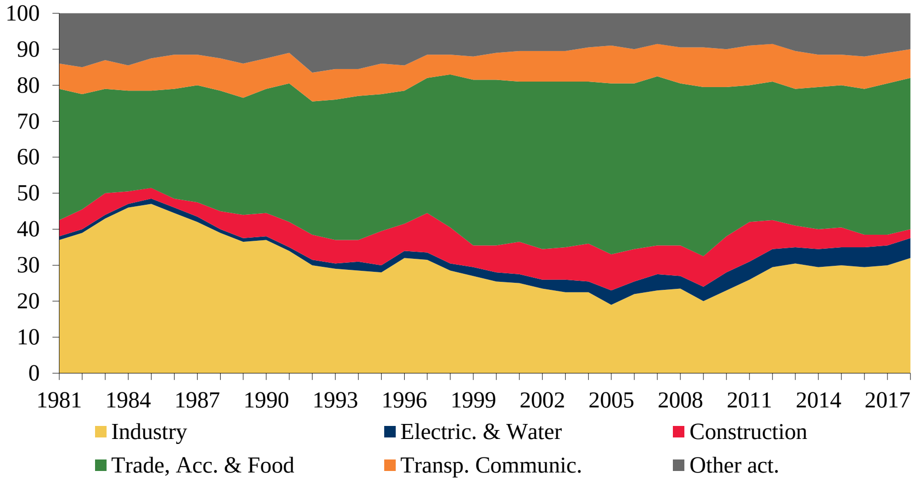


FIGURE 2: Distribution of the top 200 Portuguese firms by sector and year

Notes: The sectors presented in this figure correspond to aggregations of NACE Rev.2 sections: "Industry" (sections B and C), "Electricity and water" (sections D and E), "Construction" (section F), "Trade, accommodation and food services" (sections G and I), "Transportation and communication" (sections H and J).

The sectoral structure of the top 50 firms also provides further information (Figure 3). In 2018, "Trade, accommodation and food services" stood for nearly half of the top 50 largest firms, while representing 40% of the remaining top 51-200 firms. The relevance of network industries among the top 50 largest firms is also noticeable. "Transportation and communication" and "Electricity and water" sectors were more relevant among the top 50 largest firms (18% and 8%, respectively) than among the remaining top 51-200 firms (5% in both sectors). Conversely, firms in the "Industry" sector represented 20% of the top 50 largest firms and 36% of the remaining top 200 firms. "Industry" firms in the top 50 mostly operate in fuel, transport equipment and components, as well as food and drinks industries, while the activities of "Industry" firms in the remaining top 51-200 are more disperse. The structural differences between the top 50 largest firms and the remaining top 51-200 are observable throughout the time span covered, as shown when comparing the situation in 1981 and 2018.

Another dimension of analysis is the number of years that each firm belongs to the top 200. The average number is 9 years and the distribution is highly skewed to the right, which means there is a large number of firms that take part of the ranking for a relatively small number of years (Figure 4). The average is quite misleading to describe the individual developments of firms in the economy (Altomonte *et al.* 2011). The median firm was part of the top 200 for 6 years; on the other hand, 6% of the firms (46 out of 835 firms) were part of the top 200 for at least 30 years.

(A) 1981 (B) 2018

FIGURE 3: Distribution of the top 200 Portuguese firms by position within the ranking (top 50 vs top 51-200) and sector

Notes: The sectors presented in this figure correspond to aggregations of NACE Rev.2 sections: "Industry" (sections B and C), "Electricity and water" (sections D and E), "Construction" (section F), "Trade, accommodation and food services" (sections G and I), "Transportation and communication" (sections H and J).

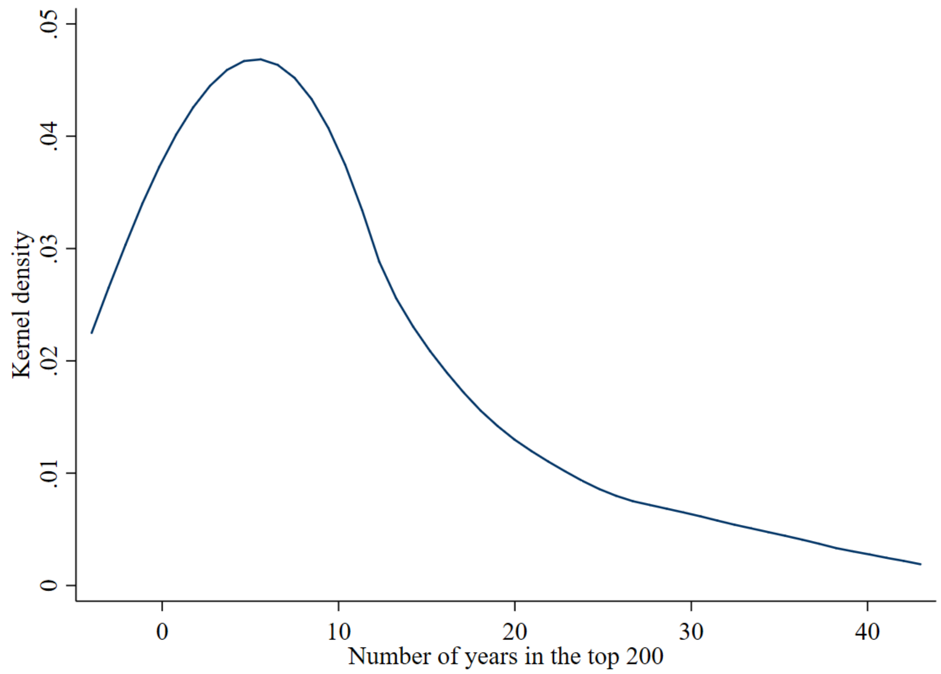
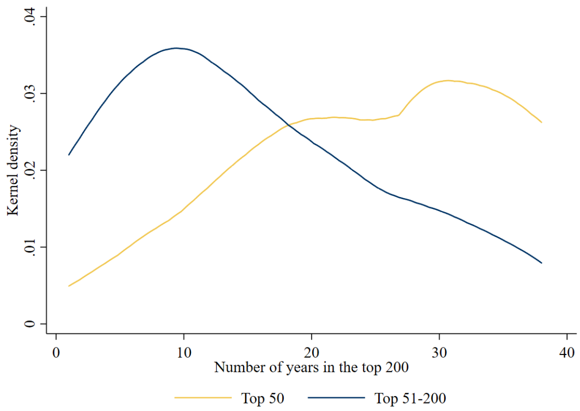


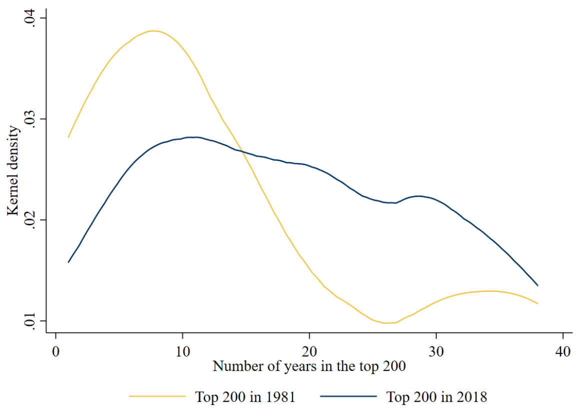
FIGURE 4: Distribution of the number of years in the top 200

Notes: The number of years in the top 200 was uniquely computed for each of the 835 distinct firms that were part of it between 1981 and 2018. The Kernel density estimation is a nonparametric method of estimating the probability density function of a variable. In the literature, for continuous variables, these density estimates are considered preferable to histograms because they smooth the distribution.

As previously mentioned, the top 50 firms show a different profile than the rest of those in the top 200 (Figure 5, panel A). As expected, it is noticeable that the former firms stay in the top 200 longer than the rest. In the 38 years covered in this analysis, firms in the top 50 were a part of the ranking, on average, 17 years (median of 18 years), while the remaining ones have been part of the top, on average, almost 8 years (median of 5 years). Furthermore, firms that were part of the top 200 in the later years of the sample remain there for longer periods than those identified in the earlier years (Figure 5, panel B). The median number of years in the ranking increased from 10, for the top 200 firms in 1981, to 18, for the top 200 firms in 2018, thus pointing towards a higher stability in the ranking in the last decades. These features are further developed in the next two sections by analysing transition matrices between classes of positions in the ranking and by estimating survival functions.



(A) By class of positions



(B) By year

FIGURE 5: Distribution of the number of years in the ranking

Notes: The number of years in the top 200 was uniquely computed for each of the 835 distinct firms that were part of the top 200 between 1981 and 2018. In Panel A, the “Top 50” correspond to those firms whose most frequent position in the ranking is in the top 50 positions. “Top 51-200” are the remaining firms. Panel B describes the distribution of the number of years in the top for those firms that were observed in the 2018 and 1981 top 200 firms.

3. Results

The dynamics in the ranking of the 200 largest firms according to turnover is analysed along two complementary approaches. Firstly, transition matrices for different time spans inform on the conditional probability of firms moving across pre-established ranking intervals. Secondly, considering only firms whose entrance in the ranking is identified in a given moment in time, estimated survival functions inform on the likelihood of them remaining in the ranking within different time spans.

3.1. Transition matrices

The results presented in this subsection are based on eight classes of positions in the ranking of the 200 largest firms according to turnover, as described in subsection 2.1. The eight classes considered correspond to positions: [1-25]; [26-50]; [51-75]; [76-100]; [101-125]; [126-150]; [151-175]; [176-200], to which is added the status "Out" of the ranking. In each year, the firms that belong to the "Out" category correspond to those that have at least once been included in the ranking but do not belong to the ranking in that specific year.

The initial analysis assesses the overall dynamics of firms among classes. The shape of the distribution of the number of changes in class of positions for each firm, considering different time spans of permanence in the ranking, is presented in Figure 6. As expected, the median number of changes increases for larger time spans. The median number of changes between classes is 4 for firms that belong to the ranking between 6 and 10 years and increases to 11 and 9 for those that belong to the ranking between 26 and 30 years and between 31 and 38 years, respectively. Moreover, the median and the percentile 25 are lower for those firms that belong to the ranking between 31 and 38 years comparing to those that remain in the ranking between 26 and 30 years. This points towards more stability when firms are a part of the ranking for longer periods.

The previous analysis can be developed by explicitly taking into account the movements observed between specific classes. Benefiting from the long time horizon available in the database, transition matrices are consecutively computed for intervals between 1 and 20 years. As an illustration, Table A.2 in Appendix represents the transition matrix for a time horizon of 10 years. Rows identify the starting position of the firm in moment t and columns refer to its position in the period $t + 10$, thus, each row is a conditional distribution adding up to 100%. According to Table A.2, the 200 largest firms tend to remain in the same ranking class 10 years later, i.e., the probabilities in the main diagonal of the transition matrix are higher, particularly in the upper classes. For example, 48.4% of firms in the interval [1-25] in a given year remain within that same set of positions 10 years later. Other firms fall within the ranking: 7.4% move to the interval [51-75] and 23.4% exit the ranking. As anticipated, an important feature in the transition matrix is that firms in the lower intervals have a lower probability of remaining in the ranking. For example, only 4.3% of firms in the interval [176-200] remain in that class 10 years later, while 80.1% of them exit the ranking.

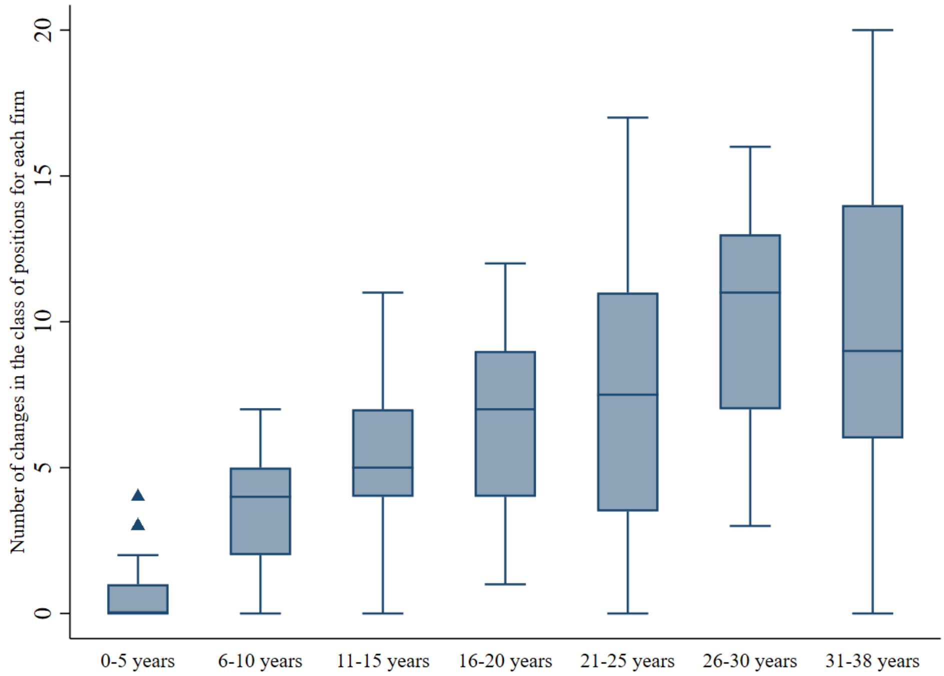


FIGURE 6: Distribution of changes between classes of positions for each firm, by the number of years in the top 200

Notes: In each year, it is assumed that a firm changes in the class of positions if the class to which it belongs in that year is different from the class to which it belonged in the previous year. If the firm belongs to the ranking for only one year, there are no changes. The classes considered correspond to positions: [1-25]; [26-50]; [51-75]; [76-100]; [101-125]; [126-150]; [151-175]; [176-200], to which is added the status "Out" of the ranking. In the box plot, the central box represents the values of the 25th percentile to 75th percentile (interquartile range) and the horizontal line correspondes to the median of the distribution (50th percentile). The vertical line extends from the minimum to the maximum value, excluding outliers (values lower than the difference between the 25th percentile and 1.5 times the interquartile range, or higher than the sum of the 75th percentile and 1.5 times the interquartile range). The triangles correspond to outliers.

The information contained in the transition matrices for different time spans allows for the identification of other stylized facts. Figure 7 presents the probability of firms placed in each class of the ranking to remain in the same class of positions up to 20 years later.[2] As illustrated in Appendix A, these "post entry" probabilities correspond to the cells in the main diagonal for the successive transition matrices from 1 up to 20 years. For example, results indicate that firms in the class [1-25] have a slowly decreasing probability of keeping this position in the future, standing at 87% and 81% after 1 and 2 years, respectively, and standing at around 29% after 20 years. For the firms placed in the class [26-50] the probabilities of keeping their position also decrease as the time span widens, reaching approximately 17% after 20 years, while in the lower classes of the ranking the probabilities decay faster and stand below 5% after 20 years. Overall, one important result that emerges is the relative resilience of the top firms in the ranking. The Spearman's rank correlation coefficient is an alternative measure of the stability of firms between classes of positions in the ranking. On average, the correlation coefficient

2. Longer transitions could be considered but the number of underlying firms used in the computation would be smaller and results would thus be less robust.

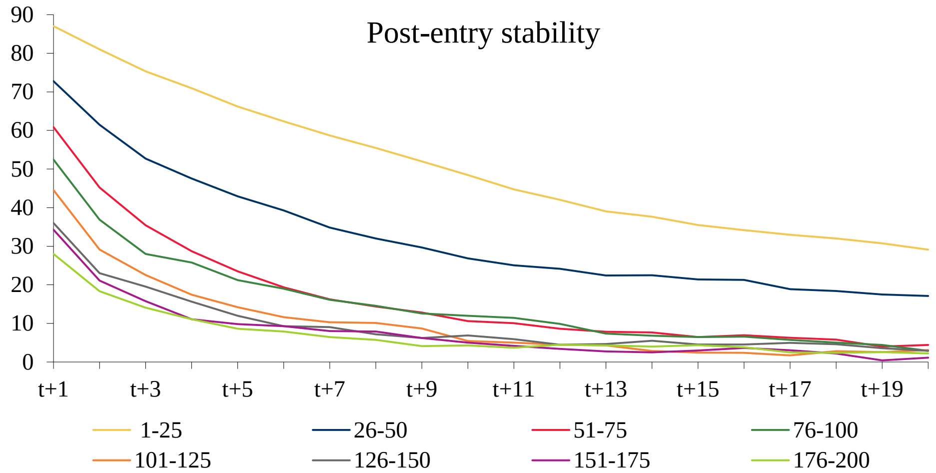


FIGURE 7: Probability of remaining in the same class of positions in the ranking between year t and year t+x, with x varying between 1 and 20 years

Notes: For each line, values correspond to those of the main diagonal of each of the 20 transition matrices. For instance, for the top 25 firms, in each year, the probability of remaining in that category 1 year later corresponds to the first cell of Table A.1 (87%). Similarly, the probability of these firms remaining in the top 25 after 10 and 20 years corresponds to the first cell of Table A.2 (48.4%) and Table A.3 (29.1%), respectively.

between the firms' class in year t and year $t + x$ tends to be lower as x widens, thus confirming the relative stability for shorter horizons (Figure B.1 in appendix).

Another perspective is to assess the probability of firms moving to a higher class in the ranking in periods from $t + 1$ up to $t + 20$, depending on the class where they start. The probability of firms "rising" in the ranking corresponds, in each line of the transition matrix, to the horizontal sum of cells to the left of the main diagonal, along the different time horizons. Results are presented in Figure 8. The probability of rising in period $t + 1$ coming from the class [26-50] is about 9.7%, it increases to 13.3% nine years later and drops to 5.1% after 20 years. Conversely, the probability to rise starting from the lowest ranking class [176-200] in period $t + 1$ is 25.6%, decreases to 15.6% in $t + 10$ and declines until 9.3% in $t + 20$. Therefore, as expected, it is easier to rise when starting from below but this feature is not as strong in larger horizons.

The dynamics of firms falling or exiting from the ranking is described in Figure 9. In this case, the probability of fall or exit from the ranking in time horizons from t up to $t + 20$, when starting from each class, is equivalent to the horizontal sum of the cells to the right of the main diagonal (i.e., including the "Out" category) for each line.[3] Results show that the probability of fall or exit from the ranking by firms in interval [1-25] is 13% in period $t + 1$, 51.6% in $t + 10$ and 70.9% in $t + 20$. Conversely, the probability of exiting for firms in the interval [176-200] (falling is not possible) is 46.4% in period $t + 1$, 80.1% in $t + 10$ and 88.4% in $t + 20$.

At this point it is relevant to highlight that, for each starting class and different time transitions, the probability of falling or exiting the ranking is larger than that of rising,

3. It is worth noting that, for each class of positions, the sum of the probabilities of stability (Figure 7), rise (Figure 8) and fall or exit (Figure 9) corresponds to 100%.

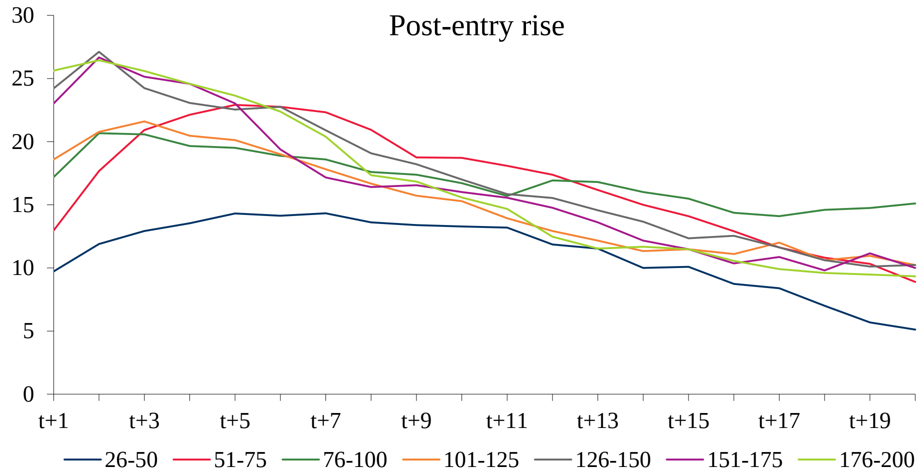


FIGURE 8: Probability of rising in the ranking in year $t + x$, given that in year t the firm belongs to each of the indicated class of positions, with x varying between 1 and 20 years

Notes: For each line, values correspond to the horizontal sum of the transition probabilities to the left of the main diagonal. For instance, considering that in year t the firm belonged to the top 26 to 50, the probability of rising to the top 25 is 9.7% after 1 year (Table A.1), 13.3% after 10 years (A.2) and 5.1% after 20 years (A.3). Similarly, for firms in the top 51 to 75 in year t, the probability of rising in the ranking is 13% after 1 year (Table A.1), 18.7% after 10 years (Table A.2) and 8.9% after 20 years (Table A.3). Given the classes of positions considered, it is not possible to rise when firms are already in the top 25.

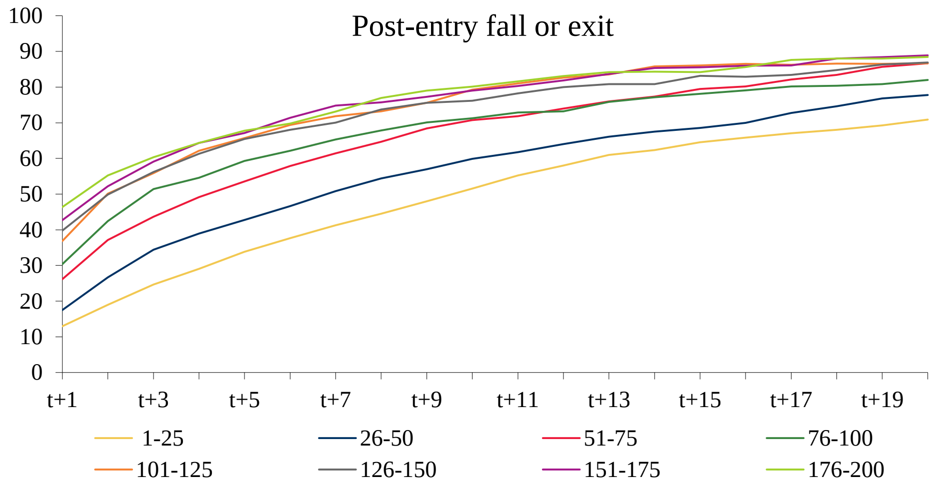


FIGURE 9: Probability of falling or exiting the ranking in year $t + x$, given that in year t the firm belonged to each class of positions, with x varying between 1 and 20 years

Notes: For each class of positions in the ranking, the values in the lines correspond to the sum of the transition probabilities to the right of the main diagonal. In each year, the firms that belong to the "Out" category correspond to those that have at least once been included in the ranking of the 200 largest firms, but do not belong to the ranking in that year. For instance, considering that in year t the firm belonged to the top 25, the probability of fall or exit from the ranking 1 year later is 13% (Table A.1) and 51.6% and 70.9% after 10 and 20 years, respectively (Table A.2 and Table A.3, respectively).

i.e., in each line, the sum of cells to the left of the main diagonal is smaller than the sum of cells to the right. This strong regularity is an important result and has a bearing on the perception about the dynamics of the largest firms in the market. Even if reaching the ranking signals success, the sword of Damocles is always hanging over their head.

Complementary, we focus on the path of firms entering the ranking. The probability of moving to the different intervals in the ranking in periods $t+1$ up to $t+20$ when starting from the situation "Out" corresponds to the bottom line in the different transition matrices, as signaled in Appendix A. Results are represented in Figure 10 and each line identifies the probability of an outside firm to move to the corresponding ranking interval in each time horizon. The probability of moving to each interval is smaller the higher the classes in the ranking. Moreover, the probability of ascending to each class increases along time. Nevertheless, it should be noted that the probabilities are relatively low in all horizons.

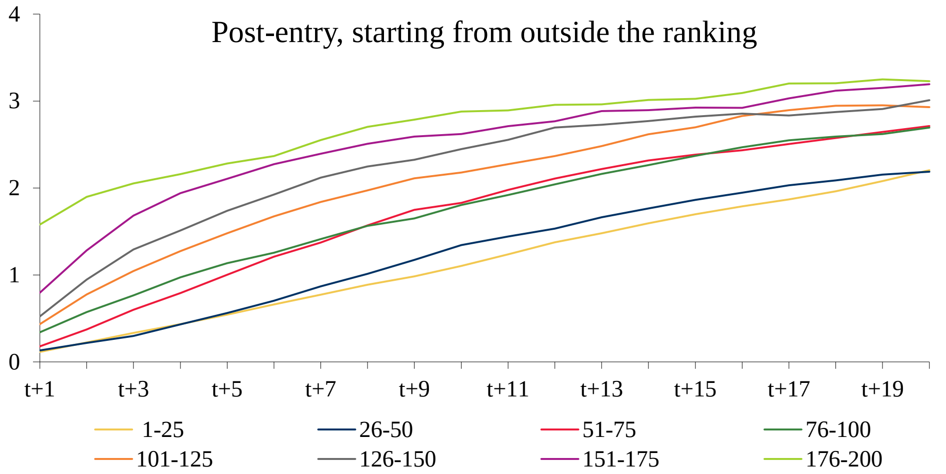


FIGURE 10: Probability of belonging to each class of positions in year t+x, given that in year t the firm was out of the ranking, with x varying between 1 and 20 years

Notes: For each of the 20 transition time horizons considered, the values correspond to those of the line highlighted in yellow (see Tables A.1, A.2 and A.3 as examples). In each year, the firms that belong to the "Out" category correspond to those that have at least once been included in the ranking of the 200 largest firms, but do not belong to the ranking in that year. In each year, 0.1% of the firms that were out of the ranking transitioned to the top 25 and 1.6% to the top 176 to 200 after 1 year (Table A.1). These probabilities are 1.1% and 2.9% for a 10-year transition horizon (Table A.2) and 2.2% and 3.2% for a 20-year transition horizon (Table A.3).

Finally, the dynamics of firms before exiting the ranking ("pre-exit") is described in Figure 11. In this case, taking firms that exit the ranking in moment t we analyse the probability of those firms being in each interval in periods from $t-1$ up to $t-20$. This information corresponds to the latest column in the set of the successive transition matrices, as signaled in Appendix A. Results show that the probability of exit from the ranking by firms that in the year before were in classes [1-25] and [26-50] is smaller (2.6% and 2.4%, respectively) increasing up 39.8% and 48.9% when the time horizon recedes to 20 years before exit. Conversely, firms present in the class [176-200] in the moment prior to exit have a probability of 46.4% of exiting, which increases to 88.4% if they depart from this same class 20 years earlier. This confirms the result of stability in the higher positions of the ranking, i.e. larger firms have a relatively higher probability of maintaining their positions.

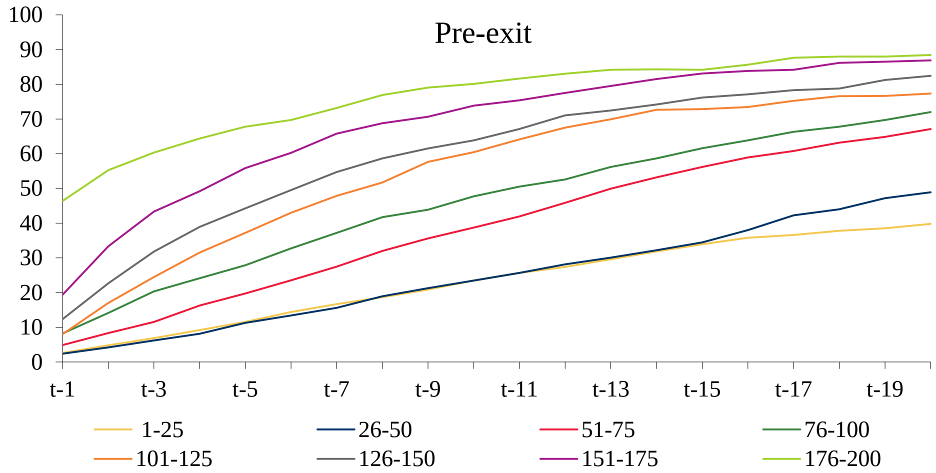


FIGURE 11: Probability of belonging to each class of positions in year $t - x$, given that in year t the firm left the ranking, with x varying between 1 and 20 years

Notes: For each of the 20 transition time horizons considered, the values correspond to those of the column highlighted in red (see Tables A.1, A.2 and A.3 as examples). In each year, the firms that belong to the "Out" category correspond to those that have at least once been included in the ranking of the 200 largest firms, but do not belong to the ranking in that year. In each year, 2.6% of the firms that were in the top 25 in the previous year leave the rank (Table A.1). This probability increases to 23.4% for a 10-year transition horizon (Table A.2) and to 39.8% for a 20-year transition horizon (Table A.3).

The previous results show that the probability of firms changing classes in the ranking is higher for those placed in the lower end and changes are mainly downwards as time goes by. A complementary analysis relies on the net changes in each class (Figure 12). For each class of ranking positions, the net changes correspond to entries minus exits, disaggregated by direction of the move – from upper classes, lower classes or outside the ranking. By construction, since the number of firms in each class is fixed, the net moves cancel out (their sum is zero in all classes). However, it is relevant to notice that the number of upward and downward movements in the ranking is not necessarily symmetric, depending instead on the magnitude of the movement. For example, an upward movement of four classes by a single firm pushes four other firms to the class immediately below. This effect explains why rises are less probable than falls or exits in our database and it is also present when firms enter the ranking (sometimes to intermediate positions). Figure 12 shows that direct entries to intermediate classes are relevant (positive red bars), classes are fed by net moves from upper ones (positive blue bars) and they feed the lower ones (negative yellow bars). Classes [26-50] and [101-125] are those where the contribution to net entry is mostly associated with firms moving from outside the ranking. In addition, in the lowest class [176-200] there is a net move from upper classes that adds to a large number of firms that exit the ranking, movements which are not compensated by entries coming from outside the ranking.

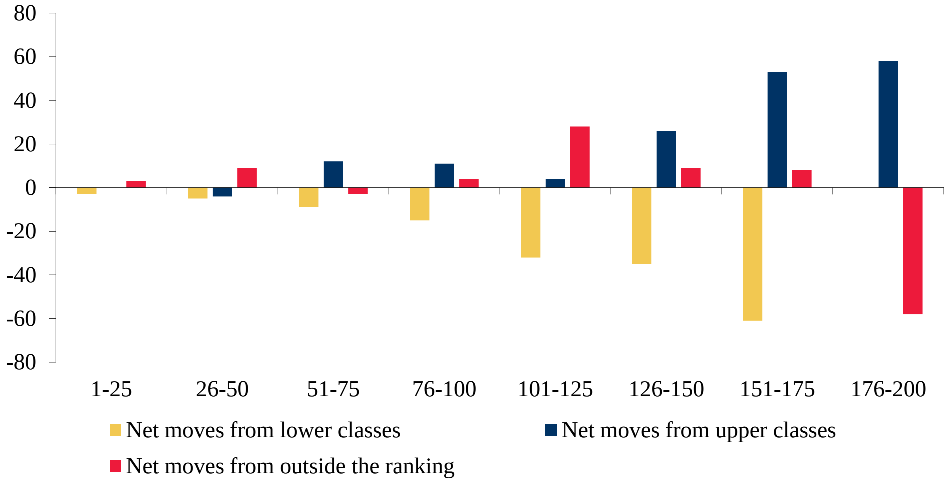


FIGURE 12: Net changes between classes of the ranking from t to $t+1$

Notes: Net changes represent the difference between the number of firms that enter a class of positions (from other classes or from the "Out" category) and the number of firms that exit the same class of positions (to other classes within the ranking or by exiting the ranking). Therefore, "Net changes from lower/upper classes" correspond to entrances from lower/upper classes minus exits to lower/upper classes and "Net changes from outside the ranking" correspond to new entries in the top 200 minus exits from the top 200.

3.2. *Nonparametric analysis of duration*

In this subsection we use duration analysis methods to estimate firms' probability of remaining in the ranking of the 200 largest firms ("survive") after different time intervals. The event of interest corresponds to firm's exit from the ranking (failure event). In addition, we compare the "survival" experiences across different sectors of activity and size classes.

3.2.1. *Methodology and sample characterization*

Considering T a non-negative variable denoting the time elapsed between firm entry and exit of the ranking, the survival function is thus represented by:

$$S(t) = 1 - F(t) = Prob(T > t) \tag{1}$$

where the F(t) is the cumulative distribution function for T. The survival function reports the probability of a firm remaining in the ranking beyond t, i.e., the probability that there is no exit prior to t.[4] The most common nonparametric estimate of the survival function is the Kaplan-Meier estimator (López-Garcia and Puente 2006).

For a dataset with k distinct failure times observed in the data, t_1, ...,t_k, the Kaplan and Meier (1958) estimate at any time t is given by:

4. The survival function is a monotone non-increasing function of time. The function is equal to 1 at t = 0 and decreases towards 0 as t goes to infinity (Cleves *et al.* 2010).

$$\hat{S(t)} = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \tag{2}$$

where n_j is the number of firms at risk (those that remain in the ranking) at time t_j and d_j is the number of failures (those firms that left the ranking) at time t_j. The product is taken over all observed failure times, departing from time t.

Since this estimator is a step function, the estimate of the pth percentile of survival horizons, t_p, is given by:

$$\hat{t_p} = min\left\{ t_i | \hat{S(t_i)} \leq 1 - \frac{p}{100} \right\} \tag{3}$$

for any p between 0 and 100, as described by Cleves *et al.* (2010).

For this estimation, several procedures had to be implemented over the original database. Firstly, firms that were already in the ranking in the first year observed (1981) were excluded, i.e., only firms that entered the ranking in 1982 or later were considered for this specific analysis. Firms that were in the database in 1981, and thus discarded from the sample, represent 24.0% of the total number in the database (835 firms). Out of these 200 firms, 7% belong to the ranking over the entire time horizon (1981-2018).

Secondly, firms with two or more one-year gaps, i.e., those that leave the ranking at least two times and re-enter, and firms with a gap greater than one year were dropped. Firms with two or more one-year gaps represent 4.9% (41 firms) of the total and those with a gap greater than one year correspond to 8.3% (69 out of 835 firms). For firms absent from the ranking only during one year, say year t, we assume that they remain in the ranking and attribute for that year the average of the ranking positions in $t-1$ and $t+1$. These gaps represent only 0.7% of the observations in our database (7600 observations) and are associated to 54 firms.

Thirdly, we assume that a firm does not survive in year t if it is absent from the sample in year $t+1$. Since the last year of the sample is used to identify the firms that exit the ranking in 2017, we restricted the sample to those that entered the ranking between 1982 and 2017 (only 5 firms entered the ranking in 2018). Overall, the sample used in this section takes information for 520 firms over the years from 1982 to 2018 (3583 observations).

3.2.2. Survival functions

The results of the Kaplan-Meier survival estimate for the firms in the sample are plotted in Figure 13. The maximum duration in the ranking of the 200 largest firms is 36 years. Approximately 74.6% of the firms remain in the ranking one year after entry and the estimated median duration is 4 years, meaning that 50% of the firms are expected to remain in the ranking for 4 or less years. After 36 years in the ranking, only about 15.7% of firms "survive".

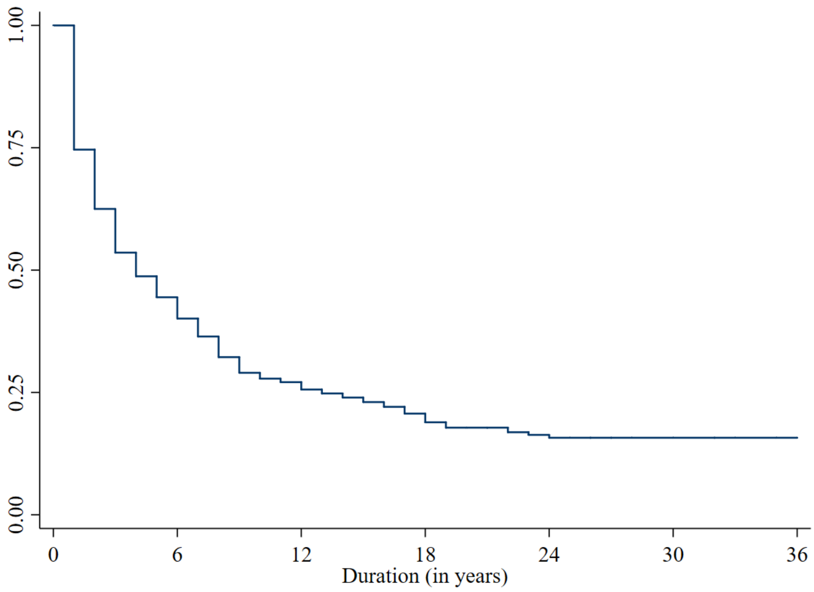


FIGURE 13: Kaplan-Meier survival function for the total sample

The "survival" in the ranking differs according to the sector in which the firm operates (Figure 14).[5] Firms operating in the sector "Electricity and water" have the highest survival probabilities up to the 25th year.[6] The median duration is higher in firms operating in the sectors "Trade, accommodation and food services" (7 years), "Other activities" (5 years) and "Transportation and communication"(4 years). By contrast, firms operating in the sectors "Construction" and "Industry" are those with lower median duration (2 and 3 years, respectively). Nevertheless, after 36 years, only about 21.2% of firms in the "Trade, accommodation and food services" sector remain in the ranking.

In addition, as expected, the smallest firms (in the 1st quartile of the distribution) are less likely to remain in the ranking (Figure 15).[7] By contrast, the largest firms (in the 4th quartile of the distribution) clearly have the highest survival probabilities after 2 years and up to the 35th year in the ranking. The estimated median duration for the smallest firms is 3 years, in sharp contrast with 7 and 8 years for the intermediate classes, respectively.[8] Nevertheless, after 36 years, only about 25.5% of firms in the 2nd quartile of the distribution remain in the ranking.

This additional set of results confirms the conclusions of the sections above regarding the resilience of specific firms in the ranking. The largest firms, which are by construction

5. Both Log-rank and Wilcoxon tests allow for the rejection of the hypothesis of equal survival among sectors.

6. For the sector "Electricity and water", it is not possible to estimate the median duration because the survival function becomes flat in $S(t) = 0.65$, i.e., 65% of the firms in this sector have not "failed" yet.

7. As for sectors, the tests allow for the rejection of the hypothesis of survival equality among classes of firms' size.

8. For the largest firms, it is not possible to estimate the median duration because the survival function has become flat at $S(t) = 0.51$, i.e., more than 50% of the largest firms have not "failed" yet.

those in the top classes of the ranking, have a much higher likelihood of remaining in top positions, i.e., current success seems to enhance future success.

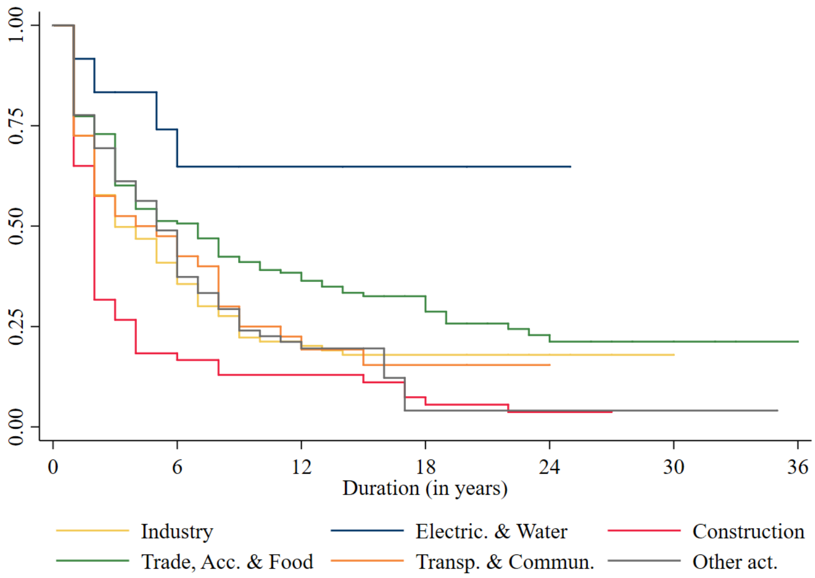


FIGURE 14: Kaplan-Meier survival function by sector

Notes: The sectors presented in this figure correspond to aggregations of NACE Rev.2 sections: “Industry” (sections B and C), “Electricity and water” (sections D and E), “Construction” (section F), “Trade, accommodation and food services” (sections G and I), “Transportation and communication” (sections H and J).

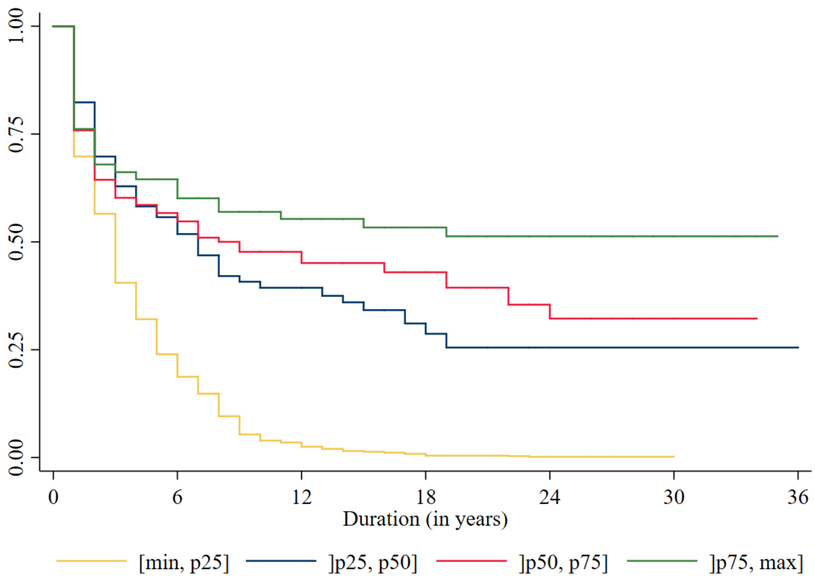


FIGURE 15: Kaplan-Meier survival function by firms’ size

Note: Each class of firms’ size was defined according to the distribution of turnover in each year. For instance, a firm belongs to the upper class in a year if its turnover was higher than the 75th percentile of the turnover distribution of all the firms in the database in that year. For this analysis, it was considered the modal size class for each firm.

4. Final remarks

This article uses a new database that identifies the top Portuguese firms in the last four decades according to their annual turnover, and establishes some stylized facts regarding their dynamics and survival in the ranking. The results are obtained by computing transition matrices between classes of positions in the ranking for different time horizons and by estimating survival functions.

The empirical literature on firms' demography has provided a rich set of results. However, there is limited evidence relatively to the dynamics of top firms in very long periods of time. We conclude that there is more stability for firms in the top positions of the ranking, showing that size is associated to resilience. In addition, the probability of rising in the ranking in different time horizons for firms in all classes is lower than the probability of falling or exiting. The fact that, on average, the rise in the ranking is harder than the fall is not contradictory with stability in top positions. Although all firms face a sizable risk of dropping out of the ranking, those that have reached the highest positions are comparatively more stable than other top firms that lay in secondary positions. These results are corroborated by survival estimates.

The obstacles to firm growth and their resilience in top positions are important aspects from the perspective of public policies. The rise of firms can be made difficult by different types of regulatory burdens or restrictive competition practices. The fall of firms can be the result of inadequate business models or triggered by unexpected events like technological transformations that turn existing products outdated or by the transfer of a firm's control between generations when management is not separated from property. The analysis and quantification of the determinants of the rise and fall of top firms is a promising avenue for further research.

References

Altomonte, Carlo, Giorgio Barba Navaretti, Filippo Di Mauro, and Gianmarco Ottaviano (2011). "Assessing competitiveness: how firm-level data can help." Tech. rep., Bruegel Policy Contribution.

Audretsch, David B, Enrico Santarelli, and Marco Vivarelli (1999). "Start-up size and industrial dynamics: some evidence from Italian manufacturing." *International Journal of Industrial Organization*, 17(7), 965–983.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen (2017). "The Fall of the Labor Share and the Rise of Superstar Firms." CEP Discussion Papers dp1482, Centre for Economic Performance, LSE.

Cabral, Sónia, Carlos Melo Gouveia, and Cristina Manteu (2020). "The granularity of Portuguese firm-level exports." *Banco de Portugal Economic Studies*.

Cleves, Mario, William Gould, Roberto Gutierrez, and Yulia Marchenko (2010). *An Introduction to Survival Analysis using Stata*. Stata press.

Dunne, Timothy, Mark J Roberts, and Larry Samuelson (1989). "The growth and failure of U. S. manufacturing plants." *The Quarterly Journal of Economics*, 104(4), 671–698.

Farinha, Luísa, Sónia Félix, and João Santos (2019). "Bank funding and the survival of start-ups." *Banco de Portugal working papers*.

Gabaix, Xavier (2011). "The granular origins of aggregate fluctuations." *Econometrica*, 79(3), 733–772.

Kaplan, Edward and Paul Meier (1958). "Nonparametric estimation from incomplete observations." *Journal of the American Statistical Association*, 53(282), 457–481.

López-Garcia, Paloma and Sergio Puente (2006). "Business demography in Spain: determinants of firm survival." *Banco de Espana Research Paper No. WP-0608*.

Mata, Jose and Pedro Portugal (1994). "Life duration of new firms." *The journal of industrial economics*, pp. 227–245.

Mata, José and Pedro Portugal (2002). "The survival of new domestic and foreign-owned firms." *Strategic Management Journal*, 23(4), 323–343.

Mata, José, Pedro Portugal, and Paulo Guimaraes (1995). "The survival of new plants: Start-up conditions and post-entry evolution." *International Journal of Industrial Organization*, 13(4), 459–481.

OECD (2018). "Market Concentration: Issues paper by the Secretariat." Issues paper, Organisation for Economic Co-operation and Development.

Schumpeter, Joseph (1911). *Theori der Wirtschaftlichen Entwicklung*. Berlin: Duncker and Humbolt.

Schumpeter, Joseph (1942). *Capitalism Socialism and Democracy*. New York: Harper.

Wagner, Joachim (1994). "The post-entry performance of new small firms in German manufacturing industries." *The Journal of Industrial Economics*, 42(2), 141–154.

Appendix A

		t+1								
		1-25	26-50	51-75	76-100	101-125	126-150	151-175	176-200	Out
	1-25	87.0	9.3	0.5	0.2	0.1	0.0	0.2	0.0	2.6
	26-50	9.7	72.8	13.7	1.0	0.1	0.3	0.0	0.0	2.4
	51-75	0.1	12.9	60.9	17.2	2.8	0.9	0.3	0.1	4.9
	76-100	0.1	0.8	16.3	52.4	16.0	4.4	1.1	0.6	8.2
t	101-125	0.0	0.4	2.6	15.6	44.5	21.4	5.5	1.9	8.0
	126-150	0.0	0.5	0.8	3.2	19.7	36.0	21.6	5.8	12.3
	151-175	0.1	0.0	0.4	1.1	3.7	17.7	34.3	23.4	19.4
	176-200	0.0	0.0	0.2	0.6	2.1	5.9	16.8	28.0	46.4
	Out	0.1	0.1	0.2	0.3	0.4	0.5	0.8	1.6	95.9

TABLE A.1. Transition matrix with 1 year time horizon

Notes: The rows reflect the initial category (class of positions in the ranking), and the columns reflect the category after 1 year. For instance, each year, 87% of the top 25 firms remained in the top 25 in the next year. The remaining 13% moved to a lower position (10.4%) or left the ranking (2.6%).

		t+10								
		1-25	26-50	51-75	76-100	101-125	126-150	151-175	176-200	Out
	1-25	48.4	14.7	7.4	2.4	1.9	0.3	1.3	0.1	23.4
	26-50	13.3	26.9	15.1	11.3	5.4	2.3	1.6	0.7	23.4
	51-75	5.1	13.6	10.6	12.0	6.9	5.0	5.1	3.0	38.7
	76-100	2.1	5.6	9.0	12.0	9.7	6.4	3.7	3.7	47.7
t	101-125	1.4	2.6	4.1	7.1	5.4	9.1	5.4	4.3	60.4
	126-150	0.6	1.7	4.0	4.9	5.9	6.9	6.7	5.6	63.9
	151-175	0.7	0.3	1.9	2.3	5.9	5.0	5.0	5.1	73.9
	176-200	0.3	0.6	1.4	2.1	3.7	2.9	4.6	4.3	80.1
	Out	1.1	1.3	1.8	1.8	2.2	2.4	2.6	2.9	83.8

TABLE A.2. Transition matrix with 10 years time horizon

Notes: The rows reflect the initial category (class of positions in the ranking), and the columns reflect the category after 10 years. For instance, each year, 48.4% of the top 25 firms remained in the top 25 after 10 years. The remaining 51.6% moved to a lower position (28.1%) or left the ranking (23.4%).

		t+20								
		1-25	26-50	51-75	76-100	101-125	126-150	151-175	176-200	Out
	1-25	29.1	10.7	7.6	5.1	2.0	3.1	1.8	0.9	39.8
	26-50	5.1	17.1	7.1	6.4	5.1	4.7	3.1	2.4	48.9
	51-75	4.2	4.7	4.4	7.1	5.1	3.6	1.8	2.0	67.1
	76-100	1.1	6.7	7.3	2.9	1.8	2.2	2.9	3.1	72.0
t	101-125	1.8	3.1	2.2	3.1	3.1	2.0	3.6	3.8	77.3
	126-150	1.3	1.3	1.3	3.1	3.1	2.9	2.9	1.6	82.4
	151-175	0.9	0.7	0.9	2.0	3.1	2.4	1.1	2.0	86.9
	176-200	0.4	0.2	0.2	1.8	2.2	2.7	1.8	2.2	88.4
	Out	2.2	2.2	2.7	2.7	2.9	3.0	3.2	3.2	77.8

TABLE A.3. Transition matrix with 20 years time horizon

Notes: The rows reflect the initial category (class of positions in the ranking), and the columns reflect the category after 20 years. For instance, each year, some 29% of the top 25 firms remained in the top 25 after 20 years. The remaining 71% moved to a lower position (31.2%) or left the ranking (39.8%).

Appendix B

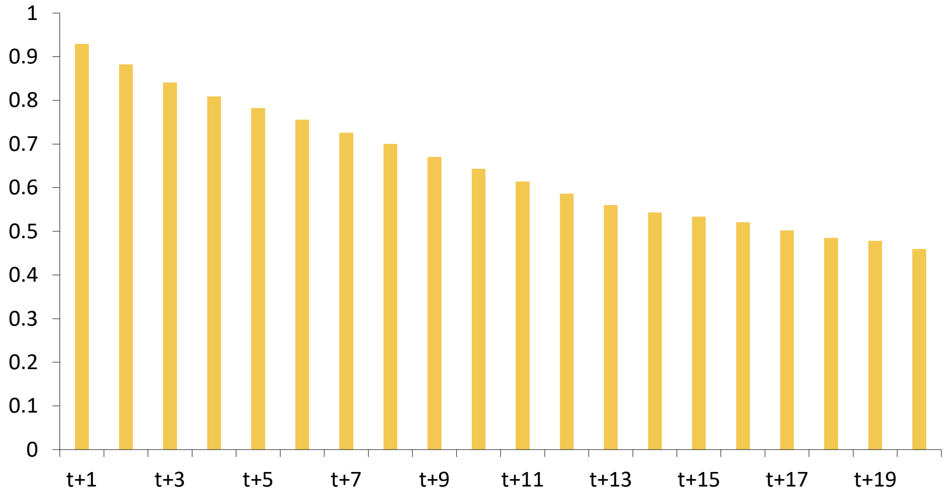


FIGURE B.1: Spearman's rank correlation coefficient between the firms' class in year t and year $t + x$

Non-technical summary

October 2020

House price forecasting and uncertainty: Examining Portugal and Spain

Robert Hill, Rita Lourenço and Paulo M. M. Rodrigues

The aim of this article is to discuss forecasting models and the impact of variables proxying for uncertainty on residential property prices in Portugal and Spain. Three potential sources of uncertainty need to be acknowledged when forecasting. On the one hand, economic uncertainty, which reflects the doubts that economic agents have about any future event. Thus, in our work besides the macroeconomic determinants typically used to explain residential property prices (such as GDP, income, residential investment, labour, unemployment rate, interest rates or housing loans) we also include business and consumer confidence and stock market volatility in an attempt to capture economic uncertainty. On the other hand, uncertainty regarding the model used for forecasting and uncertainty about the model parameters also needs to be considered.

To accommodate these three sources of uncertainty, in this paper we resort to Dynamic Model Averaging (DMA) which is a useful approach for forecasting because it inherently allows for the latter two types of uncertainty and for economic uncertainty through the augmented predictor set. In fact, we are able to track which predictors are more relevant over the forecast period and obtain some interesting conclusions regarding the relevance of all the predictors used.

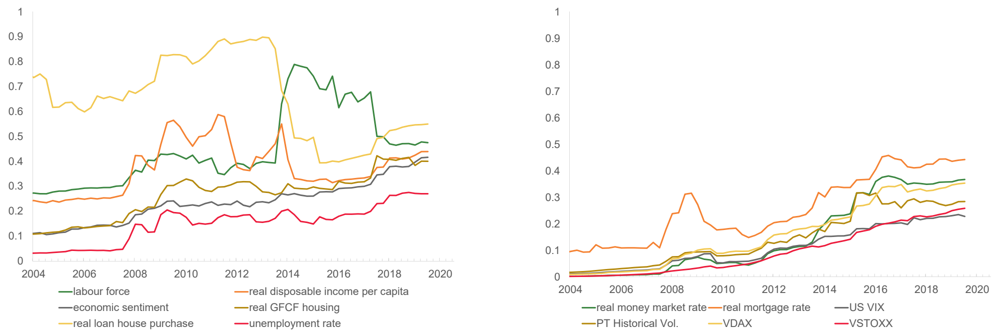


FIGURE 1: Portugal - Relevance of predictors for four periods ahead forecasting

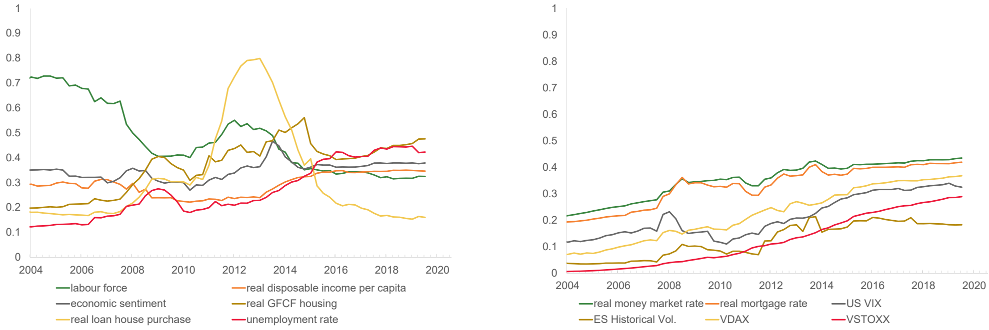


FIGURE 2: Spain - Relevance of predictors for four periods ahead forecasting

We find (Figures 1 and 2) that different predictors have varying inclusion probabilities (relevance) for both Portugal and Spain. For Portugal, most predictors (including the uncertainty measures) appear to have some value when it comes to forecasting changes in house prices. Each predictor's importance appears to increase over time. For Spain, most real economy predictors appear to be useful for forecasting, and there appears to be less variation in each predictor's importance over time. Economic uncertainty proxies appear to be more important in Portugal than in Spain for predicting house prices.

House price forecasting and uncertainty: Examining Portugal and Spain

Robert Hill
Nova SBE

Rita Lourenço
Banco de Portugal

Paulo M. M. Rodrigues
Banco de Portugal and Nova SBE

October 2020

Abstract
In this paper we apply dynamic model averaging (DMA) to forecast Portuguese and Spanish house prices. DMA is a useful method for forecasting because it inherently allows for uncertainty in both the combination of predictors (model uncertainty), as well as in the marginal effect of each predictor (parameter uncertainty). In doing so we are able to track which predictors are relevant over the forecast period. Besides fundamental macroeconomic determinants to house prices dynamics we also include as predictors business and consumer confidence and financial markets volatility. We find that different predictors have varying inclusion probabilities for both Portugal and Spain. In Portugal, most predictors appear to have some value when it comes to forecasting changes in house prices, including volatility and consumer confidence. Furthermore, each predictor's importance appears to increase over time. For Spain, most economic predictors appear to be useful for forecasting, and there appears to be less variation in each predictor's importance over time. However, volatility measures appear to be less important in Spain than in Portugal for predicting house prices. (JEL: C22, C53, R31)

1. Introduction

House prices have received considerable attention in recent years. The housing market and its developments can affect economic activity through the credit channel and through the impact that housing wealth has on consumption. Empirical evidence indicates that real estate is the main asset of households (Costa *et al.* (2020), ECB (2020) and EFF (2019)), and that changes in the value of wealth in housing can affect homeowners' consumption (Englund *et al.* (2002) and Case *et al.* (2005)). The impact on the economy resulting from changes in housing wealth may be greater than that resulting from movements in share prices (Helbling and Terrones (2003)). For an interesting overview of the dynamics of house prices in Europe, see e.g. Lourenço and Rodrigues (2015).

Acknowledgements: We thank Carlos da Silva Costa, Pedro Duarte Neves and participants of the Exchange seminar at Banco de Portugal for useful comments and suggestions. The opinions expressed in the article are those of the author(s) and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the author(s).

E-mail: robert.hill@novasbe.pt; rmlourenco@bportugal.pt; pmrodrigues@bportugal.pt

Economic uncertainty is not observable and reflects the doubts that economic agents have, be they consumers, entrepreneurs or policy makers, about any future event, be it economic (e.g. GDP or house price growth) or non-economic (e.g. a natural disaster).[1] There is no consensus among economists on how to measure it and in the economic literature an extensive set of proxies have been used to measure economic uncertainty dynamics (Bloom (2013)). These include stock-market (or financial market) volatility, GDP and income volatility, forecaster disagreement (i.e. the standard-deviation across economic forecasts from a number of different institutions), news mentions of the term 'uncertainty' and other related terms (Baker *et al.* (2015)), and differences between the actual release values for variables such as GDP and their pre-release expected values. Other measures include unemployment expectations (Carroll and Dunn (1997)), sentiment indicators (Bachmann *et al.* (2013); Ling *et al.* (2015)), and internet searches for terms related to uncertainty (Dzielinski (2012)).

According to Bloom (2013), economic uncertainty is generally caused by the same events that cause recessions, such as oil-price shocks and credit crunches. This is further compounded by the fact that recessions themselves increase uncertainty, meaning that as economic growth deteriorates uncertainty is endogenously increased further. Pástor and Veronesi (2012) and Kozeniauskas *et al.* (2016) argue that it is the unfamiliarity of recessions which leads to an increase in uncertainty. In these situations, fiscal and monetary policy become more unpredictable as policy-makers attempt innovative ideas in order to boost economic growth, and find it more difficult to forecast something different from the usual pattern of positive growth.

The main link between uncertainty and house prices is that uncertainty leads consumers to be more cautious when making purchases. This is especially so for residential property, which involves a large outlay of money and in most cases a bank loan. Moreover, house purchases are very difficult to reverse, and, unlike purchases of necessities, can be delayed through a 'wait and see' approach. Consumers also tend to put away more savings as a precautionary measure in periods when uncertainty over their future income is high. Bertola *et al.* (2005) concluded that an increase in uncertainty reduces consumers' durable expenditure, while Ling *et al.* (2015) argued that house prices are affected by changes in sentiment among important market participants.

Identifying individual sources of uncertainty is difficult, and most commonly the total expected forecast uncertainty is reported. Typically this is presented as either standard deviations (usually with an underlying assumption of normality), or fan charts (densities). Calculation methods differ across the major forecasting institutions. Most common uncertainty measures are explicitly based on past forecasting errors and include those linked to mean-absolute forecast errors (MAFE) or root mean squared forecasting errors (RMSFE). These can be derived from a static specification, but are more commonly based on recursive model estimates and are usually simple to calculate and interpret. Such measures are used by a large number of forecasters – for example,

1. An economic distinction between uncertainty and risk was proposed by Knight (1921). According to Knight, 'risk is present when future events occur with measurable probability. Uncertainty is present when the likelihood of future events is indefinite or incalculable.'

OECD in their Interim Outlook, FOMC and FEDs, Bank of England, Bank of Canada, Sveriges Riksbank and ECB/ESCB or Bundesbank. The main limitations of the simplest approach are the normality assumption, proneness to large outliers (the ECB and OECD exclude some particularly large outliers from the calculations) and lack of relationship to most recent developments.

From the point of view of speculators, who purchase property solely for investment purposes, the return-risk ratio is negatively impacted by increased uncertainty over expected returns, and increased costs of financing from banks unwilling to lend in an environment of higher default risk.

The aim of this paper is to discuss forecasting models and the importance of variables proxying for economic uncertainty as predictors of residential property prices in Portugal and Spain. We use a forecasting methodology known as dynamic model averaging (DMA) applied to house price dynamics encompassing a wide set of variables. These include macroeconomic determinants, such as income, GDP, labor force, unemployment and interest rates but also shorter-term drivers, such as housing investment, housing loans, business and consumer confidence and financial markets volatility. DMA owes its success in part to its inherit flexibility, not only by incorporating uncertainty across different forecasting models, but also uncertainty pertaining to each parameter within any given forecasting model. This is done through the use of model averaging, and the usage of two forgetting factors that reflect uncertainty in both parameters and models.

The paper is organized as follows. Section 2 briefly describes the DMA methodology used in the forecasting exercise of house prices in Portugal and Spain. Section 3 discusses the data and evaluates the forecast performance of the DMA methodology. This evaluation includes a discussion on the usefulness of each predictor for forecasting at any given time; and provides a further analysis using factors extracted from the predictors using principal component analysis to forecast house prices, thereby reducing the dimension of the predictor set. Lastly, Section 4 concludes.

2. Methodology

The forecasting methodology employed in this analysis is known as dynamic model averaging (DMA). Initial work on the methodology was done by Raftery *et al.* (2010) who applied it to an industrial context. Later Koop and Korobilis (2012) adapted it to forecast inflation. Koop and Korobilis found evidence suggesting that DMA was a superior forecasting method when compared to several alternatives, including other time varying parameter models. Since then a number of studies have used the methodology in a variety of contexts. In terms of house price forecasting, Bork and Møller (2015) analyzed the performance of DMA to forecast average house prices of US states, Risse and Kern (2016) applied the method to European house prices, and Hill and Rodrigues (2020) used DMA to forecast house prices of major economies using a new dynamic forgetting (DF) strategy. Overall, DMA has been shown to be a valuable tool for macroeconomic forecasting. For other applications of the DMA methodology see, for instance, Moretti *et al.* (2019) and Nicoletti and Passaro (2012).

We will briefly discuss the DMA approach, emphasizing its relevance for uncertainty. We apply different tuning parameters to the DMA which reflect the flexibility that the methodology has in terms of model averaging and time variation of model coefficients. We also use dynamic model selection (DMS), which can be seen as a special case of the DMA approach described below. For more technical details the reader is referred to Raftery *et al.* (2010), Koop and Korobilis (2012) and Hill and Rodrigues (2020).

2.1. Between Model Uncertainty

The DMA procedure is initiated with the researcher specifying a set of potential models. In practice, this usually means selecting a group of predictor variables and generating a set of linear models with all possible combinations of predictors. For instance, for K predictors there would be 2^K different linear models. DMA then uses Bayesian model averaging of each model's forecast to generate the forecasts. The averaging is Bayesian in the sense that weights assigned to each model are based on how well each model performed in the past. Let $\hat{y}_t$ be the forecasted variable of interest, in our case house prices and let each of the 2^K models be labeled as M_k, $k \in (1, ..., 2^K)$. The weighted average is computed as,

$$\hat{y}_t = \sum_{k=1}^{2^K} P(model_t = M_k \mid \mathcal{F}_{t-1}) \hat{y}_{t(k)} \tag{1}$$

where $\hat{y}_{t(k)}$ is the forecast from model k, $\mathcal{F}_{t-1}$ represents the information set available at the time of the forecast and $model_t$ refers to the forecasts generating model. The posterior probability weight $P(.)$ changes according to how well one of the k models forecasts in comparison with all the other available models. Weights are updated after each iteration. The update involves prior probabilities of model k, as well as a normal likelihood with mean $\hat{y}_{t(k)}$ and the predicted variance evaluated at the actual y_t. One important contribution of Raftery *et al.* (2010) was the use of a forgetting factor, labeled α, that reflects the degree of model uncertainty. The parameter α dictates how much uncertainty we wish to attach to the posterior weight as it is updated, i.e. becomes the prior in the next iteration. With fixed forgetting factors, the researcher can set α between 0 and 1, with lower values reflecting more model uncertainty. In practice α is usually set somewhere between 0.95 and 1. With $\alpha < 1$, models that perform better than the average receive proportionally less weight than they would if $\alpha = 1$, while weights of the models performing worse than average receive a higher weight. The lower α the stronger this effect. Therefore, DMA allows for initial uncertainty by allowing the researcher to be unsure of the data generating model, and α flattens the distribution over all possible models.

We also conduct analysis using a model selection framework based on dynamic model selection (DMS). This is shown in Table 1 below as dynamic model selection, or Bayesian model selection (DMS or BMS). In this setting there is no averaging over models as described above, instead the forecast comes from the model with the largest weight. This is a special case of DMA, in which the weights are 1 for $\max_k(P(model_t = M_k \mid \mathcal{F}_{t-1}))$ and 0 for all the other $K-1$ models. The model with the largest weight acts as the *selected* model and gives the exclusive forecast for a particular period. In the following period, weights are adjusted according to how well each model performed in the past, with α dictating, as indicated above, how much memory is involved in the process. This gives DMS an advantage when there appears to be one model that outperforms, while other models are confounding.

2.2. *Parameter Uncertainty*

Parameter uncertainty in the DMA framework is addressed through the use of state space methods, namely the Kalman filter. We can formulate this with a measurement and a state equation as,

$$y_t = \mathbf{x}'_{t-1}\boldsymbol{\theta}_t + \varepsilon_t \tag{2}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\nu}_t \tag{3}$$

where ε_t and $\boldsymbol{\nu}_t$ are normally distributed error terms with ε_t being a scalar and $\boldsymbol{\nu}_t$ a vector of the same dimension as $\boldsymbol{\theta}_t$, $\boldsymbol{\theta}_t$ is a vectors of coefficients, and $\mathbf{x}_t$ a vectors of predictors. Raftery *et al.* (2010) provide a more detailed explanation, here we simply state that parameter uncertainty is accommodated via the coefficient vector $\boldsymbol{\theta}$, that evolves as a random walk. The Kalman filter can be thought of as a recursive least squares approach, which iteratively solves an OLS problem, thus giving a series of coefficient estimates. To avoid the filter converging on a specific $\boldsymbol{\theta}$, the DMA includes a second forgetting factor defined as λ, which effectively places more weight on recent observations of the recursive OLS problem, thereby allowing for some uncertainty in the coefficients. There are a number of ways of considering λ. Raftery *et al.* (2010) and others consider it to be a constant parameter which is set by the researcher *a priori*. As with α, λ in practice takes values between 0.95 and 1, with 1 indicating recursive OLS where recent and past observations carry the same weight. A lower λ increases the flatness of the coefficient covariance matrix implying more uncertainty over the generating process of $\boldsymbol{\theta}_t$. A drawback of a lower λ is that it makes the system more susceptible to noise which causes the filter to over-adjust.

Hill and Rodrigues (2020) explore a solution to the over-adjustment problem by employing predictor specific dynamic forgetting factors. This allows the filter to permit more uncertainty in the process, without over-fitting noise. The individual forgetting factors also decrease (originating more forgetting) when forecast errors are large, implying that, for instance, a structural break in the generating process will increase forgetting across all predictors. The main idea in Hill and Rodrigues is to limit the size of the covariance matrix from above and below, so as to allow forgetting, without the drawback of over-sensitivity (we present results using fixed as well as dynamic forgetting in Table 1). The role of the forgetting factor λ is the same under model averaging or model selection as its role is related to the rate of variation of parameters over time within each model.

3. Forecasting House Prices

3.1. Data

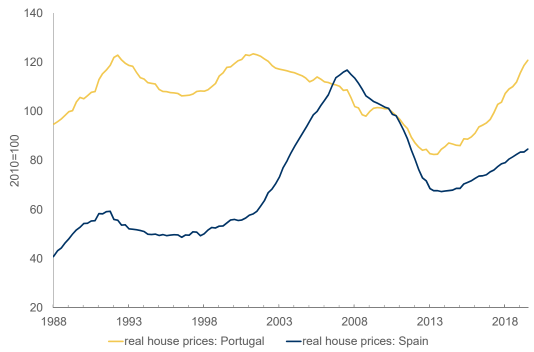
Real estate market dynamics have gained particular interest in recent years, following the US sub-prime collapse in 2007 which quickly spread worldwide and led to a significant impact of housing markets on the economy. Understanding the price determination process in real estate markets is of foremost importance if we want to forecast. Determinants of housing demand include growth in household disposable income and gradual shifts in demographics, such as the relative size of older and younger generations. Permanent features of the tax system that might encourage home ownership as opposed to other forms of wealth accumulation also matter, as well as the average level of interest rates possibly related to the long-term behavior of inflation. The availability and cost of land, as well as the cost of construction and investments in the improvement of the quality of existing housing stock are also relevant (Poterba *et al.* (1991) and Tsatsaronis and Zhu (2004)). For instance, the growth of the housing stock can be constrained in the short run as a result of a number of factors that include the length of the planning and construction. There could also be shorter-term drivers related to constraints in the growth of the housing stock, prevailing conditions in the provision of housing loans, or uncertainty about future prospects. Higher GDP and disposable income, more confidence in the economy, less unemployment, more labor and an increase in mortgage lending are expected to have a positive impact on the housing market. In contrast, higher interest rates are expected to drive borrowing costs up and demand down leading to a subsequent fall in house prices and make alternative applications of wealth more interesting. The same with residential investment, if it increases prices may go down.

The predictors we use in our analysis consist of fundamental macroeconomic covariates, such as real money market rate, labor force, real disposable income *per capita*, real GDP *per capita*, real mortgage rates, real gross fixed capital formation (GFCF) in housing, real loans for house purchases, and the unemployment rate. However, we also include other variables that attempt to gauge uncertainty, such as, business and consumer confidence and financial markets volatility which we also expect to have a positive impact on house prices.

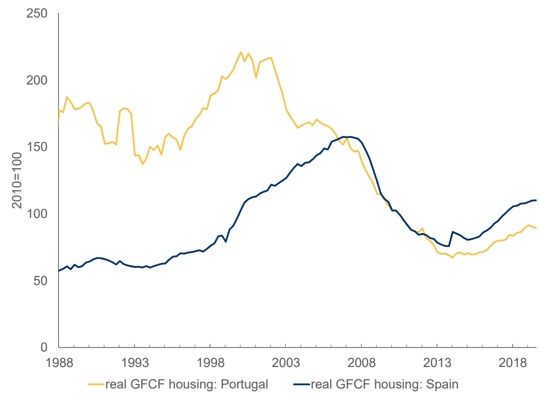
Our data set comprises quarterly time series from 1988:Q1 to 2019:Q3 for Portugal and Spain. Data on house prices, real GDP, real GFCF in housing, disposable income, labor force, unemployment, population and private consumption deflator were collected from the OECD, the Eurostat, Statistics Portugal and Banco de Portugal, while loan for house purchases, short-term interest rates and mortgage rates were taken from the European Central Bank. Short-term interest rates correspond to 3-month inter-bank money market yield rates. Mortgage rates correspond to the interest rate on loans for house purchase. Confidence data refer to the Economic Sentiment Indicator of the European Commission Surveys. Historical volatility from the PSI-20 and IBEX 35 is the annualized standard deviation of 60-day average of daily volatility. The VIX, VDAX and

VSTOXX are market indexes representing the market's expectation of 30-day forward-looking volatility based on the price inputs of the S&P 500, DAX and EuroStoxx50 index options. These were taken from Refinitiv. All series in real terms were computed using the private consumption deflator. GDP and GFCF in housing are chain linked volume. House price indices correspond to seasonally unadjusted series constructed from national data from a variety of public and/or private sources, such as, national statistical services, mortgage lenders and real estate agents. House price series may differ in terms of dwelling types and geographical coverage. For Portugal and Spain they are country-wide and refer to newly and existing apartments. The house price indexes are based on hedonic approaches to price measurement characterized by valuing the houses in terms of their attributes (average square meter price, size of the dwellings involved in transactions and their location).

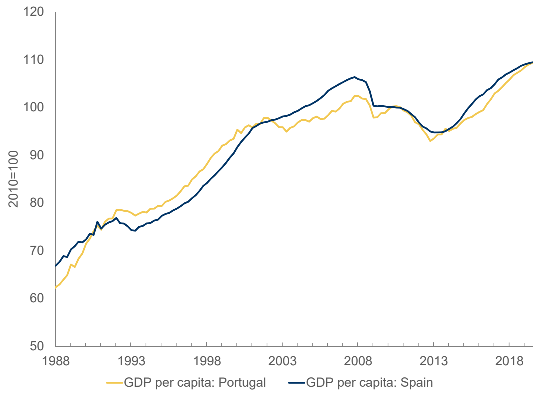
Before analyzing the empirical results it is important to briefly describe the evolution of the real estate markets, house prices and macroeconomic variables. During two decades, until the beginning of the financial crisis in 2007, house prices grew on average less than 1 per cent per year in real terms in Portugal and 7 per cent in Spain (Figure 1a)). Since the crisis and until the end of 2019 house prices fell 2 per cent on average in Spain and increased 1 per cent in Portugal. However, this masks a highly differentiated evolution over the past decade. House prices declined in both economies, though more in Spain, between 2008 and until the recovery in 2013, and increased in both countries over the past five years, especially so in Portugal. In terms of activity, there was a major difference between Portugal and Spain from the late 90's and until 2007, associated to the impact of immigration flows to Spain resulting in a significant increase of active population at the beginning of the XXI century, which probably contributed to an increase in housing demand (Lourenço and Rodrigues (2014)). During this period, Spanish residential investment grew at an average annual rate of about 8 per cent, while in Portugal it recorded a 2 per cent contraction (Figure 1b)). In turn, GDP accelerated slightly in both economies, although less in the Portuguese case (Figure 1c)). In the five years following the financial crisis and until the recovery in 2013, both countries saw a similar contraction in GDP and housing investment, although most strongly in terms of investment, over 11 per cent compared to 1 per cent in GDP. The unemployment rate increased sharply and labor force declined, which may be related to emigration flows (Figures 1d) and 1e)). Between 2014 and 2019, amidst increasing confidence, GDP accelerated 2 per cent in Portugal and in Spain and residential GFCF increased 4 and 6 per cent, respectively (Figure 2a)). Given its relevance for the housing sector and the impact it may have on the cost of financing it is also important to analyze credit in detail. Data on bank lending indicate the existence of episodes of very high growth in mortgage loans between the mid-1990s and 2007 (Figure 2b)). This annual growth was about 15 per cent on average in Portugal and in Spain, in the context of declining costs of bank loans and high and sustained growth in household disposable income, which was reflected in an increase of indebtedness of families (Figures 1f) 2b) and 2c)). The significant deceleration of credit to housing from 2010 onward should be seen in the context of the international financial crisis which had a negative impact on the supply, given a significant tightening in lending conditions, and on housing credit demand. The volatility variables (Figure 2d)) display spikes during the crises (e.g. subprime crisis followed by a recession and sovereign debt crisis).



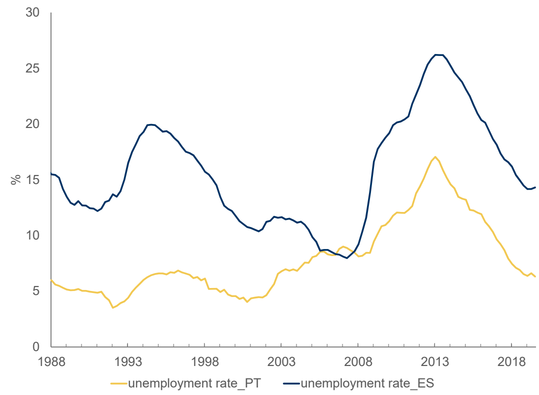
a) House Price Index



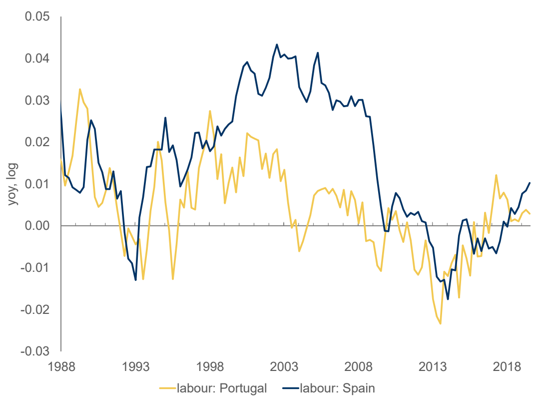
b) Housing GFCF



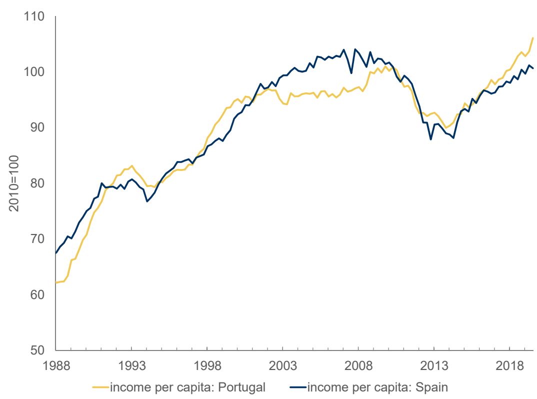
c) GDP



d) Unemployment Rate

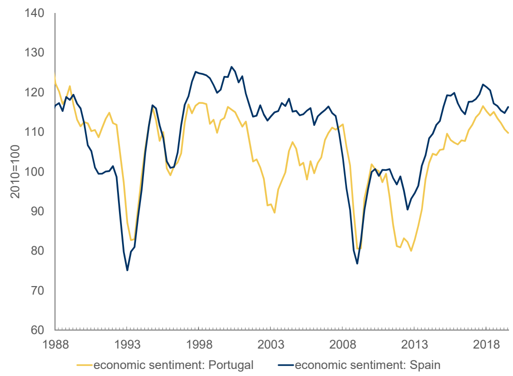


e) Labour Force

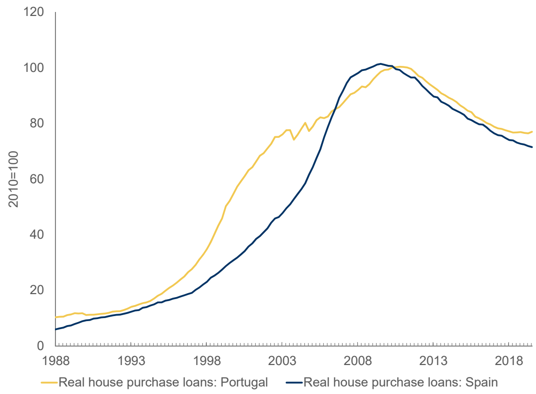


f) Real Income

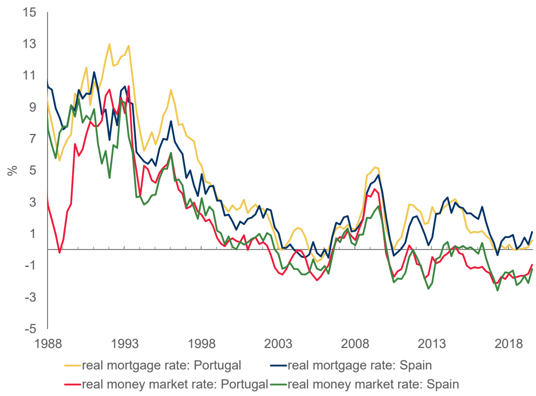
FIGURE 1: Plots of variables used in the analysis



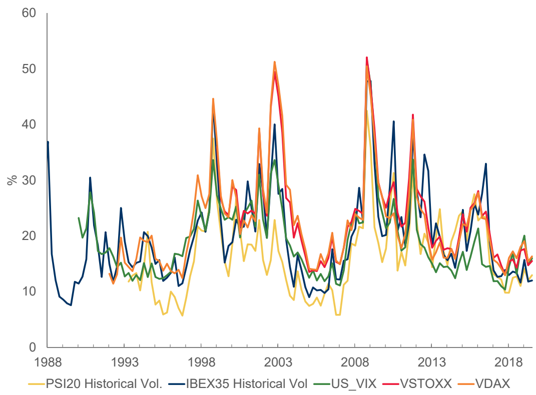
a) Confidence Indicator



b) Loans



c) Mortgage Rates



d) Volatility Measures

FIGURE 2: Plots of variables used in the analysis

3.2. *Empirical Results - All Variables*

In our analysis we consider forecasts from a number of DMA specifications, from DMS and from a first order autoregressive model (AR(1)). The forgetting factor α is generally fixed between 0.95 and 1 in most applications of DMA in the literature. We find that varying α within this range does not alter forecasts significantly. Therefore for simplicity, we fix α at 0.97 in our analysis and allow λ to vary. We select four different specifications for λ; 0.95, 0.99, 1 and the dynamic forgetting of Hill and Rodrigues (2020), denoted by the superscript DF.

We measure the performance of each approach using the mean squared forecast error (MSFE), as well as the mean absolute forecast error (MAFE) for the pseudo out-of-sample period starting from T_0 =2009 Q1 to the end of the sample at T =2019 Q3. The MSFE is computed as $\sum_{t=T_0}^{T}(y_t - \hat{y}_t)^2/T_{os}$ and the MAFE as $\sum_{t=T_0}^{T}|y_t - \hat{y}_t|/T_{os}$, where T_{os} is the number of out-of-sample periods. We also report the p-values of the Clark and West (2007) test of equal predictability performance and the out-of-sample R^2 (R^2_{os}) given by

$$R^2_{os} = 1 - \frac{\sum_{t=T_0}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=T_0}^{T}(y_t - \bar{y}_t)^2}$$

where $\bar{y}_t$ is the historical average of the y_t series and $\hat{y}_t$ is the forecast from our model in question. The R^2_{os} is positive if the forecasting model beats the historical average, while the opposite is true if the R^2_{os} is negative.

3.2.1. Forecast Performance

Table 1 shows that in general, across all time periods, the dynamic forgetting (DF) approach appears to have an advantage over fixed forgetting for the one period ahead forecasts. For longer periods, the results are mixed. For Spain the baseline AR(1) model seems to out-perform all competing models when $h = 4$ is considered. It is not clear which parametrization of λ offers the best forecast performance across both Portugal and Spain. This suggests the need for a forgetting scheme that is adaptable to different data generating processes.

Aside from the dynamic forgetting, at one period ahead, a low λ indicating more forgetting appears to have smaller forecast errors. This suggests that parameters in equation 3 provide better forecasts when we increase their variance, in other words increasing the uncertainty of the parameter estimates and thus preventing the Kalman filter from stabilizing provides us with lower forecast errors. In terms of discounting past data, a λ of 0.95 means that data at $t - 4$ carry about 80% of the weight as data at time t. For two periods ahead, the results for the two countries are more mixed. Dynamic model selection, in which the model weights are 1 for the best performing model and 0 otherwise, has the lowest forecast error for Spain, indicating that the data generating process may closely resemble one specific model, while other potential models tend to miss the mark. In this case the models that are selected for most of the out of sample period are the models that include only the lagged dependent variable and intercepts and the model that includes lagged real GDP *per capita* and loans for house purchases. In both Portugal and Spain, the two period ahead forecast horizon is not dominated by a single specification, instead both model selections with dynamic forgetting appear to do well. At four periods ahead, the high forgetting specification and dynamic forgetting do well in the case of Portugal, whereas in Spain, the AR(1) outperforms the DMA and DMS.

Portugal

	h=1				h=2				h=4			
forecasting method	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$
DMA $\alpha = 0.97, \lambda = 0.95$	0.2160	1.1123	**0.0000**	0.4141	0.2464	1.2520	0.0002	0.2724	**0.2994**	**1.4475**	0.0003	-0.0078
DMA $\alpha = 0.97, \lambda = 0.99$	0.2170	1.1131	0.0001	0.4086	0.2412	1.2152	0.0001	0.3027	0.3028	1.5247	0.0014	-0.0305
DMA $\alpha = 0.97, \lambda = 1$	0.2174	1.1107	0.0002	0.4062	0.2398	1.2025	0.0001	0.3106	0.3011	1.5231	0.0020	-0.0190
DMA $\alpha = 0.97, \lambda^{DF}$	**0.2149**	1.1104	0.0000	**0.4197**	0.2411	1.2146	0.0001	0.3034	0.2996	1.4949	0.0010	-0.0086
AR1	0.2225	1.1161	-	0.2654	0.2540	1.2574	-	0.1986	0.3207	1.6266	-	**0.0521**
DMS $\alpha = 0.97, \lambda = 0.95$	0.2188	1.1717	0.0002	0.3986	0.2314	**1.1214**	**0.0001**	0.3579	0.3132	1.5167	**0.0001**	-0.1026
DMS $\alpha = 0.97, \lambda = 0.99$	0.2192	1.1065	0.0031	0.3964	0.2378	1.1909	0.0030	0.3223	0.3087	1.5215	0.0035	-0.0714
DMS $\alpha = 0.97, \lambda = 1$	0.2163	**1.0885**	0.0043	0.4124	0.2426	1.2202	0.0061	0.2947	0.3045	1.5094	0.0049	-0.0425
DMS $\alpha = 0.97, \lambda^{DF}$	0.2154	1.0897	0.0009	0.4170	**0.2307**	1.1310	0.0008	**0.3622**	0.3099	1.5033	0.0013	-0.0793

Spain

	h=1				h=2				h=4			
forecasting method	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$
DMA $\alpha = 0.97, \lambda = 0.95$	0.1949	0.9768	0.0050	0.7160	0.2222	1.0400	0.0113	0.6394	0.3864	1.8016	**0.0072**	-0.0677
DMA $\alpha = 0.97, \lambda = 0.99$	0.1948	0.9820	0.0036	0.7163	0.2139	1.0333	0.0110	0.6660	0.3463	1.6295	0.0085	0.1426
DMA $\alpha = 0.97, \lambda = 1$	0.1980	0.9957	**0.0035**	0.7070	**0.2107**	1.0325	0.0095	**0.6758**	0.3330	1.5751	0.0089	**0.2069**
DMA $\alpha = 0.97, \lambda^{DF}$	**0.1935**	**0.9731**	0.0043	**0.7200**	0.2169	**1.0224**	0.0125	0.6566	0.3609	1.6834	0.0082	0.0683
AR1	0.2283	1.0577	-	0.1025	0.2213	1.0470	-	0.1382	**0.3212**	**1.3483**	-	0.0729
DMS $\alpha = 0.97, \lambda = 0.95$	0.2070	1.0206	0.0067	0.6798	0.2251	1.0359	0.0194	0.6299	0.4022	1.8920	0.0119	-0.1569
DMS $\alpha = 0.97, \lambda = 0.99$	0.2146	1.1079	0.0076	0.6557	0.2223	1.1120	0.0129	0.6391	0.3683	1.7581	0.0088	0.0299
DMS $\alpha = 0.97, \lambda = 1$	0.2092	1.0885	0.0084	0.6730	0.2129	1.0646	**0.0077**	0.6689	0.3643	1.7419	0.0097	0.0508
DMS $\alpha = 0.97, \lambda^{DF}$	0.2017	1.0204	0.0067	0.6958	0.2272	1.0884	0.0142	0.6232	0.3770	1.7803	0.0084	-0.0163

TABLE 1. Results when all predictors are considered

It is important to note that recent figures do not yet reflect the impact of the Covid-19 pandemic on the world economy. The unprecedented nature of this crisis makes it challenging to gauge its repercussions on the predictors used in this analysis, and of course on house prices themselves. The impact will, to a certain extent, depend on changes in fundamentals that support the housing market, such as, banks funding lines, interest rates, housing shortages within key locations, and unemployment, where particularly the latter may be one of the main stress factors in the coming months.

The pandemic will likely result in structural changes in the housing market and in the marginal effects of house price predictors. However, the model averaging as well as the time varying parameter characteristics of DMA should grant the forecasting method adequate flexibility to incorporate these structural changes. The speed at which DMA will be able to react to the changes brought on by the pandemic will depend on the tuning parameters α and λ, that represent the forecasters uncertainty pertaining to the set of predictors and the marginal effect of each predictor in this set.

3.2.2. *Posterior Probability of Inclusion Plots*

An interesting feature of the DMA approach is that using the posterior probability distributions (PIPs) from each model, one can construct inclusion probabilities for each predicting variable. Every model in the DMA model set that contains a particular variable is given a PIP upon propagation of the Kalman filter. The total probability attached to each of these models is then used as a posterior probability of inclusion for a given predictor. The PIPs are presented in Figures 3 and 4 for Portugal and Figures 5 and 6 for Spain. The inclusion probabilities have been divided between economic and financial/volatility predictors. The PIPs for the lagged autoregressive predictor, as well as for the constant are not included, given that these have been very high and relatively stable over the whole period considered. Each line represents the probability that the corresponding predictor is included in the applied model for a given period. In other words, each line represents the relative importance of a predictor for forecasting house prices. As shown in Figures 3 to 6 some of the variables change significantly over time. This indicates that a forecasting framework that incorporates model uncertainty is justified.

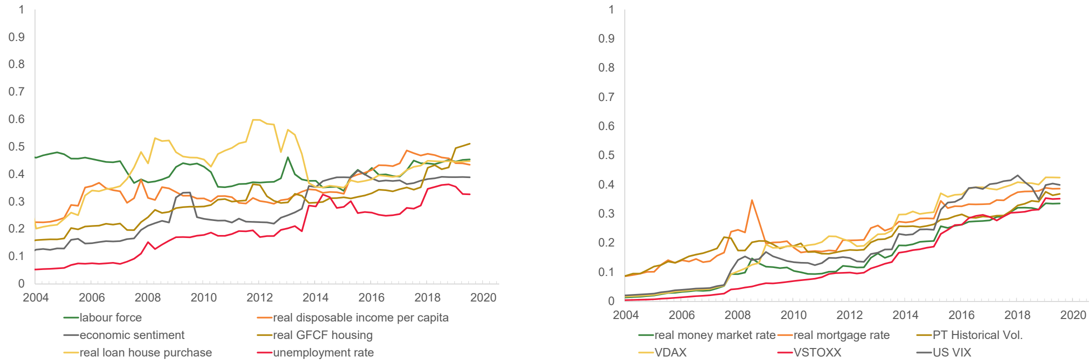


FIGURE 3: Portugal - one period ahead forecast horizon

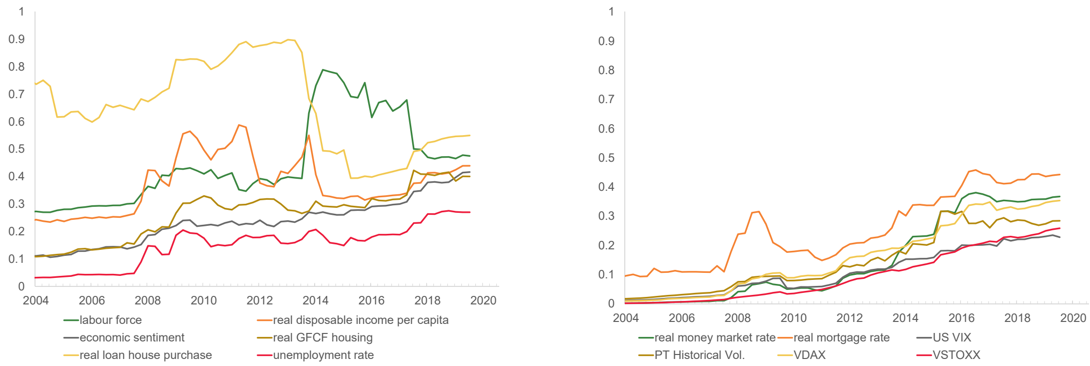


FIGURE 4: Portugal - four periods ahead forecast horizon

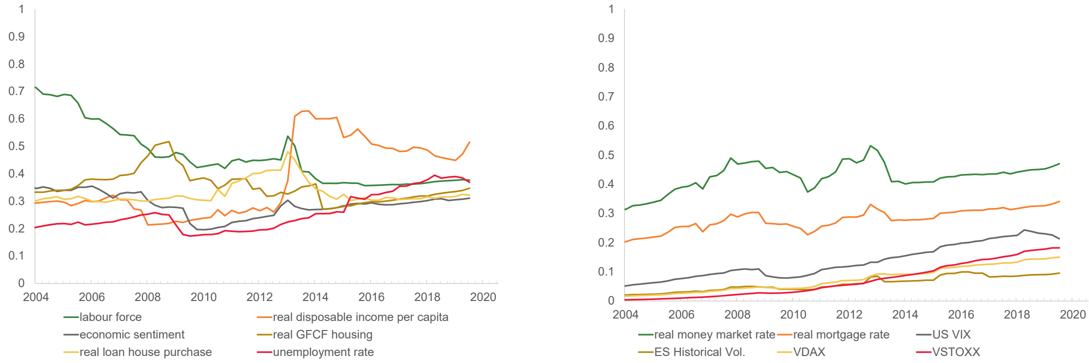


FIGURE 5: Spain - one period ahead forecast horizon

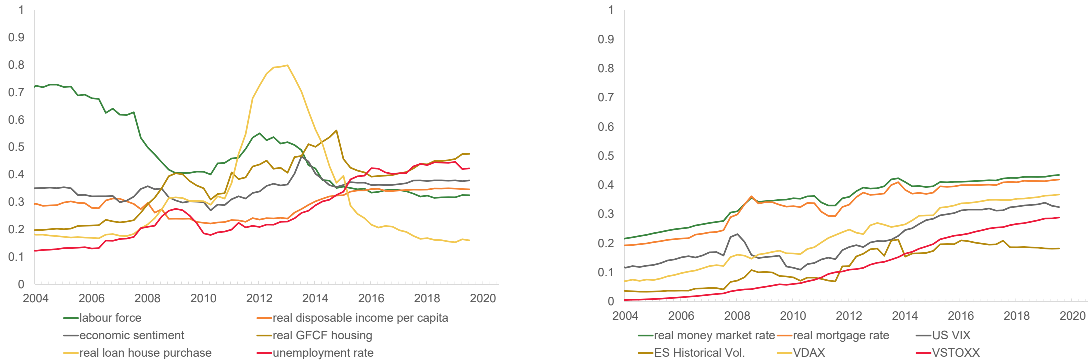


FIGURE 6: Spain - four periods ahead forecast horizon

The PIPs should be regarded as a measure of a particular predictor's importance for forecasting relative to other competing models. This is because the construction of the PIPs is based on the model weights used in the model averaging, which all sum to one and increase or decrease in proportion to the model's performance relative to the forecast performance of all other models. Therefore, when we see the financial and volatility predictors increasing over time as a group in Figures 3 and 4, this suggests a steady increase in importance of volatility and interest rate variables relative to other predictors. This includes the first lag of real house prices, which although has a PIP of 1, since it is included in all the models which contain other predictors, has a model averaging weight that could be decreasing in time and does not show up in the chart of PIPs.

For example, in the one period ahead PIPs for Portugal we notice the set of financial predictors all being relatively clustered together and increasing over time. The clustering of the PIPs for financial and volatility predictors likely stems from their strong correlation. We also see a slightly weaker increase in the PIPs of the real economic variables. Since all PIPs are increasing, they may be doing so at the expense of the lag of house prices. A possible interpretation of this is that the autoregressive property of the house price series is decreasing in favour of predictability from other variables, particularly financial/volatility variables.

For the four quarter forecast horizon, we notice more separation between the inclusion probabilities of predictors, particularly for housing loans. This predictor displays strong predictive power for the 2004 to 2013 period. It is later replaced by labour force in terms of significance after around 2013. The steady increase of PIPs for the financial/volatility predictors are, as per the one quarter ahead forecast, likely the result of the autoregressive property of the differenced house price series weakening. Given the DMA's lower performance for longer forecast horizons, it could be the case that the decrease in the autoregressive predictability of differenced real house prices is not being compensated by an improving predictive power of other predictors. Agents on the supply and demand side could be reacting quicker to changes in fundamentals and driving price changes in short horizons as opposed to longer ones.

Both sets of PIPs for Portugal suggest some volatility in terms of model switching around 2013. By itself, this is not sufficient evidence for a regime change in terms of the drivers of house prices, but it does suggest that dynamics may have shifted around this period. This question warrants further investigation; see e.g. Lourenço and Rodrigues (2017) and section 3.3.

The PIPS from Spain's one quarter ahead forecasts suggest that the labour force predictor was a relatively important variable for forecasting house prices up until 2013. This is the case for both one and four quarter ahead forecasts. However, after roughly 2013, in both forecast horizons the importance of labour force drops. Its importance is replaced by real disposable income for one quarter ahead, while for the four quarter ahead forecast there does not appear to be any variable that stands out aside from real house loans which jumps around 2013 but fall shortly afterwards. The financial and volatility predictors for Spain co-move over the forecast period, again likely due to the strong correlation of those predictors. Interestingly however, the real money market rate maintains a steady importance relative to other predictors over the forecast period. It is interesting to note that, similar to the case of Portugal, the shifts in the PIPs for one quarter ahead forecasts suggest some significance surrounding the year 2013. We also note that this coincides with the beginning of Spain's economic recovery. This corroborates work done on the subject of the Spanish housing market by Cuestas and Kukk (2019) who identify Q2 2013 as a break date in their analysis of drivers of house prices in Spain.

The differences in the dynamics of PIPs between Portugal and Spain is not clear. Both housing markets were affected by external factors over the sample period. However real loans for housing in Portugal appear to have played an important role pre-2013 for both long and short forecast horizons. Loans became an important driver in Spain during the bust period between 2008 and 2013. After that, real disposable income plays an important role. During the bust period, the availability of loans decreased as more restrictions were placed on lending thereby leading to a fall in housing loans observed in both countries, this also coincides with a fall in housing price and results in the increased importance of the housing loans predictor in both cases. The difference in the importance of the volatility predictors between the two countries is also interesting. Volatility, as a proxy for economic uncertainty, seems to play a bigger role in Portugal than in Spain. This could be due to a number of reasons. Portuguese lenders and buyers may be more cautious during higher volatility than their counterparts in Spain, and the composition of buyers could also be different.

3.2.3. *Posterior Size Plots*

A further interesting feature of the DMA methodology is the concept of posterior size probability. Each of the models which are averaged over in the DMA process contains a certain number of predictors. Taking the number of predictors of each model, and producing a weighted average using the posterior predictive probability of that model as a weight, provides an indication of the number of variables actually used to predict the change in real house prices (Koop and Korobilis (2011). For both Portugal and Spain, the number of predictors appears to increase over time. This is consistent with the posterior probability of inclusion plots, which show the probability of inclusion for most variables increasing over time. The distinct increase in the size of the best performing models indicates that they are changing over time. This suggests that the *a priori* model uncertainty was justified, since the 'optimal' model changes over time, i.e., it seems that there was not a single model specification that was appropriate over the whole sample. Hence, the implicit uncertainty, and updating in the DMA process was utilized in the forecasts. Had the lines in these charts been more or less constant, there would be less evidence for model change throughout the sample.

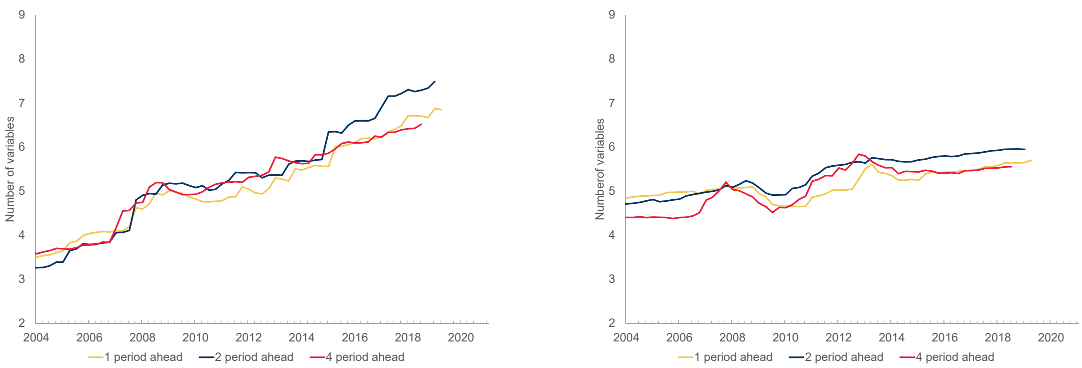


FIGURE 7: Left graph is for Portugal and the right graph for Spain

3.3. *Empirical Results - Factors*

Factor models have been used in a wide variety of forecast applications and have been found to be useful for dimension reduction which often improves out-of-sample forecasting. An added benefit is that the computational burden is drastically reduced when 12 predictors are replaced by 3 factors as we have done. We follow Koop and Korobilis (2011) and create block factors. We divide the predictors into three blocks; an economic uncertainty block consisting of the volatility indices; a financial block consisting of the real money market and mortgage rates, and a real economy block consisting of the remaining predictors. Thus, we extract the common variation in each predictor block, and use the resulting factor in place of the original predictors. This leaves us with $2^3 = 8$ models to average over in the DMA/DMS procedure. We extract factors using an eigenvalue decomposition of the standardized block predictor matrix. We confirm that one principal component captures most of the within block variation by examining the relative size of the largest eigenvalue and finally use these eigenvalues to extract the factors for the block matrix.

The analysis of the results suggests that there are gains in using factors instead of a large number of predictors. This is demonstrated by lower forecast errors in most cases across both countries and forecast horizons, in particular at longer forecast horizons. This suggests the there was perhaps some mild over fitting occurring in the DMA using all predictors, although the difference is not substantial enough to markedly alter the forecasts in this case.

Portugal

	h=1				h=2				h=4			
forecasting method	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$
DMA $\alpha = 0.97, \lambda = 0.95$	0.2142	1.1166	0.0000	0.4237	0.2401	1.1937	0.0000	0.3087	0.2826	1.4199	0.0018	0.1023
DMA $\alpha = 0.97, \lambda = 0.99$	0.2139	1.0834	0.0001	0.4250	0.2397	1.1827	0.0002	0.3113	0.2866	1.4499	**0.0006**	0.0771
DMA $\alpha = 0.97, \lambda = 1$	0.2138	1.0763	0.0002	0.4255	0.2388	1.1793	0.0008	0.3165	0.2853	1.4510	0.0009	0.0851
DMA $\alpha = 0.97, \lambda^{DF}$	0.2131	1.1186	**0.0000**	0.4293	0.2391	1.1972	**0.0000**	0.3146	0.2775	1.3671	0.0015	0.1343
AR1	0.2225	1.1161	-	0.2654	0.2540	1.2574	-	0.1986	0.3207	1.6266	-	0.0521
DMS $\alpha = 0.97, \lambda = 0.95$	0.2134	1.0926	0.0001	0.4282	0.2371	1.1641	0.0003	0.3259	0.2801	1.3958	0.0037	0.1180
DMS $\alpha = 0.97, \lambda = 0.99$	0.2128	**1.0706**	0.0002	0.4309	0.2371	1.1657	0.0000	0.3259	0.2870	1.4511	0.0018	0.0743
DMS $\alpha = 0.97, \lambda = 1$	0.2133	1.0719	0.0005	0.4286	0.2360	1.1690	0.0001	0.3323	0.2853	1.4400	0.0033	0.0854
DMS $\alpha = 0.97, \lambda^{DF}$	**0.2120**	1.0881	0.0001	**0.4353**	**0.2336**	**1.1502**	0.0002	**0.3460**	**0.2742**	**1.3543**	0.0028	**0.1548**

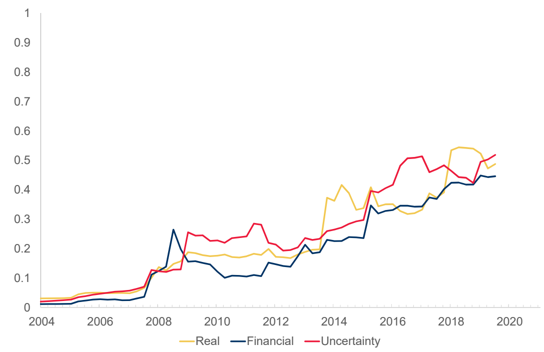
Spain

	h=1				h=2				h=4			
forecasting method	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$	**MSFE**	**MAFE**	**CW test**	$\mathbf{R^2_{os}}$
DMA $\alpha = 0.97, \lambda = 0.95$	0.1860	0.8860	0.0183	0.7414	0.2129	0.9550	0.0280	0.6690	0.3459	1.5850	0.0193	0.1445
DMA $\alpha = 0.97, \lambda = 0.99$	0.1912	0.9102	0.0227	0.7265	0.2090	0.9611	0.0367	0.6809	0.3170	1.4260	0.0294	0.2815
DMA $\alpha = 0.97, \lambda = 1$	0.1954	0.9403	0.0199	0.7146	0.2085	0.9687	0.0402	0.6826	0.3081	1.3816	0.0317	0.3211
DMA $\alpha = 0.97, \lambda^{DF}$	**0.1842**	**0.8756**	0.0166	**0.7463**	0.2119	0.9443	0.0275	0.6720	0.3543	1.6434	0.0133	0.1021
AR1	0.2283	1.0577	-	0.1025	0.2213	1.0470	-	0.1382	0.3212	**1.3483**	-	0.0729
DMS $\alpha = 0.97, \lambda = 0.95$	0.1867	0.9002	0.0149	0.7393	0.2087	0.9529	0.0290	0.6819	0.3384	1.5504	0.0160	0.1809
DMS $\alpha = 0.97, \lambda = 0.99$	0.1911	0.9140	0.0218	0.7270	**0.2080**	0.9812	0.0289	**0.6841**	0.3127	1.4165	0.0204	0.3008
DMS $\alpha = 0.97, \lambda = 1$	0.1995	0.9783	**0.0141**	0.7024	0.2094	1.0080	**0.0248**	0.6798	**0.3029**	1.3576	0.0223	**0.3439**
DMS $\alpha = 0.97, \lambda^{DF}$	0.1855	0.8904	0.0145	0.7426	0.2080	**0.9410**	0.0313	0.6840	0.3458	1.5962	**0.0117**	0.1448

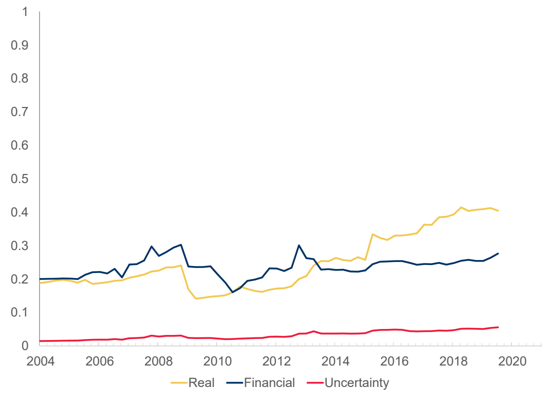
TABLE 2. Results when factors are used as predictors

3.3.1. *Posterior Probability of Inclusion Plots*

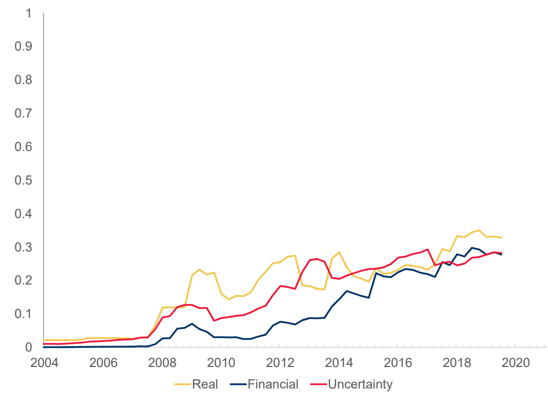
Regarding the posterior probability of inclusion plots (Figure 11), we see that extracting the common variance and reducing the dimensions changes some of the inclusion probabilities. We notice that for the 1 quarter ahead forecasts for Portugal, the uncertainty block factor becomes relatively more important towards the end of the sample, indicating that the common fluctuations of volatility indices have predictive power for house prices post 2012, albeit with a four quarter delay. Moreover, each factor has an increasing probability of being included in the data generating model, with a bump for each factor around the time of the financial crisis, across all forecast horizons. All factors tend to have roughly similar inclusion probabilities for Portugal, whereas in Spain we notice that the uncertainty block factor is less important for forecasting than the financial and real economy factors. The use of factors to reduce the dimension of the set of predictors is helpful as it illustrates clearly the usefulness of each set of factors for forecasting house prices.



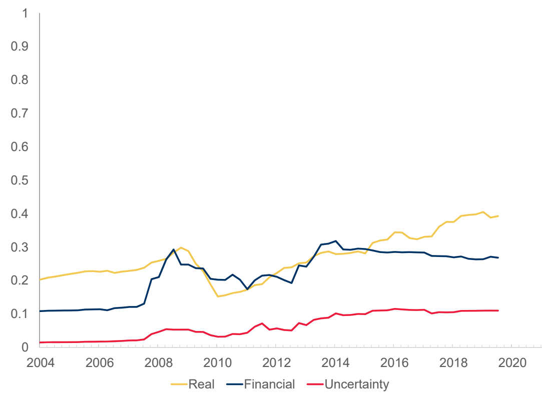
a) Portugal - one period ahead forecast horizon



b) Spain - one period ahead forecast horizon



c) Portugal - four periods ahead forecast horizon



d) Spain - four periods ahead forecast horizon

FIGURE 8: Posterior probability plots - Factors

In order to gain further insight we conduct a Quandt tests for each of the models in the factor model set and record the F test for a structural break. Although the limiting distribution of the Quandt test is not known precisely, we do see high values of the F statistics around 2013 for each of the models in the (factor predictor) model space. Although more research is needed in order to uncover what exactly is going on around 2013, we can say that there appears to be a structural break around that year.

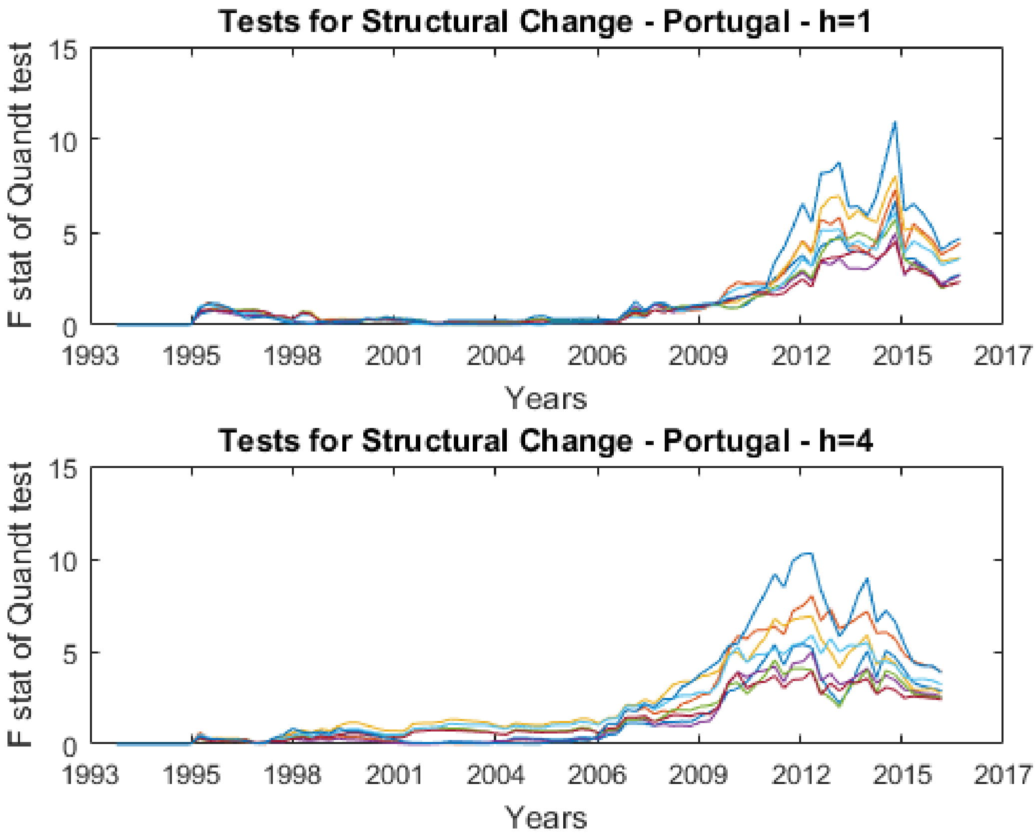


FIGURE 9: Portugal Quandt test - lines represent F statistics from a Quandt test at a given date for all 8 factor forecasting models. High values for the F statistic occur around the 2012 to 2014 period, suggesting a break date within that time frame.

4. Conclusion

Dynamic model averaging is a useful method for forecasting as it inherently permits uncertainty in both the combination of predictors as well as in the marginal effect of each predictor. Through the use of the two forgetting factors discussed above, DMA avoids converging onto a specific set of predictors and parameter estimates thereby allowing parameters and the predicting model to shift over time. These factors can be interpreted as mirroring the forecasters uncertainty towards estimated parameter and model distributions, with less 'forgetting' the forecaster is able to have more confidence in filtered parameter and model distributions. With more forgetting, the estimated distributions at each iteration of the filter are flattened reflecting the forecaster's uncertainty towards the estimates. This makes it a particularly useful approach for forecasting house prices for large out of sample periods, as we expect relevant predictors, and their marginal effects to change over time. In this paper, we applied DMA to forecast Portuguese and Spanish house prices, in doing so we are also able to track which predictors are relevant over the forecast period.

We experiment with different values for each forgetting factor and also apply a dynamic forgetting approach which attempts to minimize excess instability in estimating the coefficients of each model, while still permitting them to move quickly over time. We find that while there is no one-size-fits-all forgetting scheme, dynamic forgetting appears to offer lower forecast errors in most cases. We also carry out the analysis with block factors instead of a set of individual predictors. To acquire these factors, predictors were organized into real economy, financial, and volatility blocks. A specific factor was extracted from each block of predictors via the first principal component. We find that this dimension reduction technique provided gains in terms of forecast errors and should be considered in future forecasting exercises using DMA.

We find that different predictors have varying inclusion probabilities for both Portugal and Spain. The shifts in PIPs for Portugal (both sets) and Spain (one-quarter ahead) indicate some volatility in terms of model switching around 2013. Although by itself, this is not sufficient evidence for a regime change in terms of the drivers of house prices, nonetheless it does suggest that dynamics may have shifted around the beginning of the economic recovery. In Portugal, most predictors (including the economic uncertainty proxies) appear to have some value when it comes to forecasting changes in house prices. Furthermore, each predictor's importance appears to increase over time. For Spain, most real economy predictors appear to be useful for forecasting, and there appears to be less variation in each predictor's importance over time. Volatility measures appear to be more important in Portugal than in Spain for predicting house prices. This could be due to a number of reasons, for example it might be the case that Portuguese lenders and buyers may be more cautious during higher volatility than their counterparts in Spain or that the composition of buyers could also be different.

References

Bachmann, R., S. Elstener, and E. Sims (2013). "Uncertainty and Economic Activity: Evidence from Business Survey Data." *American Economic Journal: Macroeconomics*, 5(2), 217–249.

Baker, S., N. Bloom, and S. Davis (2015). "Measuring economic policy uncertainty." *Centre for Economic Performance*, p. CEP Discussion Paper No. 1379.

Bertola, Giuseppe, Luigi Guiso, and Luigi Pistaferri (2005). "Uncertainty and Consumer Durables Adjustment." *The Review of Economic Studies*, 72(4), 973–1007.

Bloom, N. (2013). "Fluctuations in uncertainty." *Centre for Economic Performance*, p. Occasional Paper 38.

Bork, Lasse and Stig V Møller (2015). "Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection." *International Journal of Forecasting*, 31(1), 63–78.

Carroll, Chris and Wendy Dunn (1997). "Unemployment Expectations, Jumping (S,s) Triggers, and Household Balance Sheets." *in NBER Macroeconomics Annual 1997*, 12, Bernanke and Rotemberg.

Case, Karl E., John M. Quigley, and Robert J. Shiller (2005). "Comparing Wealth Effects: the Stock Market versus the Housing Market." *Advances in Macroeconomics*, 5(1), Article 1.

Clark, Todd E and Kenneth D West (2007). "Approximately normal tests for equal predictive accuracy in nested models." *Journal of econometrics*, 138(1), 291–311.

Costa, S., L. Farinha, L. Martins, and R. Mesquita (2020). "Portuguese Household Finance and Consumption Survey: results for 2017 and comparison with the previous waves." *Bank of Portugal Economic Studies*, (Volume VI), 25–49.

Cuestas, Juan Carlos and Merike Kukk (2019). "The Spanish housing market: is it fundamentally broken?" *Applied Economics Letters*, pp. 1–5.

Dzielinski, Michal (2012). "Measuring economic uncertainty and its impact on the stock market." *Finance Research Letters*, 9(3), 167 – 175.

ECB (2020). "The Household Finance and Consumption Survey: Results from the 2017 wave." Statistics paper series no 36 / march 2020, ECB.

EFF (2019). "Encuesta financiera de las familias (EFF) (2017): métodos, resultados y cambios desde 2014." 4/2019 boletín económico, Banco de Espanha.

Englund, Peter, Min Hwang, and John M. Quigley (2002). "Hedging Housing Risk"." *Journal of Real Estate Finance and Economics*, 24, 167–200.

Helbling, Thomas and Marco Terrones (2003). "Real and Financial Effects of Bursting Asset Price Bubbles." *IMF World Economic Outlook*, Chapter 2, April.

Hill, Robert and Paulo M. M. Rodrigues (2020). "Flexible Dynamic Model Averaging Forecasts." Working paper, mimeo.

Knight, F.H. (1921). "Uncertainty and Profit." *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.*

Koop, Gary and Dimitris Korobilis (2011). "UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so?" *Economic Modelling*, 28(5), 2307–2318.

Koop, Gary and Dimitris Korobilis (2012). "Forecasting inflation using dynamic model averaging." *International Economic Review*, 53(3), 867–886.

Kozeniauskas, Nicholas, Anna Orlik, and Laura Veldkamp (2016). "The Common Origin of Uncertainty Shocks." NBER Working Papers 22384, National Bureau of Economic Research, Inc, URL `https://ideas.repec.org/p/nbr/nberwo/22384.html`.

Ling, David Ngo Chek, Joseph T. L. Ooi, and Thao T. Le (2015). "Explaining House Price Dynamics: Isolating the Role of Nonfundamentals." *Journal of Money, Credit and Banking*, 47, 87–125.

Lourenço, R.F. and P.M.M. Rodrigues (2014). "The Dynamics and Contrast of House Prices in Portugal and Spain." *Bank of Portugal Economic Bulletin*.

Lourenço, R.F. and P.M.M. Rodrigues (2015). "House prices: Bubbles, exuberance or something else? Evidence from Euro Area countries." Banco de Portugal, Working Paper 17.

Lourenço, R.F. and P.M.M. Rodrigues (2017). "House prices in Portugal - what happened since the crisis?" *Bank of Portugal Economic Studies*, (Volume III), 41–57.

Moretti, Laura, Luca Onorante, and Shayan Zakipour Saber (2019). "Phillips curves in the euro area." Tech. rep., ECB Working Paper Series No 2295 / July 2019.

Nicoletti, Giulio and Raffaele Passaro (2012). "Sometimes it helps: the evolving predictive power of spreads on GDP dynamics." Tech. rep., Working Paper Series 1447, European Central Bank.

Pástor, Lubos and Pietro Veronesi (2012). "Uncertainty about Government Policy and Stock Prices." *The Journal of Finance*, 67(4), 1219–1264.

Poterba, JM, DN Weil, and R Shiller (1991). "House price dynamics: the role of tax policy and demography." *Brookings papers on economic activity*, (2), 143–203.

Raftery, Adrian E, Miroslav Kárnỳ, and Pavel Ettler (2010). "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, 52(1), 52–66.

Risse, Marian and Martin Kern (2016). "Forecasting house-price growth in the Euro area with dynamic model averaging." *The North American Journal of Economics and Finance*, 38, 70–85.

Tsatsaronis, K. and H. Zhu (2004). "What drives Housing Price Dynamics: Cross Country Evidence." Tech. rep., BIS Quarterly Review, BIS.

Non-technical summary

October 2020

Inputs, technology and efficiency: The Portuguese economy in the last three decades

João Amador and António R. dos Santos

In this article we analyse the contribution of the accumulation of inputs and total factor productivity (TFP) do the economic growth of the Portuguese economy between 1990 and 2017. The contribution of TFP to growth is associated to the sum of technological progress and changes in efficiency. We compare results obtained for Portugal with those for the average of the EU28 and distinguish the block of countries that joined until the mid-nineties (EU15) from those that acceded later (EU13). In addition, we compare Portugal with five countries of similar size in terms of capital and labour (Austria, Belgium, Sweden, Czech Republic and Greece).

The methodological approach is based on some hypothesis. If we assume that every economy can have access to the world technology, which evolves over time for different capital-labour combinations, it is possible to estimate the maximum production that can be obtained out of different levels of inputs, given the existing technology, and hence compute the contribution of technological progress (shifts in the frontier) and efficiency developments (changes in the distance to the frontier). In conceptual terms, technological progress corresponds to more *productive* techniques, for example associated with innovations, which are not captured by the conventional methods of accounting the stock of inputs. In parallel, improvements in efficiency correspond to improved institutional and organizational arrangements, i.e., more *efficient* use of the current level of inputs and technology. Therefore, for given levels of capital and labour, an economy benefits from the world technological progress, though these gains may not entirely materialize due to efficiency developments.

The article considers five separate 10-year periods and results are presented in terms of average contributions to GDP growth. In order to increase the robustness of conclusions, periods with overlapping years are considered: [1990–1999], [1995–2004], [1999–2008], [2004–2013], [2008–2017].

As in all empirical work, the results are sensitive to the assumptions made and the data available. In this latter respect, the database used, although it may occasionally differ from national sources, offers long comparable time series for different countries. As for methodological hypotheses, the choice of translog production function and the assumption of a linear trend for technological progress are relevant points. As a robustness test, the frontier was estimated as a Cobb-Douglas production function and results remain qualitatively unchanged. Another relevant point is the quality of inputs, related with human capital levels and types of investment made. By not being explicitly incorporated into the methodology, these issues play an important role in comparing countries in terms of TFP contributions to economic growth.

Results suggest a modest performance of the Portuguese economy relatively to the EU28 average. In the last 15 years, with the exception of Greece in the last two periods considered, the overall performance of gross domestic product (GDP) in Portugal was lower than that of countries of similar size in terms of capital stock and number of workers. The contribution of total input accumulation in Portugal has decreased along the period considered, reflecting low levels of investment and adverse demographic developments. The relatively low capital-labour ratios and the high impact of capital accumulation on output highlight the importance of investment as a driver for Portuguese economic growth. Moreover, the change in TFP contributions in Portugal was qualitatively similar to that of the average of the EU28 but clearly lower in levels. This performance mostly resulted from efficiency developments, which had negative contributions in all decades. In this context, Portugal emerges as a country whose output is quite distant from the international technological frontier. More than in other segments of the technological frontier, some of Portugal's peer size countries have comparatively very favorable performances. In this context, Portugal appears with a low level of efficiency, which in the last two periods considered corresponds to the minimum value of the EU28 (Figure 1).

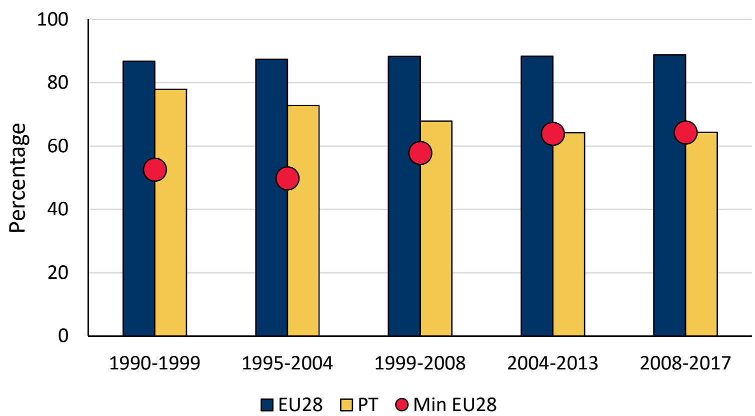


FIGURE 1: Efficiency levels in Portugal and in the EU28

Inputs, technology and efficiency: The Portuguese economy in the last three decades

Joao Amador
Banco de Portugal
Nova School of Business and Economics

António R. dos Santos
Banco de Portugal
Nova School of Business and Economics

October 2020

Abstract
This article estimates a common dynamic stochastic production frontier for the European Union countries taking several decades within the period 1990–2017. These frontiers are the starting point to analyse developments in the Portuguese economy, notably through a growth accounting exercise that disentangles the total contribution of inputs' accumulation and total factor productivity (TFP) to GDP growth. In addition, TFP contribution is broken down into technological progress and changes in efficiency. Moreover, the computation of the elasticities of capital and labour to GDP make it possible to disentangle total input's accumulation into the contributions of capital and labour. Results reflect a modest performance of the Portuguese economy along the last decades, particularly in terms of the contribution of efficiency developments. (JEL: C11, O47, O52)

1. Introduction

The expansion of total factor productivity (TFP) reflects the ability of an economy to grow over and above the accumulation of inputs like capital and labour and it is typically obtained as part of a growth accounting exercise. Therefore, the analysis of TFP developments is a relevant part of the debate on Portuguese and European economic growth. However, in order to better understand economic performance, GDP growth must be disentangled in such a way that TFP is not obtained as a simple residual, i.e., not just in terms of what is not explained by the accumulation of inputs. If we assume that every economy can have access to the world technology, which evolves over time for different capital-labour combinations, it is possible to estimate an international stochastic production frontier and decompose TFP as the contribution of technological progress (shifts in the frontier) and efficiency developments (changes in the distance to the frontier).

These two components represent different dimensions to be considered in TFP developments. In conceptual terms, technological progress corresponds to more *productive* techniques, for example associated with innovations, which are not

Acknowledgements: The authors are thankful to Nuno Alves, António Antunes, José Ramos Maria e Pedro Duarte Neves for usefull comments and suggestions. The analyses, opinions and conclusions expressed herein are the sole responsibility of the authors and do not necessarily reflect the opinions of Banco de Portugal or the Eurosystem.

E-mail: jamador@bportugal.pt; ammsantos@bportugal.pt

captured by the conventional methods of accounting the stock of inputs. In parallel, improvements in efficiency correspond to improved institutional and organizational arrangements, i.e., more *efficient* use of the current level of inputs and technology. Therefore, for given levels of capital and labour, an economy benefits from the world technological progress, though these gains may not entirely materialize due to efficiency developments. In practical terms, the best performers within the set of countries in the sample determine the international frontier, also meaning that technology can deteriorate if all countries perform worse for each combination of inputs. Although the direct causes for efficiency developments are not identified in this type of methodological approach, growth accounting exercises based on stochastic technological frontiers are a step forward in understanding the drivers of economic developments in different periods of time.

The seminal contribution in empirical growth literature is that of Solow (1957), which decomposes GDP growth along input's accumulation and TFP. The application of dynamic stochastic production frontiers to growth accounting, notably through Bayesian statistical methods, was suggested by Koop *et al.* (1999), using a set of developed economies. Our exercise closely follows this methodological approach and updates the work of Amador and Coimbra (2007), maintaining all its priors and assumptions, while using another database and a different set of countries. The time period covered is 1990–2017 and the set of countries corresponds to the EU28. This group of countries faces a similar set of institutional constraints, making it even more likely that they potentially have access to a common technology. Amador *et al.* (2019) present a short exercise using this methodology and database for the same set of countries but for a shorter period, 1995–2014.

The main question underlying this article concerns the relative performance of the Portuguese economy comparatively to the EU28 average along the last three decades with an emphasis on the role of inputs, technology and efficiency. The results suggest that the experience of EU28 countries is quite different, with Portugal showing a modest performance in relative terms. This is associated with decreasing contributions from capital and labour accumulation and basically no positive effects emerging from TFP developments. Most strikingly, efficiency developments have been systematically negative, hinting at the existence of structural problems that have not been solved.

The article is organized as follows. The next section briefly describes the methodology and the database used. Section 3 presents the contributions of the growth accounting components to Portuguese and EU28 GDP growth in the five 10-year periods considered. Moreover, we highlight the low efficiency levels in Portugal and their low contributions to GDP growth, also in comparison with similar EU28 countries in terms of capital stock and employment. Section 4 offers some concluding remarks.

2. Methodology and data

The estimation of stochastic production functions (or stochastic frontiers) is standard in the literature, though the utilization of Bayesian methods is somewhat less common. In this section we briefly overview the estimation methodology and present the database. As in all empirical work, results are dependent on the data available, which is a challenge when, as in our study, long comparable time series for different countries are required. The data for the most recent years are necessarily prone to revisions.

2.1. *The stochastic frontier approach*

The basic underlying hypothesis in the methodology is the existence of a common EU28 stochastic dynamic production frontier, which can be statistically identified because there are countries lying in its different segments. Conceptually, it means that, since all countries have access to the same technology, if two of them have equal labour and capital endowments the one with higher GDP is more efficient, i.e., it stands closer to the frontier.

The validity of the assumption on the existence of a EU28 production frontier is worthwhile discussing. Although it is accepted that knowledge about production techniques and about the relative value of goods and services is widely accessible across Member States, its dissemination may take a long time to materialize, for example due to specificities of domestic institutions or other barriers. For instance, heavy licensing procedures or other regulatory costs may deter the entry of firms with new technologies. In this vein, Basu and Weil (1998) discuss the speed of international dissemination of technological progress and its implications in terms of growth, arguing that it occurs at a slower pace than the diffusion of knowledge. In this context, the time that elapses until a country effectively adopts the technological innovations in the production systems is reflected in its relative productive efficiency. Overall, the dissemination of knowledge is faster within a group of countries that is homogeneous in terms of institutional setup and with geographical proximity, which supports the decision to estimate the EU28 stochastic production frontier.

Another issue is the assumption regarding the pace of technological progress. The assumption taken is that it evolves in a linear way. In connection with what was referred in the previous paragraph, this implicitly means that there is an average speed for the adoption of new technologies across countries and specific lags or leads are captured by the efficiency component. Koop *et al.* (1999) tested alternative formulations for the dynamics of the production function, namely a time specific model, where frontiers are totally independent in time, a quadratic trend model and a linear trend model under constant returns to scale. Each of these alternatives presents advantages and limitations. The time specific model is very flexible but implies the sampling of numerous parameters, which is computationally heavy. The linear and quadratic trend models are less demanding in terms of parameters but impose some rigidity in the dynamics of technical progress. The quadratic trend is more flexible than the linear one, which makes it preferable if long periods of time are analysed. In turn, the linear trend constrained to a constant returns technology imposes too much structure. Taking the set of alternatives, the linear trend model seems to offer the best compromise, with good results in terms of the in-sample fit and useful balance between flexibility and parsimony.

Our article considers five separate 10-year periods (9 annual rates) and results for the growth accounting exercises are presented in terms of average contributions to GDP growth. In order to increase the robustness of conclusions, we take a rolling window perspective with partial overlapping, considering periods [1990–1999], [1995–2004], [1999–2008], [2004–2013], [2008–2017]. It should be noted that the length of the periods considered is sufficient to average out short-run fluctuations in the macroeconomic variables.

Regarding the functional form of the production function, we use a translog specification. This formulation encompasses, as a special case, the logarithmic transformation of the Cobb-Douglas production function and it is more flexible than the latter. In fact, a major limitation of the logarithmic transformation of the Cobb-Douglas production function is the absence of interaction terms between labour and capital. Temple (2006) argues that the assumption of a Cobb-Douglas specification may lead to spurious results in economical and statistical terms. The problem is magnified because traditional growth accounting exercises treat TFP as unobservable (omitted variable). Conversely, if the researcher identifies a good proxy for TFP and the data are actually generated by a translog, a suitable specification accurately recovers the original parameters and rejects the Cobb-Douglas. Nevertheless, as a robustness test, we estimated the frontier as a Cobb-Douglas production function and results remain qualitatively unchanged.

Econometric principles allow for the estimation of stochastic production functions through maximum likelihood methods.[1] However, the Bayesian methods are suitable when samples are small because they allow for inferences without relying on asymptotic approximations. In addition, most importantly, Bayesian methods make it possible to rationally combine observed data with economically meaningful initial assumptions (priors). In practice, observed data is combined with priors to generate a posterior distribution function.

In our exercise, the prior for the efficiency parameter is an asymmetric positive distribution. The rationale behind this assumption is twofold. Firstly, this parameter measures the distance relatively to the production frontier so it should be positive. Secondly, there is a smaller probability of finding observations as we move further inside the production frontier. This assumption is common for the estimation of stochastic frontier functions but the specification of the asymmetric distribution remains an open question. We opted for a normal-gamma model (normal distribution of the residual component and gamma distribution for the efficiency component). Its relative advantages versus other alternatives, such as normal-half normal and normal-exponential models, are discussed in Greene (2000) and Tsionas (2000).

1. For references on non-bayesian estimation methods of stochastic production functions see, for example, Aigner *et al.* (1977), Meeusen and der Broeck (1977) and Kumbhakar and Lovell (2004).

2.2. *The model*

The model considered for the decomposition of the GDP growth closely follows Koop *et al.* (1999) and takes the form:

$$Y_{ti} = f_t\left(K_{ti}, L_{ti}\right)\tau_{ti}w_{ti}, \tag{1}$$

where Y_{ti}, K_{ti} and L_{ti} stand for the real output, the real capital stock and labour in period t $(t = 1, ..., T)$ in country i $(i = 1, ..., N)$, respectively. Furthermore, τ_{ti} $(0 < \tau_{ti} \leqslant 1)$ is the efficiency parameter and w_{ti} represents the measurement error in the identification and its stochastic nature. As mentioned above, the basic model assumes a flexible translog production function:

$$y_{ti} = x'_{ti}\beta_t + v_{ti} - u_{ti} \tag{2}$$

where:

$$x'_{ti} = \left(1, k_{ti}, l_{ti}, k_{ti}l_{ti}, k^2_{ti}, l^2_{ti}\right) \tag{3}$$

$$\beta_t = \left(\beta_{t1}, ..., \beta_{t6}\right)' \tag{4}$$

and lower case letters indicate natural logs of upper case letters. The logarithm of the measurement error v_{ti} is iid $N(0, \sigma^2_t)$ and the logarithm of the efficiency parameter is one sided to ensure that $\tau_{ti} = \exp\left(-u_{ti}\right)$ lies between zero and one. The prior for u_{ti} is taken to be a gamma function with a time specific mean λ_t.

The contribution of input endowment, technology change and efficiency change to GDP growth are defined in a simple way. The GDP growth rate in country i in period $t + 1$ is:

$$y_{t+1,i} - y_{t,i} = \left(x'_{t+1,i}\beta_{t+1} - x'_{t,i}\beta_t\right) + \left(u_{t,i} - u_{t+1,i}\right), \tag{5}$$

where the first term includes technical progress and factor accumulation and the second term represents efficiency change. In practice, shifts in the frontier correspond to changes in the betas between two moments in time, while changes in inputs correspond to changes in k and l, which are considered in vector x. Therefore, the first term can be further broken down as:

$$\frac{1}{2}\left(x_{t+1,i} + x_{ti}\right)'\left(\beta_{t+1} - \beta_t\right) + \frac{1}{2}\left(\beta_{t+1} + \beta_t\right)'\left(x_{t+1,i} - x_{ti}\right) \tag{6}$$

The technical change for a given level of inputs results from the first term of the previous equation and is defined as:

$$TC_{t+1,i} = \exp\left[\frac{1}{2}\left(x_{t+1,i} + x_{ti}\right)'\left(\beta_{t+1} - \beta_t\right)\right] \tag{7}$$

and the input change defined as the geometric average of two pure input change effects, relatively to the frontiers in consecutive years:

$$IC_{t+1,i} = \exp\left[\frac{1}{2}\left(\beta_{t+1} + \beta_t\right)'\left(x_{t+1,i} - x_{ti}\right)\right] \tag{8}$$

The efficiency developments (change in the distance to the frontier) are defined as:

$$EC_{t+1,i} = \exp(u_{ti} - u_{t+1,i}) = \frac{\tau_{t+1,i}}{\tau_{t,i}} \tag{9}$$

As previously mentioned, the structure of technological change is assumed to evolve in a linear way. Therefore:

$$\beta_t = \beta^* + t\beta^{**} \tag{10}$$

$$\sigma_t^2 = ... = \sigma_T^2 = \sigma^2 \tag{11}$$

Thus the model is written as:

$$y = X^* \times \beta - u + v \tag{12}$$

with

$$y = \left(y_1^{'}...y_T^{'}\right), u = \left(u_1^{'}...u_T^{'}\right), v = (v_1...v_T)', \beta = \left(\beta^{*'}\beta^{**'}\right)', \tag{13}$$

where β is a 12×1 vector and:

$$X^* = \begin{bmatrix} X_1 & X_1 \\ . & . \\ X_t & tX_t \\ . & . \\ X_T & TX_T \end{bmatrix} \tag{14}$$

where X_t is a 28 (countries) by 6 vector.

The sequential Gibbs sampling algorithm defined by equations A.2 to A.6 was run with 1,020,000 iterations for each of the five 10 year periods considered, with a burn-in of the first 20,000 iterations to eliminate possible start-up effects (see Casella and George 1992). In addition, the results for the contributions of technology (ATC), input (AIC) and efficiency (AEC) are presented in terms of geometric averages for each period. The details of the likelihood function and the formulas for capital and labour elasticities are presented in Appendix A.

An important methodological aspect is to verify that the algorithm converges to a stable distribution for each parameter, thus providing robust posterior estimates. In this context, we have computed the classic Geweke (1992) algorithm convergence criteria. Geweke's statistic is a convergence diagnostic for Markov chains based on a test for equality of the means of the initial and final parts of the chain, which has an asymptotically standard normal distribution. More specifically, if the two samples are drawn from the stationary distribution of the chain, the corresponding means should equalize. In our exercise, the Z scores for all parameters reject the probability of the difference between the means of the samples associated with the first and second half of the iterations to be different from zero.

2.3. *Database*

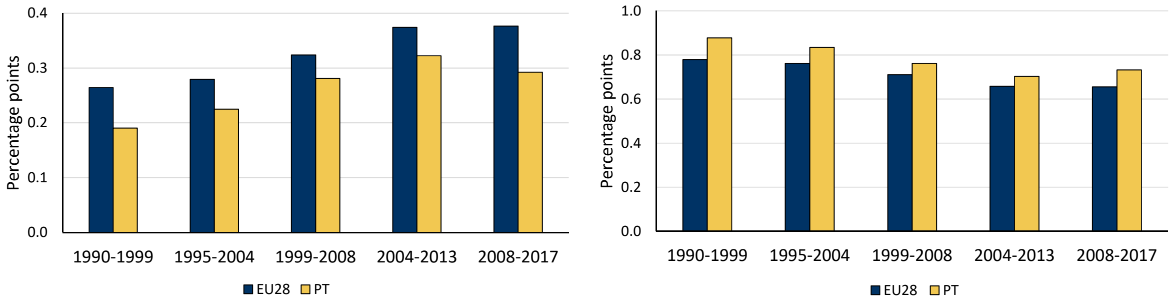
The data set used collect information for employment in number of persons, capital stock and GDP for the overall economy from 1990 until 2017 for the set of EU28 countries collected from the 9.1 version of the Penn World Table (Feenstra *et al.* 2015). Growth accounting exercises depend upon reliable data and, when the aim is to estimate a stochastic production frontier, this data have to be comparable across countries.

The Penn World Table has set a standard for high quality in historical cross-country economic aggregates, thus it is suitable to provide an accurate insight into the size and contributions to income differences in the EU28. The latest version of the database is more robust and has expanded the scope of information available relatively to the previous ones, notably in what concerns measures of physical capital. Nevertheless, GDP growth rates for Portugal in 2016 and 2017 in this version of the database are distant from those recorded in the national accounts data. Therefore, we replaced GDP growth rates for these years. As for all other years and for employment levels the database is very close to the national accounts, which is also the case for the remaining countries considered in the exercise.

3. Results

The basic results of the sequential Gibbs sampler are the posterior means and medians for the set of 12 parameters in the production function. These sets of parameters in the five periods considered can be used to compute the elasticity of capital and labour for Portugal and for the remaining EU28 countries (Figure 1). The values obtained for the elasticities of capital and labour, used to breakdown the contribution of inputs to GDP growth, are different from those usually considered in classic growth accounting exercises, with the elasticity of capital being higher in our exercise. Some facts concur to explain these differences. The translog specification is more flexible than the Cobb-Douglas as elasticities depend on the specific levels of capital and labour, thus deviating from the fixed share of inputs on GDP. The estimation of the EU28 production frontier with a Cobb-Douglas production function conveys capital elasticities close to those of labour, in a background where labour shares have been lower in recent decades. Furthermore, the estimated production functions exhibit mild increasing returns to scale. The sum of the elasticities of capital and labour is approximately 1.03 for all decades considered.

Labour elasticities increased in Portugal and in the EU28 until the decade 2004–2013, while capital elasticities decreased up to this period. In the decade 2008–2017 the Portuguese economy was characterized by capital elasticities slightly higher than those of the EU28 average (0.73 and 0.66, respectively), meaning that, in the segment of the international production function where Portugal stands, further capital accumulation has an impact on GDP levels that is not distant from that of the EU28 but smaller than in the past. In practice, this result highlights the importance of investment as a driver for Portuguese economic growth.



(A) Labour (B) Capital

FIGURE 1: Estimated labour and capital elasticities for Portugal and EU28 average

In this context, it should be added that capital-labour ratios in the Portuguese economy have been relatively low in the context of the EU15 and slightly above those prevailing in the EU28 average. Nevertheless, the Portuguese capital-labour ratio became closer to the EU15 average in the decade 2008–2017, also via the reduction of the denominator, driven by the strong job destruction that took place during the 2011–2014 economic and financial adjustment program. Amongst other drivers, the relatively low capital-labour ratios cannot be dissociated from the reduced qualifications of the Portuguese labour force. These features put the Portuguese economy in a less favourable segment of the international production function, posting lower GDP per worker and expanding relatively less in a context of capital biased technological progress.

Tables 1 and 2 report the results of the detailed growth accounting decomposition for Portugal and for the average of the EU28, respectively. The Portuguese economy presented a modest performance in the decade 2008–2017 with an average GDP growth of -0.01 per cent. The average posterior Bayesian estimate is very close to this number (0.03 per cent). Economic growth in this period was affected by the 2008 global economic and financial crisis and by the following euro area sovereign debt crisis. The subsequent recovery starting in 2014 puts Portugal's GDP in 2017 approximately at the same level reported in 2008. The sharp correction in the macroeconomic imbalances prevailing in the Portuguese economy, associated with the sudden stop in external financing, had a negative impact on investment and led to a strong destruction of jobs. Therefore, the contribution of total input accumulation was small, -0.13 percentage points (p.p.), with capital representing 0.59 p.p. and labour -0.72 p.p. These contributions were the lowest of the five 10-year periods considered.

It is also worth noting that the contribution of capital and labour to Portuguese economic growth has decreased along the five 10-year periods considered. Although the contribution of labour to inputs' accumulation is affected by cyclical developments, the long-term trend reflects adverse demographic developments associated with the ageing of the population. This is one of the important challenges affecting the Portuguese economy going forward, as well as several other European countries. As for capital accumulation in Portugal, the progressive lower contribution to GDP growth results from the combination of lower elasticities and subdued investment associated with the high debt levels that have been prevailing in the economy. The drop in investment

Decades ending	Observed GDP	Expected GDP	Input			Total Factor Productivity		
			Total	Capital	Labour	Total	Technology	Efficiency
1999	2.75	3.20	3.21	2.98	0.23	-0.01	0.46	-0.47
2004	2.61	3.26	3.01	2.75	0.26	0.25	1.54	-1.29
2008	1.36	1.73	2.03	1.94	0.09	-0.30	1.30	-1.60
2013	-0.36	-0.34	0.32	0.75	-0.43	-0.67	-0.52	-0.14
2017	-0.01	0.03	-0.13	0.59	-0.72	0.16	0.26	-0.10

TABLE 1. Growth accounting results for **Portugal**

Note: Observed and expected GDP are presented as percentage average decade growth rates, while inputs and total factor productivity are presented as percentage points (geometric) average decade contributions. Expected GDP and contributions from inputs and total factor productivity result from the bayesian estimation.

Decades ending	Observed GDP	Expected GDP	Input			Total Factor Productivity		
			Total	Capital	Labour	Total	Technology	Efficiency
1999	1.33	1.57	1.24	1.58	-0.35	0.33	0.71	-0.38
2004	3.50	3.58	1.81	1.86	-0.05	1.77	1.58	0.19
2008	3.65	3.76	2.14	1.87	0.28	1.62	1.34	0.28
2013	1.41	1.31	1.40	1.37	0.03	-0.09	-0.24	0.15
2017	1.07	1.25	0.86	0.90	-0.05	0.40	0.60	-0.21

TABLE 2. Growth accounting results for the **European Union 28**

Note: Observed and expected GDP are presented as percentage average decade growth rates, while inputs and total factor productivity are presented as percentage points (geometric) average decade contributions. Expected GDP and contributions from inputs and total factor productivity result from the bayesian estimation.

was particularly strong during the economic and financial assistance program, reaching levels lower than those of the depreciations and thus leading to reductions in the level of the capital stock. The rebound of investment is another key challenge in the Portuguese economy, a task made harder by the negative prospects related with the prevailing indebtedness and the expected negative impacts of the COVID-19 pandemic.

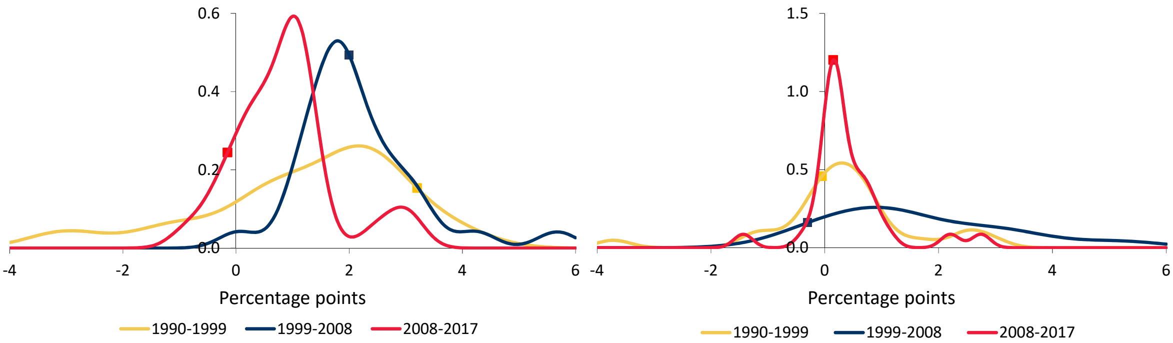
Developments in the TFP contribution to Portuguese GDP growth were different in the five 10-year periods considered, also in terms of contributions from technological progress and efficiency developments. This information is contained in the last three columns of Table 1. TFP contributions have been systematically low in Portugal, reaching values of -0.30 and -0.67 p.p in the decades ending in 2013 and 2017, respectively. In addition, contributions in Portugal were always lower than those observed in the average of the EU28, where the TFP was only contributed negatively in the decade 2004–2013 (-0.09 p.p.). Therefore, the increase in TFP, at least in terms of keeping up with the EU28 developments, constitutes another important challenge for the Portuguese economy in the next years.[2]

The comparison of the Portuguese growth accounting results with those for the average of the EU28 are very useful for benchmarking purposes. In this latter region, the decade 2008–2017 also featured the lowest average GDP growth rate of the five 10-year periods considered, with a weakening of inputs accumulation but a small improvement in TFP contribution relatively to the decade ending in 2013. Nevertheless, the EU28 average combines realities from two distinct sets of countries. Although with some differences, the countries whose accession took place until 1995 (EU15) started with an underlying situation that was quite different from that of the countries that joined later, in the aftermath of the fall in the Berlin wall (EU13). These differences in performance are visible in Tables B.1 and B.2 presented in Appendix B.

The EU13 countries experienced a negative performance in the decade 1990–1999, associated with the transition from central planning to market economies. In this period there was a strongly negative contribution from labour accumulation, due to job destruction and emigration, coupled with sizeable efficiency losses. However, in the other decades the EU13 countries performed much better than the EU15 group. Almost all the new EU member states have been catching up strongly and converging towards the EU15. The largest contributions for this good performance are attributed to TFP developments, which reached 2.79 and 2.86 p.p. in the decades ending in 2004 and 2008, respectively. These results highlight both the existence of cross-country differences but also the key role of TFP for economic growth.

Figure 2 illustrates cross-country differences in the EU28 by plotting the kernel distributions of total inputs and TFP contributions to GDP growth rates in the decades 1990–1999, 1999–2008 and 2008–2017, with the position of the Portuguese economy signalled by coloured squares. The distribution of input's contributions (panel A) presents a heavier left tail in the decade ending in 1999, due to what occurred in the Central and Eastern European countries. In the remaining decades the distribution is closer to a Gaussian curve and it has shifted left in the most recent period. The Portuguese position deteriorates, starting at the right tail of the 1990-1999 distribution and finishing at the left tail of the most recent distribution. Furthermore, the distribution of TFP contributions (panel B) reflects different underlying structural conditions in terms of quality of inputs and functioning of markets and institutions. In the decade 2008–2017 most of the density in the distribution lies between zero and 1 p.p., with Portugal posting a marginally positive figure (0.16 p.p.). In fact, Portugal reports contributions close to zero in all decades.

2. The results are in line with those in the Special Issue "Real convergence in the European Union and the relative performance of the Portuguese economy" (Banco de Portugal 2019).



(A) Total input accumulation (B) Total factor productivity

FIGURE 2: Distribution of growth accounting contributions in the EU28

Note: The squares indicate Portugal's position in the respective distribution. Kernel density estimation is a non-parametric way of estimating the probability density function of a variable.

As previously mentioned, one of the strengths of the methodology is the possibility of breaking down TFP contributions along technological progress and efficiency developments, which are conditional on the position of the estimated international stochastic frontier. In this respect, what strikingly stands out in the Portuguese case is the systematic negative contribution from efficiency developments to average GDP growth in the five 10-year periods considered. These contributions reached -1.29 and -1.60 p.p. in the decades ending in 2004 and 2008, respectively. Moreover, these contributions were mostly compensated by the positive contribution of technological progress, which is associated with the performance of best performing countries with similar levels of inputs. Next, in the decade that ends in 2013, as previously mentioned, since most countries were negatively affected by the global economic and financial crisis, the contribution of technological progress is negative, which adds up to a smaller contribution from efficiency developments (-0.14 p.p.). Finally, in the decade 2008–2017 the contribution of technology returned to positive territory but that was not the case for efficiency developments.

The change in TFP contributions in the average of the EU28 followed a pattern similar to that of Portugal, but the levels of contributions were clearly higher. Again, most of the difference is attributed to efficiency developments. The analysis of the subgroups of countries in the EU15 and EU13 provides further detail. Comparing the last column of Tables B.1 and B.2 in Appendix B shows that, except in the decade 1990–1999, efficiency developments in the group of the most recent EU member countries were better than in the EU15, where they were marginally negative.

A complementary view consists in computing the level of efficiency of the economy. This is defined as the percentage of output actually produced relatively to the output level at the frontier, measured at the exact combination of employment and capital stock levels existing in the country. Figure 3 plots the efficiency levels in Portugal and in the average of the EU28 in the five 10-year periods considered. Efficiency levels in the average of the EU28 were relatively constant in all periods around 88 percent. In contrast, Portuguese efficiency levels stood close to 80 percent in the decade 1990–1999 and dropped to 64 in the period 2008–2017, which is a very low number. In fact, the striking result is that Portugal stands as the country with the lowest efficiency level in the set of the EU28 countries in the two final 10-year periods considered. This is in accordance with the efficiency developments reported in the growth accounting exercise and highlights the existence of ample room for improvement in the utilization and allocation of resources available in the Portuguese economy.[3]

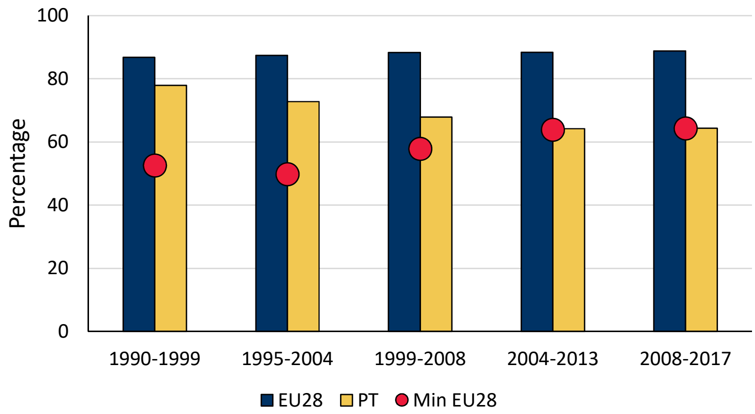
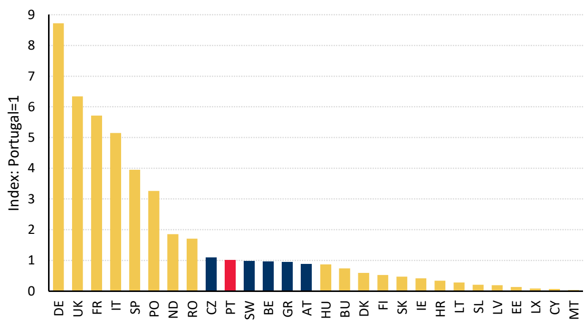
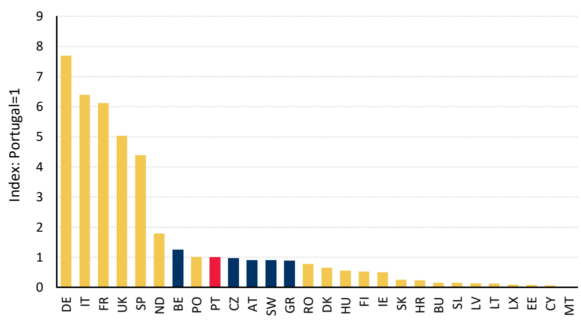


FIGURE 3: Efficiency levels in Portugal and in the EU28

In order to further document the unfavourable efficiency developments in the Portuguese economy we chose a set of five EU28 countries with a similar size in terms of employment and capital stock. Figure 4 motivates the choice of Austria, Belgium, Sweden, the Czech Republic and Greece as "peer size" countries. This group contains countries that joined the EU at different moments in time and covers different geographical areas. As a side result, Figure 4 also highlights the gap between the six largest EU28 economies and the remaining members. By choosing countries with similar levels of labour and capital we focus the analysis on a comparable segment of the EU28 production frontier.



(A) Labour



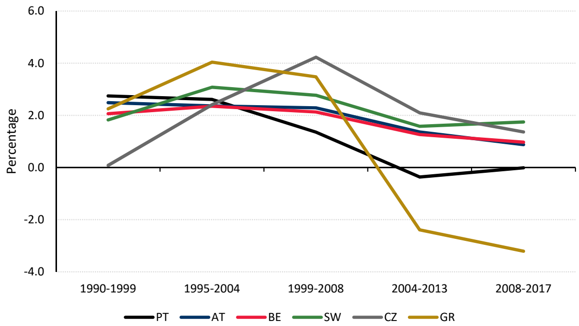
(B) Capital

FIGURE 4: Selected peer countries basing on total employment and capital stock
Note: Average values for the decade 2008–2017.

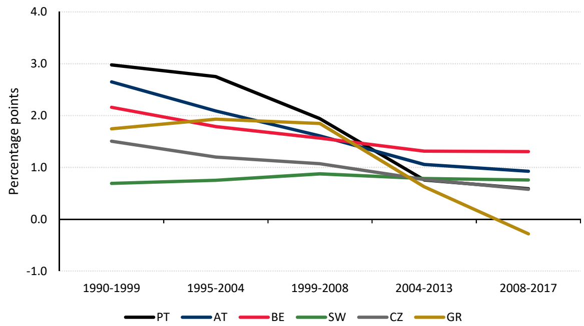
3. A robustness check using the Cobb-Douglas production function leads to the same relative efficiency levels for Portugal.

There are relevant results emerging from the comparison across peer size countries in panels of Figure 5. In the last three 10-year periods the overall GDP growth performance in Portugal was worse than in peer size countries, except for Greece in the last two periods (panel A). Conversely, the Czech Republic has shown a good GDP performance. In addition, in contrast with other peer countries, there is a sharp reduction of the contribution of capital accumulation to GDP growth in Portugal across decades (panel B). This development is paired with the negative contributions from labour in the last two decades, a feature also visible in Greece and in contrast with positive contributions in other peer countries (panel C). As for the contribution of total factor productivity in the different periods (panel D), differences are visible across countries. The Czech Republic has recorded good contributions along the time, except for the initial decade, while Portugal has posted low contributions, which were only more unfavourable for the case of Greece in the last two 10-year periods. Next, as for the contribution of technological progress for GDP growth, it is quite similar in all peer countries (panel E). This is exactly the expected result because peer countries were selected for being placed in the same segment of the stochastic production frontier. Therefore, by construction, shifts in this function affect them all in the same way. Finally, as regards the contribution from efficiency developments to GDP growth (panel F), as previously highlighted, we identify a modest performance for Portugal, translated into negative contributions, accompanied by Greece in the last two 10-year periods. These developments contrast with very small positive contributions in other peer countries, and especially with the remarkable progress observed in the Czech Republic.

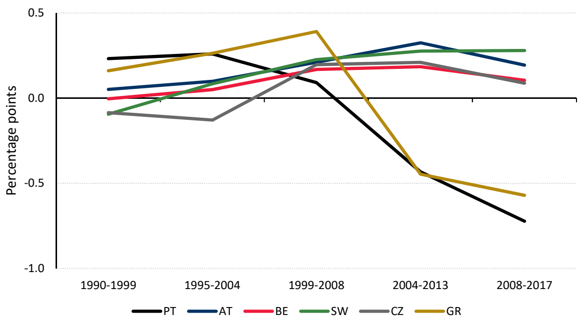
Overall, these comparisons show that countries with similar levels of inputs, which also share the EU institutional setup, may present quite different outcomes. This is related with structural conditions that go beyond the EU setup and stresses the role for good national policies and for the sharing of best practices, notably in the context of benchmarking exercises which are regularly carried out by international organizations. Moreover, the quality of inputs, related with human capital levels and types of investment made, certainly play a major role in explaining different efficiency performances.



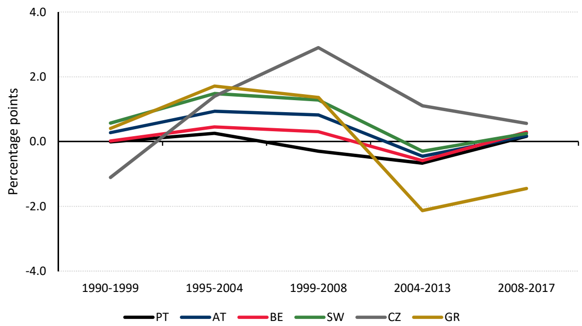
(A) GDP growth



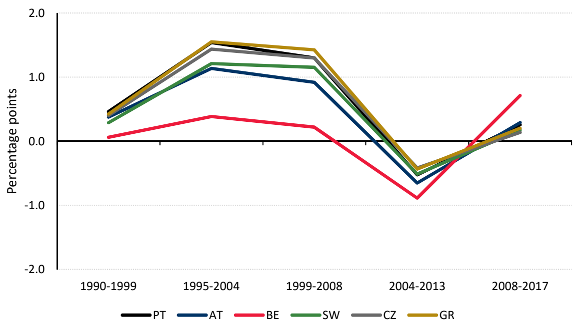
(B) Capital



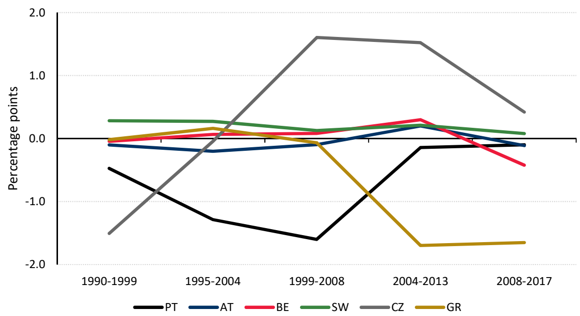
(C) Labour



(D) TFP



(E) Technology



(F) Efficiency

FIGURE 5: Growth decomposition: Portugal and peer countries regarding size

4. Final remarks

In this article we conduct a growth accounting exercise based on the estimation of a dynamic stochastic production frontier for the EU28 countries between 1990 and 2017. Growth accounting exercises are mechanical by nature but offer a useful assessment of economic performance, especially if other countries are explicitly taken as benchmarks. This comparison is possible to achieve with the stochastic production frontier approach, notably in terms of detailing TFP developments.

The results suggest a modest performance of the Portuguese economy relatively to the EU28 average. The contribution of total input accumulation in Portugal has decreased during the period considered, reflecting adverse demographic developments and low levels of investment. The relatively low capital-labour ratios and high capital elasticities highlight the importance of investment as a driver for Portuguese economic growth. Moreover, the change in TFP contributions in Portugal was qualitatively similar to that of the average of the EU28, but contributions to GDP growth were clearly lower in levels. This performance mostly resulted from efficiency developments, which had negative contributions in all decades. Moreover, efficiency levels in the Portuguese economy were lower than those for the average of the EU28 in the five 10-year periods considered, with Portugal standing as the country with the lowest efficiency level in the two final 10-year periods.

It must always be born in mind that results are sensitive to hypothesis taken and statistical data. In this latter respect, the international data for the capital stock trends are affected by different accounting measures and deflation procedures. International databases like the Penn World Tables try to offer harmonized series, though they may sometimes deviate from national sources. The alternative of fully replicating the exercise with official data for all EU28 countries is not viable due to numerous series breaks and limited time horizon. As for methodological hypothesis, it is important to underline that, although the translog production function offers substantial flexibility, this choice and the assumption of a linear trend for technological progress in each decade affects the results.

As for policy prescriptions, it is hard to go beyond the long standing references to the need to intensify the accumulation of capital and to allocate it properly, as well as to maintain progress in terms of human capital. In this article we highlight the fact that there is large room for efficiency improvements, which may materialize through the removal of unjustified regulatory barriers and the improvement of inputs' quality. The extension of the exercise that we carried out in order to explicitly consider the quality of inputs in the estimation of the technological frontier is a promising avenue for future research.

References

Aigner, Dennis, C Lovell, and Peter Schmidt (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics*, (6), 21–37.

Amador, João and Carlos Coimbra (2007). "Characteristics of the Portuguese Economic Growth: What has been Missing?" Tech. rep.

Amador, João, Carlos Coimbra, and António R. dos Santos (2019). "How have technological progress and efficiency developments contributed to Portuguese growth?" In *Portuguese Economic Growth: A View on Structural Features, Blockages and Reforms*, edited by Economics and Reasearch Department. Banco de Portugal.

Banco de Portugal (2019). "Real convergence in the European Union and the relative performance of the Portuguese economy." *Special Issue, Economic Bulletin, Banco de Portugal*, October.

Basu, W and D Weil (1998). "Appropriate Technology and Growth." *Quarterly Journal of Economics*, 113, 1025–1054.

Casella, George and Edward George (1992). "Explaining the Gibbs Sampler." *The American Statistician*, (46), 167–174.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer (2015). "The Next Generation of the Penn World Table." *American Economic Review*, 105(10), 3150–3182.

Geweke, J. (1992). "Evaluating the Accuracy of Sampling Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, edited by J. M. Bernardo et al. Oxford Clarendon Press.

Greene, W (2000). "Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function." *Quarterly Journal of Economics*, 113, 1025–1054.

Koop, Gary, Jacek Osiewalski, and Mark Steel (1999). "The Components of Output Growth: A Stochastic Frontier Analysis." *Oxford Bulletin of Economics and Statistics*, 61(4), 455–487.

Kumbhakar, Subal C. and C Lovell (2004). *Stochastic Frontier Analysis*. Cambridge University Press.

Meeusen, Wim and Van der Broeck (1977). "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review*, (18), 435–452.

Solow, Robert (1957). "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 39, 312–320.

Temple, Jonathan (2006). "Aggregate production functions and growth economics." *International Review of Applied Economics*, 20(3), 301–317.

Tsionas, E (2000). "Full likelihood inference in Normal-gamma stochastic frontier models." *Journal of Productivity Analysis*, 13, 183–205.

Appendix A: The likelihood function

The full likelihood function of the model presented in subsection 2.2 can be written as:

$$f_N^{TN}\left(y\left|X^*\beta-u,\sigma^2 I_{TN}\right.\right)p\left(\beta\right)p\left(\sigma^{-2}\right)p\left(\lambda^{-1}\right)\prod_{t=1}^{T}\prod_{i=1}^{N}f_G\left(u_{ti}\left|1,\lambda^{-1}\right.\right), \tag{A.1}$$

where f_N^{TN} stands for a multivariate $T\times N$ normal probability distribution function, f_G stands for a gamma probability distribution function and:

$$\begin{aligned} p\left(\lambda^{-1}\right) &= f_G\left(\lambda^{-1}\left|1,-\ln\left(\tau^*\right)\right.\right) \\ p\left(\sigma^{-2}\right) &= \sigma^2\exp-\frac{10^{-6}}{2\sigma^2} \end{aligned}$$

The prior for λ^{-1} assumes a gamma distribution with the first parameter equal to 1 and second parameter equal to $-ln(\tau^*)^{-1}$ such that τ^* is the prior median efficiency. Typically τ^* is chosen based on a priori expectations for the median of the efficient distribution. However, in a very heterogeneous sample of countries, the existence of large deviations from the frontier increases the sum of errors and places the randomized algorithm that generates a sequence of posteriors - the sequential Gibbs sampler - in an unstable path. For the algorithm to accommodate such a sample, this has to be compensated by a low τ^*. We assume a starting point for τ^* near zero and check the posterior median efficiencies. As for σ^{-2}, we assume the usual flat prior.

Given this prior structure, the posterior marginal distributions that compose the Gibbs sampler can be easily derived. The conditional for β is:

$$p\left(\beta\left|Data,u,\sigma^{-2},\lambda^{-1}\right.\right)\sim f_N^{2J}\left(\beta\left|\widehat{\beta},\ \sigma^2\left(X^{*\prime}X^*\right)^{-1}\right.\right), \tag{A.2}$$

where

$$\widehat{\beta}=\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\left(y+u\right) \tag{A.3}$$

The conditional for σ^{-2} to be used in the Gibbs sampler is:

$$p\left(\sigma^{-2}\left|Data,\beta,u,\lambda^{-1}\right.\right)\sim f_G \left(\sigma^{-2}\left|\frac{n_0+TN}{2},\frac{1}{2}\left[a_0+\left(y-X^*\beta+u\right)'\left(y-X^*\beta+u\right)\right]\right.\right) \tag{A.4}$$

Next, the conditional for u is a left truncated normal at zero:

$$p\left(u\left|Data,\beta,\sigma^{-2},\lambda^{-1}\right.\right)\sim f_N^{TN}\left(u\left|X^*\beta-y-\frac{\sigma^2}{\lambda}\iota,\sigma^2 I_{NT}\right.\right)\prod_{t=1}^{T}\prod_{i=1}^{N}I(u_{it}\geqslant 0), \tag{A.5}$$

whose mean is forced to be higher or equal to zero in the algorithm and ι is a $TN\times 1$ vector of ones. Finally, the marginal posterior distribution for the λ^{-1} is:

$$p\left(\lambda^{-1}\left|Data,\beta,u,\sigma^{-2}\right.\right)=f_G\left(\lambda^{-1}\left|1+TN,-\ln\left(\tau^*\right)+\sum_{t=1}^{T}\sum_{i=1}^{N}u_{it}\right.\right) \tag{A.6}$$

A final important element in the methodology is verification of regularity constraints regarding the elasticities of capital (EK_{ti}) and labour (EL_{ti}). Given the matricial formulation, these generic elements are:

$$EK_{ti} = (\beta_2^* + t\beta_8^{**}) + (\beta_4^* + t\beta_{10}^{**})l_{ti} + 2(\beta_5^* + t\beta_{11}^{**})k_{ti} \tag{A.7}$$

$$EL_{ti} = (\beta_3^* + t\beta_9^{**}) + (\beta_4^* + t\beta_{10}^{**})k_{ti} + 2(\beta_6^{**} + t\beta_{12}^{**})l_{ti} \tag{A.8}$$

Therefore, we only accept a set of posterior β parameters that translate into non-negative elasticities for all countries and periods.

Appendix B: Additional growth accounting results - UE15 and UE13

Decades ending	Observed GDP	Expected GDP	Input			Total Factor Productivity		
			Total	Capital	Labour	Total	Technology	Efficiency
1999	2.67	2.73	2.30	2.22	0.08	0.43	0.55	-0.12
2004	3.13	3.25	2.37	2.14	0.23	0.87	0.94	-0.06
2008	2.53	2.78	2.23	1.91	0.32	0.54	0.63	-0.09
2013	0.73	0.66	1.28	1.21	0.07	-0.62	-0.58	-0.04
2017	0.80	1.02	0.92	0.93	-0.01	0.10	0.37	-0.28

TABLE B.1. Growth accounting results for the **European Union 15**

Note: Observed and expected GDP are presented as percentage average decade growth rates, while inputs and total factor productivity are presented as percentage points (geometric) average decade contributions. Expected GDP and contributions from inputs and total factor productivity result from the bayesian estimation. The EU15 countries are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden and United Kingdom.

Decades ending	Observed GDP	Expected GDP	Input			Total Factor Productivity		
			Total	Capital	Labour	Total	Technology	Efficiency
1999	-0.22	0.23	0.02	0.85	-0.83	0.21	0.89	-0.68
2004	3.91	3.95	1.16	1.54	-0.38	2.79	2.32	0.47
2008	4.93	4.90	2.04	1.82	0.22	2.86	2.15	0.71
2013	2.19	2.07	1.55	1.56	-0.01	0.52	0.16	0.36
2017	1.38	1.52	0.78	0.87	-0.10	0.75	0.87	-0.12

TABLE B.2. Growth accounting results for the **European Union 13**

Note: Observed and expected GDP are presented as percentage average decade growth rates, while inputs and total factor productivity are presented as percentage points (geometric) average decade contributions. Expected GDP and contributions from inputs and total factor productivity result from the bayesian estimation. The EU13 countries are Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia.

www.bportugal.pt