1

Banco de Portugal Economic Studies

Volume V

BANCO DE PORTUGAL
EUROSYSTEM

1

Banco de Portugal Economic Studies

Volume V

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, 2019 • www.bportugal.pt

Banco de Portugal Economic Studies | Volume V – no. 1 | Lisbon 2019 • Banco de Portugal Av. Almirante Reis, 71 | 1150-012 Lisboa • www.bportugal.pt • Edition Economics and Research Department • Design Communication and Museum Department | Design Unit • ISSN (*online*) 2183-5217

Content

Editorial

Editorial

January 2019

The first issue of Banco de Portugal Economic Studies for 2019 contains three diverse essays dealing with bank credit concentration, the sensitivity of current account balances to the business cycle, and education mismatches in the labour market. The issue also presents an Economics synopsis on the inflation goals of monetary policy.

The first essay, by António R. dos Santos and Nuno Silva bears the title "Sectoral concentration risk in Portuguese banks' loan exposures to non-financial firms". After the great recession and the sovereigns' crisis in Europe, the concerns of policy makers and the public with the robustness of banks have increased markedly. One area that has received much deserved attention is the prevalence of non-performing loans and the threats they pose to the stability of the financial system. A fundamental regulatory approach to the credit risk problem is the Basel capital framework, whereby banks are required to add equity as their exposure to risks increases. The problem the authors address is that, in this approach, the capital cushion required for a specific loan only depends on the loan's own risk, regardless of the composition of the overall credit portfolio of the bank. In other words, in the currently used methodologies the risk assigned to a bank's loans does not take into account the gains in risk reduction that may come through increasing the diversification of the loans portfolio.

In order to demonstrate the quantitative importance of diversification the authors build and estimate two contrasting models for credit loss distributions. The first model has two components generating the returns of assets: an idiosyncratic risk component and a common factor, where this last component captures sensitivity to aggregate macroeconomic conditions. The second model is similar but instead of a factor common to all loans, the returns for each asset depend on an industry specific risk component. These risk factors are less than perfectly correlated across industries, which means that the aggregate measure of risk of a diversified loan portfolio will always be smaller than the risk measured ignoring diversification.

The empirical work used data on bank credit to non-financial corporations operating in Portugal between 2006 and 2017 with each firm assigned into one of thirteen industry groups. The models were estimated using one-year probabilities of default available from Banco de Portugal in-house credit assessment, individual credit exposures and sector expected default frequencies estimated from the national credit register (CRC).

Simulations compared risk measures based on the portfolio loss distributions using the two models. The results indicate that in the last years the difference in the risk measures generated by the common factor

model and the one with industry specific factors increased. In the pre-crisis period, common factor models ignoring diversification had measures of risk 40% higher than the models with gains from diversification. The differences increased since 2014 to approximately 50%. What drove this change? The results point to diversification gains in the last years thanks to a lower concentration of credit in a specific sector: construction. This is relevant because the analysis shows that the lower risks were due to this diversification and not due to a widespread allocation into sectors with lower interdependency.

The results in this paper seem to justify the introduction of improvements in financial system risk assessments, considering explicitly loan portfolio diversification.

The second essay by João Amador and João Falcão Silva is titled "Cyclically-Adjusted Current Account Balances in Portugal". In the same way that making a distinction between structural and actual government budget balances has become more relevant in recent years, we have also witnessed a growing interest in filtering out the fundamental or trend components of current account balances from the influence of business cycle fluctuations. In the Portuguese case, the current account balance evolved from a deficit of about 10 per cent of GDP in 2010 to a surplus of 0.5 per cent in 2017. Did this change result from a structural adjustment or does it mostly reflect cyclical developments? The authors deal with this question focusing on the sensitivity of imports and exports to the GDP growth of Portugal and of the export destination countries.

The basic idea that imports depend on GDP is refined to consider a model taking into account the different import intensities of domestic expenditure (C+G+I) and of exports. The relationship between imports and aggregate demand is assumed to be characterized by stable long run elasticities. This assumption allows the authors to produce estimates of what the levels of imports and exports would be if both Portugal and its trading partners had GDPs at their trend value, that is, with zero output gaps.

The empirical analysis used quarterly data from OECD for the volumes and deflators of GDP and its components from the last quarter in 1995 until the last quarter in 2017. Domestic and foreign output gaps estimates came from the IMF World Economic Outlook. Trade elasticities were estimated using information from the OECD Inter-Country Input-Output database for 2016.

The results for exports show that the gap between observed and structural exports have been relatively small, never exceeding 2 percentage points in absolute terms. Similarly, the changes in imports of goods and services as a percentage of GDP from 1996 to 2008 were largely of a structural nature. After this period, imports were systematically below the corresponding structural level, meaning that the negative output gaps prevalent in these years brought down imports significantly. The strongest cyclical adjustment of imports represented 3.4 p.p. of GDP in 2012. Putting together the results for imports

and exports (and not introducing any cyclical adjustments to the balances of primary income and secondary income) we obtain the cyclically adjusted current account balance as a percentage of GDP for the Portuguese economy. The adjusted balances have fluctuated around the non-adjusted balance over the years. Since 2012 the adjusted current account balance has been lower but the gap between adjusted and non-adjusted balances has progressively diminished to near zero in 2017. The conclusion to extract from the analysis is that, as in other countries, the adjustment of the Portuguese current account balance to the economic cycle is not very large. This means that most of the improvement observed in the Portuguese current account balance in recent years has a structural nature.

The third essay, by Ana Catarina Pimenta and Manuel Coutinho Pereira has the title "Aggregate educational mismatches in the Portuguese labour market". Portugal has a labour force with lower levels of education than in most European countries. The data shows that while the schooling levels of younger cohorts has been rising, there has also been a shift towards occupations requiring more skills as economies modernize and the weight of technology intensive industries grows. Is the prevalence of undereducation in the labour force a growing or a vanishing problem? For some, the problem is actually the opposite. As the fraction of the labour force with higher education grows, does it outpace the employment needs for highly educated workers? Are we having a problem of overeducation, where many people with higher education can only find jobs that do not make use of those investments in human capital? Regardless of their type, educational mismatches have a negative impact on firms' productivity and on job satisfaction.

The paper provides answers based on the microdata in "Quadros de Pessoal". The analysis covers more than 23 million observations for the years between 1995 and 2013. The analysis starts from the individual information in "Quadros de Pessoal" using the Portuguese classifications of occupations and the levels of completed education. The authors establish correspondences between these classifications and the International Standard Classification of Occupations (ISCO) on one hand and the International Standard Classification of Education (ISCED) on the other. A standard correspondence between occupations and required levels of education, based on work by the International labour Organization, is then used to identify situations with under or overeducation. The results of this methodology show a consistent reduction over time of undereducation in Portugal, from 64.6 per cent of the employees in 1995 down to 35 per cent in 2013. This reflects the replacement of older by younger and more educated generations. In contrast, overeducation is much less significant with a prevalence of 0.9 percent in 1995, nevertheless growing up to 5.1 percent in 2013.

The authors also use another methodology based on the modes of the workers level of education distributions in each occupational group. This mode indicator is driven only by the Portuguese labour force data and thus

reflects the gaps relative to international standards by using a lower level of required education for some occupations. According to this methodology, in 1995 the prevalence of undereducation was close to 12 percent while in 2013 it was above 20%. One final section of the paper concentrates on comparisons with other 25 European countries, using harmonized microdata from the Survey of Income and Living Conditions (EU-SILC), covering a wide range of countries on an annual basis and applying methodologies similar to those described earlier to data from 2007 and 2016. Portugal was the country with the highest incidence of undereducation in both years considered, despite a drop from 2007 to 2016. As for the prevalence of overeducation, Portugal's rates are well below the EU average.

Assuming that the results using international standards are the most relevant, one "take away" from the paper is that in Portugal the passage of time has been reducing the size of the undereducation problem, a problem that is still significant today as the international comparisons reveal.

The Economics synopsis included in this issue is authored by Bernardino Adão and titled "Why is price stability a key goal of central banks?". It is part of the modern central banks' goals to maintain a low and stable inflation rate, which is typically defined as 2% in the medium run. Does the literature provide a justification for adopting this goal? The paper surveys the literature and answers affirmatively: a low and stable level of inflation is not only efficient but also equitable.

The starting points of the analysis are the realizations that the opportunity cost of holding money is given by the nominal interest rate and that inflation is essentially a tax on money holdings. An insight, due to Milton Friedman, is that since the cost of producing money is basically zero, then efficiency requires that the cost of holding money should be zero as well. This implies that the optimal rate of inflation should be negative and with the same absolute value as the real interest rate.

The idea that the socially optimal cost of holding money is zero turned out to be robust to early optimal taxation arguments in a second best world. Even if other taxes are distortionary, it is still not efficient to tax money because efficiency requires the taxation of final goods only (a result by Nobel laureates Peter Diamond and James Mirrlees) and money is best seen as an intermediate good in a "technology" that produces transactions. However, some results in the literature show that the optimality of the zero inflation does not survive if some relevant imperfections of the tax systems are taken into account.

If tax authorities are not able to tax pure monopoly profits, if tax evasion is significant or if the tax authorities cannot tax transfers from government, then the optimal rate of inflation is not zero. However, efforts to quantify the optimal inflation rate in these cases produced very low estimates, providing evidence that the optimum cost for money holdings is not too far from zero. Other pieces of research took into account that collecting traditional taxes is costly but that the inflation tax has essentially no collection costs for the

government. In this case, the optimal inflation is not zero but quantification efforts resulted again in very low optimal inflation rates.

A different strand of the literature considers the consequences of price rigidities. Based on empirical studies, this literature assumes that prices fail to adjust timely to markets' conditions, leading to price dispersion not based on fundamentals and consequently to misallocation of resources. Price rigidities can also interact with problems in measuring the price level, more specifically with problems associated with improvements in the quality of goods. To mitigate the price rigidity problem but also to take into account the welfare of agents holding money the optimal inflation in these models is a compromise, which is a value between the minus of the real interest rate and zero, when there is no need for price changes.

Another problem is downward nominal rigidities, being the nominal wage the most important case. When the nominal wage is downward sticky, stable prices prevent adjustments to negative shocks potentially leading to excess unemployment and making positive levels of inflation desirable. However, attempts to quantify optimal inflation rates yet again reveal the optimal level to be very low and close to zero. A different situation turned out to be relevant in the last few years: the restriction that interest rates cannot be (much) lower than zero became active. This makes it more difficult to conduct stabilization policy and points to the desirability of positive inflation. Again, results trying to quantify optimal inflation rates taking these situations into account point to the optimality of positive but very low inflation rates.

A different concern is the redistributive effect of inflation. The income elasticity of the demand for money is less than one, which means that poorer households hold a larger fraction of their income in money. Since high income households are better at avoiding the inflation tax than those with low incomes it follows that inflation is a regressive tax.

One thing is the optimal level of inflation in the medium run, another its stability. A stable (and predictable) rate of inflation is good because it facilitates the use of prices in making decisions by all agents in the economy. Stable inflation improves welfare by eliminating a source of uncertainty. Indexation only offers a partial resolution to the problems caused by surprises since there is not perfect observability of inflation. Data on current inflation is not available in real time. Also there is heterogeneity in the types of indices that would best suit different economic agents. Additionally, contracts contingent on inflation would have higher transaction costs. Unexpected inflation also interacts with the progressive taxation of households' income, and increases the cost of capital to firms by raising capital gains taxation and artificially reducing tax allowances for capital depreciation.

A final interesting point in the survey deals with a situation that clearly is both under-researched and extremely relevant. Inflation rates can differ across regions in a monetary union, for example because of specific shocks. In the presence of frictions that make price changes more difficult, like downward

rigidities, adjustments within a monetary union are easier if the central bank has a higher target for inflation. That would avoid having regions with deflation when that is not optimal. Considering this problem, what might be the optimal inflation rate? The literature does not yet seem to have answers but hopefully that gap will soon be covered by future research.

Sectoral concentration risk in Portuguese banks' loan exposures to non-financial firms

António R. dos Santos
Banco de Portugal

Nuno Silva
Banco de Portugal

January 2019

Abstract
This article proposes a credit risk model to the Portuguese banks' aggregate loan portfolio of non-financial corporations (NFC). Using a one-period simulation-based multi-factor model, we estimate the loss distribution and several one-year risk metrics between 2006 and 2017. The model differentiates from the Basel IRB framework by explicitly incorporating interdependencies between economic sectors. The flexible nature of the model allows sectoral risk to be decomposed into different components. The results point to diversification gains in the last years thanks to a lower concentration in a specific sector, the construction sector, and not due to an allocation into sectors with lower interdependency. (JEL: G17, G21, G32)

Introduction

Concentration risk in a credit portfolio can arise from large exposures to specific borrowers relative to the size of the portfolio (*name concentration*) or from large exposures to groups of highly correlated borrowers. When two or more borrowers default simultaneously, the portfolio losses are more severe. The higher the correlation of defaults, the greater is the concentration risk. Default correlation can have several sources. Some of the most commonly mentioned are macroeconomic factors, geographic factors, corporate interrelations – arising either from common shareholders or supply chain relations – and economic sectors. The last decades were marked by several episodes where sector concentration played an important role. The concentration of bank credit in the energy sector in Texas and Oklahoma in the 1980s and the overexposure to the construction and property development sectors in Sweden in the early 1990s and in Spain and Ireland in the 2000s are examples of incidents of correlated defaults that jeopardized the health of many financial institutions.

Acknowledgements: I would like to thank António Antunes, Nuno Alves, Luísa Farinha, Diana Bonfim and Nuno Lourenço for their comments. The opinions expressed in this article are those of the authors and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the authors.

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

Since the implementation of Basel II, under Pillar 1 of bank capital regulation, banks can opt to either use a regulatory standardized approach to calculate credit risk capital requirements, or follow an Internal Ratings-Based (IRB) approach using their own estimated risk parameters. Either of these approaches aims at capturing general credit risk. However, they do not explicitly differentiate between portfolios with different degrees of diversification. Among other things, Pillar 2 in Basel II and in Basel III addresses this issue by providing a general framework for dealing with concentration risk. Nevertheless, banks and regulators have a large degree of freedom in choosing the quantitative tools to cover such risk (Grippa and Gornicka 2016).

The IRB formula is based on the Asymptotic Single Risk Factor (ASRF) model derived from the Vasicek (2002) model. The origins of this model can be found in the seminal work by Merton (1974). The ASRF model is based on two crucial assumptions, namely the existence of a single risk factor and portfolio granularity. Together, these two assumptions lead to portfolio invariance, i.e. the capital required for a loan only depends on its risk, regardless of the composition of the portfolio it is added to. From a regulatory perspective, this simplifies the supervisionary process allowing for the framework to be applicable to a wider range of countries and institutions. In the ASRF model, two borrowers are correlated with each other because they are both exposed to a unique systematic factor but with (potentially) varying degrees. In the specific case of the IRB approach the degree of exposure to the systematic factor is a decreasing function of the probability of default.[1] According to BIS (2005), this decreasing function is in line with the findings of several supervisory studies. Still, this can be a simplified way of capturing the interdependencies between the various debtholders in a portfolio where several other systematic risk factors might drive default events (Das *et al.* 2007; Saldías 2013). Keeping everything else equal, the IRB approach leads to the same capital charges for banks with different levels of sectoral concentration.

In this paper we implement a simulation-based multi-factor method to estimate the loss distribution for the aggregate loan portfolio of non-financial firms of Portuguese banks and derive several one-year credit risk metrics. This method differs from the IRB approach in two aspects: (i) instead of a single systematic risk factor we consider one risk factor for each sector, mirroring the asset returns correlations between sectors; (ii) instead of using a decreasing function of the probability of default, we explicitly estimate the degree of exposure to each sector-related systematic factor. Thus, the risk of default is not synchronized across sectors and the degree of exposure to the shocks varies according to the sector. The flexible nature of simulation-based

1. I.e., the IRB approach is a specific ASRF model where the implied correlation between borrowers are a function of their own risk.

methods allow us to evaluate the evolution of concentration over time and to decompose the credit risk into different components. This can help micro and macro-prudential authorities to detect sectoral risks in individual banks and in the banking system.

Methodology

The general framework relies on a structural multi-factor risk model evolved from the seminal work by Merton (1974). In this model set-up, a default is triggered when a firm's assets value is less than debt value. This implies that a default occurs when a firm standardized asset return, X_i, is below the threshold implied by the probability of default (PD) for that firm:

$$X_i \leq \Phi^{-1}(\mathrm{PD}_i), \tag{1}$$

where Φ^{-1} denotes the inverse cumulative distribution function for a standard normal random variable.

Adding on Merton's model further consider that the standardized asset return X of a firm i belonging to sector s is a linear function of an industry specific risk factor, Y_s, and an idiosyncratic risk factor, ε_i:

$$X_{si} = r_s Y_s + \sqrt{1 - r_s^2}\varepsilon_i, \tag{2}$$

$$\varepsilon_i \sim N(0,1) \quad Y_s \sim N(0,1).$$

In the above equation $r_s \in [0,1]$ is the factor weight (or factor loading), which measures the sensitivity of the asset returns to the risk factor. The standardized asset return X_i is a function of an idiosyncratic component – the risk that is endemic to a particular firm – and a sector-specific systematic component. Dependencies between borrowers arise from their affiliation with the sector and from the correlation between Y_s.[2] The risk factors dependencies are usually estimated using market sector indices. Those indices are not available for Portugal. Therefore, we use observed default frequencies to compute, under the Merton model assumptions, the implicit normalized asset returns and estimate correlations between sectors – Table B.2 in the Appendix B.[3]

A critical parameter in this exercise is the factor loading r_s. Small changes in this parameter can produce significantly different results. In Düllmann

2. For further details see Appendix A.

3. This procedure guarantees monthly frequency data offering greater consistency. Data is available between 2005m1 and 2017m12.

and Masschelein (2006) and Accornero *et al.* (2017) the factor weight is set exogenously as equal to 0.5. This value is chosen such that their benchmark portfolio capital charge equals the IRB capital charge. In the IRB approach the implied factor weight is a decreasing function of the PD and is bounded between, approximately, 0.35 (highest possible PD) and 0.49 (lowest possible PD). The objective of this study is not to evaluate the size of the Basel capital requirements but instead to recognize the likelihood of joint-defaults and how costly they are for a portfolio. Factor loadings should thus reflect by how much an extra euro borrowed by a firm i that belongs to a sector s is affected by the business cycle. We estimate the parameter endogenously with a year fixed effect regression for each sector using the implied threshold (also referred to as distance-to-default, $\mathrm{DD} = -\Phi^{-1}(\mathrm{PD}_i)$) as the dependent variable, weighted by the outstanding amount. Our goal is to capture by how much the variability of the distance-to-default is explained by time for each euro invested in sector s. The results are available in Table B.1 in the Appendix B.

The Loss distribution, L, for a given portfolio is then estimated through Monte Carlo simulations of the systematic industry specific and idiosyncratic risk factors. In each simulation/scenario, defaults are identified by comparing simulated standardized asset returns with the default threshold $\Phi^{-1}(PD_i)$:

$$L = \sum_{s=1}^{S} \sum_{i=1}^{I_s} D_{X_i \leq \Phi^{-1}(\mathrm{PD}_i)} \cdot \mathrm{EXP}_i \cdot \mathrm{LGD}_i, \tag{3}$$

where $D = 1$, when a company defaults, EXP_i is the exposure to the company i, LGD_i is the loss given default of exposure i, S is the number of sectors and I_s is the number of firms in sector s. For a given year t, the exposure of company i is the one observed in the last month of year $t-1$ and the LGD is assumed to be constant and equal to 0.5.[4] Each Monte Carlo simulation can be seen as a scenario or state of the world. Each scenario generates a particular loss for the portfolio. The frequency of various outcomes/losses after a large number of simulations generates the credit loss distribution. Figure 1 illustrates the process.

There are several risk measures that can be computed based on the portfolio loss distribution. The most commonly referred are the expected loss (EL), the value-at-risk (VaR), the unexpected loss (UL) and the expected shortfall (ES). The EL corresponds to the expected value of the portfolio loss L, which can be estimated as the mean of the simulated loss scenarios.[5] The VaR_p is the maximum possible loss if we exclude worse outcomes whose

4. In BIS (2001) the LGD is considered to be 0.5 for subordinated claims on corporates without specifically recognized collateral.

5. The EL can also be estimated as PD*LGD*EXP. The EL estimation does not depend on the model used.

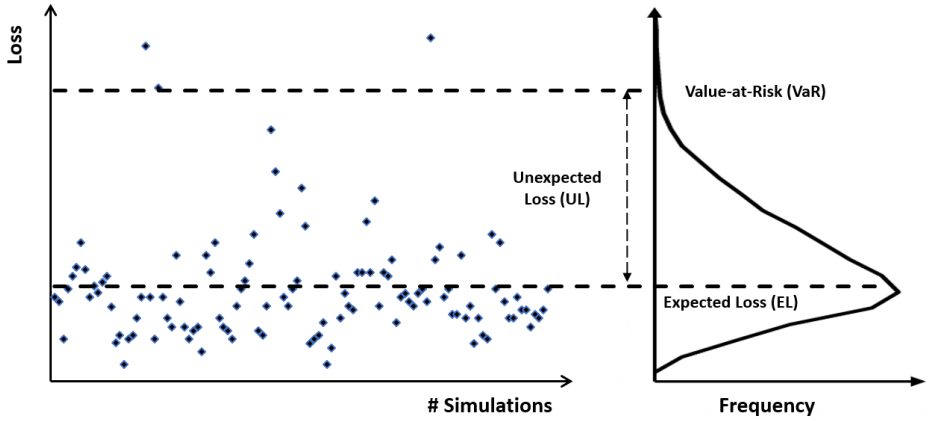


FIGURE 1: Credit Loss Distribution.

probability is less than p. The VaR is a quantile of the distribution. The UL_p is the difference between the VaR_p and the EL. In the IRB approach, it is considered that banks should have enough capital to sustain a loss with probability less than $p = 99.9\%$. The UL can thus be interpreted as the required capital to sustain such losses. In turn, the ES measures the expected loss beyond a specified quantile, the expected loss on the portfolio in the worst $p\%$ of cases. The ES is not considered under the IRB approach. However, it can be intuitively interpreted as the amount of capital required on average to sustain losses with probability above p. From now on we will consider $p = 99.9\%$, the value used in the IRB model.

The ES can be decomposed in marginal contributions of each economic sector s. According to Puzanova and Düllmann (2013) marginal contribution measures have a desirable full allocation property, i.e. they sum up to the overall ES. The marginal contribution is interpreted as the share of ES attributable to a sector, an approximation of its systematic relevance. It combines the assessment of sector risk, its weight in terms of credit exposure and its interdependency with other sectors:

$$\mathrm{MC}_s = E[L_s | L_{tot} \geq \mathrm{VaR}_q(L_{tot})]. \tag{4}$$

Data

This article uses a unique dataset with series for non-financial corporations operating in Portugal between 2006 and 2017. This dataset includes: individual credit exposures and observed sectoral default frequencies captured from the national credit register (CRC); NACE[6] groups available from IES (*Informação Empresarial Simplificada*), and one-year probabilities of default available from Banco de Portugal in-house credit assessment – SIAC (*Sistema Interno de Avaliação de Crédito*).[7]

The initial sample covers roughly the population of non-financial firms that have at least one loan granted by a resident financial institution. Nevertheless, only firms whose loans are considered to be performing are included in the analysis because only those are in risk of default in the next year. Thus, when a firm defaults at year t it is excluded from the analysis at $t + 1$ and for as long as the firm is considered as in default.[8] Therefore, we analyze approximately 77% of firms – 85% of total exposure.

The economic groups are divided based on the aggregate levels of NACE into thirteen sectors. Ideally, firms in a given group should be as homogeneous as possible in the variability of PD over time, but heterogeneous between groups. In other words, they should react in a similar way to the same factors. One possibility to increase group homogeneity would be to further divide the groups using lower levels of NACE. However, when using lower levels of NACE we could not guarantee a reasonable number of observations in each group to consistently estimate the model parameters. Thus, each firm was assigned to one of the thirteen industry groups. Figure 2 shows that more than half of the credit exposure of performing loans is concentrated in four sectors: wholesale and retail trade, manufacturing, construction and real estate activities. While the first two sectors maintained a relatively constant weight between 2006 and 2017, the aggregate exposure to the other two declined from 40% to 25% of the total portfolio. This decrease in weight was roughly equally offset by the remaining sectors, although more prominently in the transporting and storage and accommodation and food service activities.

6. Statistical classification of economic activities in the European Community.

7. See Antunes *et al.* (2016).

8. A firm is considered to be "in default" towards the financial system if it has 2.5 per cent or more of its total outstanding loans overdue. The "default event" occurs when the firm completes its third consecutive month in default. A firm is said to have defaulted in a given year if a default event occurred during that year.

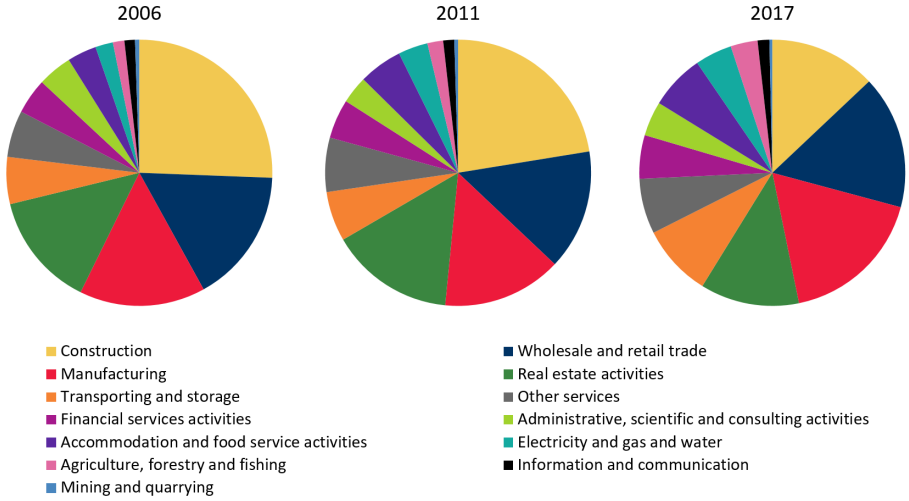


FIGURE 2: Portuguese credit portfolio of performing loans to non-financial firms – weights by activity sector.

Results

Figure 3 reports the loss distribution for the aggregate loan portfolio of non-financial firms of Portuguese banks between 2006 and 2017, presented as a percentage of the total exposure.[9] The distribution is not symmetric, being more concentrated in small losses and with a reduced frequency of large losses. The distribution is limited to the left since its best scenario is when there are no losses. It has a heavy tail and so losses can be quite extensive. Using the information from the loss distribution estimated for each year, Figure 4 shows the expected loss and the three tail credit risk measures – value-at-risk, unexpected loss and expected shortfall – at 99.9% between 2006 and 2017. In order to allow comparisons between different years, all credit risk measures are presented as a percentage of the total exposure. All measures display a similar pattern: a continuous increase between 2006 and 2013, followed by a decline until 2017. $\mathrm{VaR}_{99.9\%}$ and $\mathrm{ES}_{99.9\%}$ move in a parallel way because loss distributions are strictly monotonically decreasing in the tail. During this period the EL ranged from 1.6% to 5.3%, while the $\mathrm{UL}_{99.9\%}$ ranged from 5% to 8.8%. In 2017, the EL was approximately at levels of 2009/2010, while the UL was close to the minimum value reported in 2006. In fact, the difference between EL and UL has decreased over time. This issue will be addressed later on.

9. See dynamic graph on the PDF file.

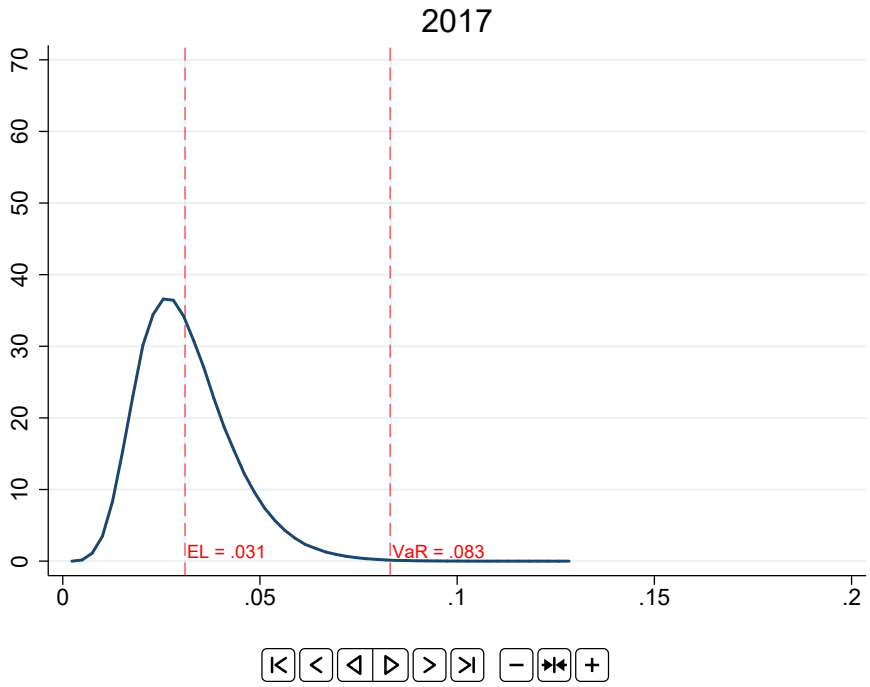


FIGURE 3: Portfolio Loss Distribution 2006-2017.

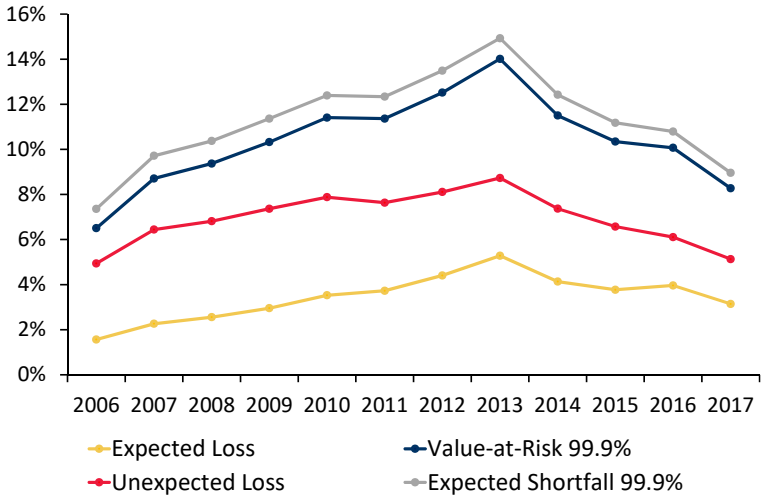


FIGURE 4: Credit risk measures based on Loss Distribution for the Portuguese loan portfolio.

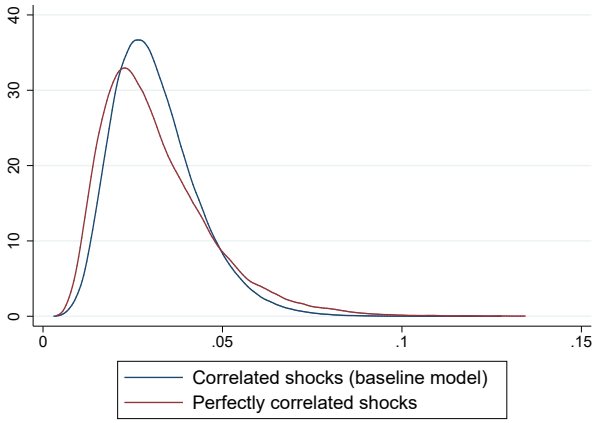
The measures presented so far are useful to assess the credit risk in a loan portfolio but they fail to quantify the role of sector concentration for portfolio credit risk. As such, we will rely on two different exercises that try to establish meaningful measures for the evolution of concentration risk. The first compares the results of our general framework (baseline model) with an ASRF model, while the second decomposes the unexpected loss. The values that are going to be presented should be interpreted with caution since they are sensible to the interdependency structure considered and to the factor weight r_s.

For the first exercise, Figure 5 (A) reports the portfolio loss distribution for 2017 under two different assumptions for the industry specific risk factor Y_s in equation (3). The model with correlated shocks (baseline model, in blue) refers to the loss distribution generated using the correlation structure presented in Table B.1 in the Appendix B, the same distribution as in Figure 3. Whereas the model with perfectly correlated shocks ignores diversification issues and can be treated as an ASRF model. The distribution in this second case (in red) is slightly to the left but it has also a heavier tail. This result is somehow expected since positive (negative) scenarios will now materialize simultaneously for all sectors. By construction the distribution in red produces higher (or equal[10]) values for the $\mathrm{VaR}_{99.9\%}$. In 2017, the unexpected loss is approximately 54% higher under this hypothesis (8.0% instead of 5.2%). In other words, if default risk was perfectly synchronized across sectors the UL for the Portuguese loan portfolio in 2017 would be 54% higher *vis-à-vis* a scenario where default risk is only partially synchronized. By repeating this exercise for all periods, the results indicate that in the last years the difference in the unexpected loss between the baseline model and the one with perfectly correlated shocks increased – Figure 5 (B). In the pre-crisis period the difference was around 40% and has increased since 2014 to approximately 50%, suggesting that the portfolio has become more diversified. But what drove this change?

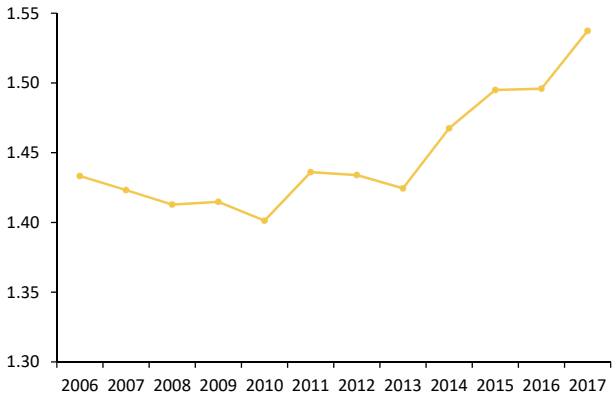
To try to answer the question we will perform a second exercise. Again, let us consider the industry specific risk factor, Y_s, in equation (3) and define three different auxiliary models: (i) a model with only idiosyncratic shocks, where all firms are independent and so each one suffers from a specific shock Y_i; (ii) a model that imposes only correlation within-sector by simulating a different Y_s for each sector s but assumes that all Y_s are independent; (iii) our baseline model that imposes both intra and inter-sector correlations. By construction each model has the same expected value but produces higher (or equal) values for the $\mathrm{VaR}_{99.9\%}$ and $\mathrm{UL}_{99.9\%}$:

$$\mathrm{UL}^{(i)}_{99.9\%} \leq \mathrm{UL}^{(ii)}_{99.9\%} \leq \mathrm{UL}^{(iii)}_{99.9\%}\,. \tag{5}$$

10. The portfolio exposure is concentrated in only one sector or in perfectly correlated sectors.



(A) Portfolio Loss Distribution 2017.



(B) Unexpected Loss – ratio between the perfectly correlated shocks model and the baseline model.

FIGURE 5: Model under the hypothesis of perfectly correlated shocks *vis-à-vis* the baseline model.

Figure 6 decomposes the UL between 2006 and 2017 based on its risk drivers, notably, an independent firm contribution, a contribution arising from within-sector correlation and a contribution arising from between-sector correlation. This is done using the three models before mentioned. From the figure, it is possible to see that, despite slightly increasing, the independent firm contribution plays a very minor role. Most of the unexpected loss is justified by within and between sector correlations. The relative contribution from each of these sources of correlations to UL has however changed during the last years. While in the pre-crisis period, the within-sector correlation explained most of the UL, this role is now played by the between-sector correlation. An interesting additional metric to understand this dynamic is

the ratio between unexpected and expected loss (UL / EL). Figure 7 shows this ratio and decomposes it into the same contributes as Figure 6. Based on Figure 7 it is possible to see that the referred ratio decreased steadily from 2006 until 2015 and remained constant afterwards. This ratio is especially affected by interdependency in borrowers' defaults. The between-sector contribution to the ratio remains fairly constant over time while the within-sector contribution dictates the ratio's trend. The results indicate that the possible diversification gains in the last years are caused by a lower concentration in specific sector(s) and not due to an allocation into sectors with lower dependency *vis-à-vis* other sectors. Otherwise the between-sector contribution would have decrease. This trend is also found in the Herfindahl Index that measures the size of activity sectors in relation to the overall portfolio (normalized to 2006). So which sector or sectors are driving this result?

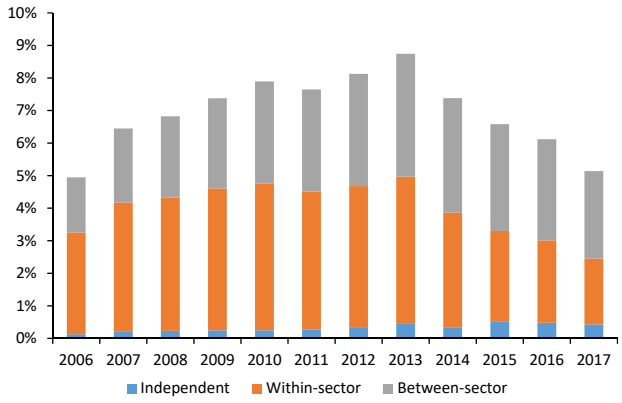


FIGURE 6: Contributions for the Unexpected Loss.

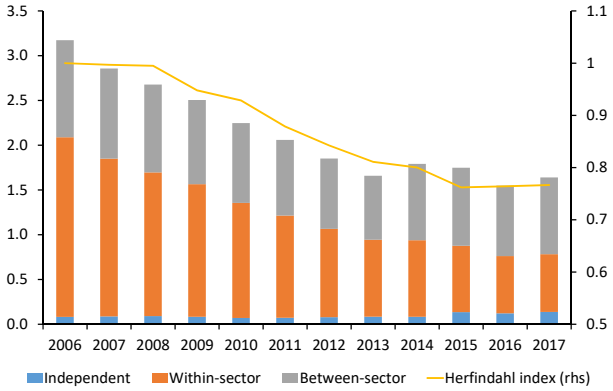


FIGURE 7: Contributions for the ratio UL/EL and Herfindahl Index (normalized to 2006).

Figure 8 reports the contributions of each sector to the expected shortfall for the baseline model in three different periods. Tail risk is significantly concentrated in two sectors, namely construction and real estate activities, which account for more than half of the ES. Still, while the contribution of the real estate sector remains fairly constant, the contribution of the construction sector decreases from approximately 55% to 30% between 2006 and 2017. Thus, the diversification gains documented before are apparently a result driven by the construction sector. Its marginal contribution for the tail risk is decreasing over time, mainly because its weight in the overall portfolio is also decreasing. This decrease results, *inter alia*, from the very significant number of defaults observed in this sector. Moreover, in Figure 9 we observe that the construction sector has, on average, the highest contribution for the EL but an even higher contribution for the ES. In contrast, sectors such as manufacturing and wholesale and retail trade, have a low contribution to the ES (approximately 13%) when compared with their importance to the EL (approximately 24%).[11] This difference suggests the existence of potential diversification gains.

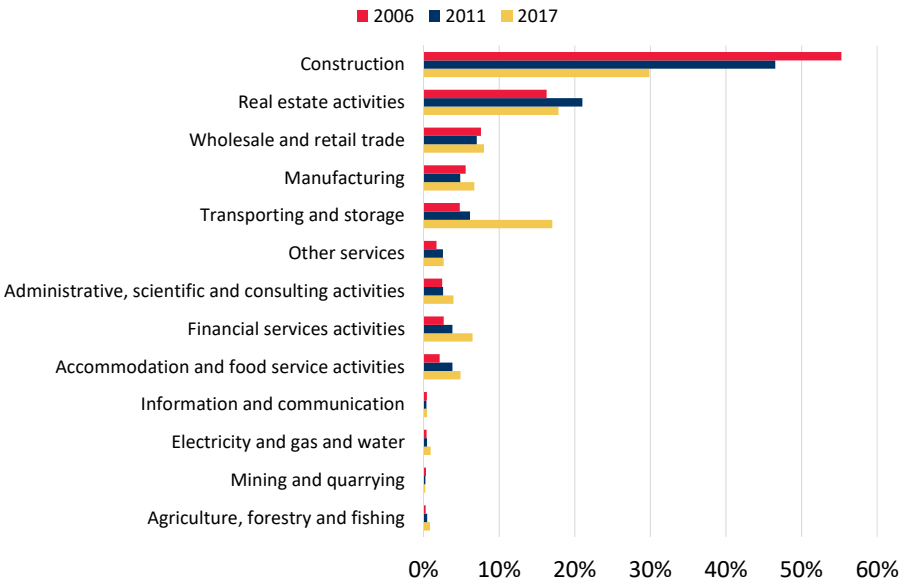


FIGURE 8: Contributions to $ES_{99.9\%}$.

For each year contributions must sum up 100%.

11. The magnitude of this difference depends significantly from the factor loading parameterization. Whenever one considers r=0.5, the homogenous factor loading proposed in Düllmann and Masschelein (2006), this effect is considerably mitigated.

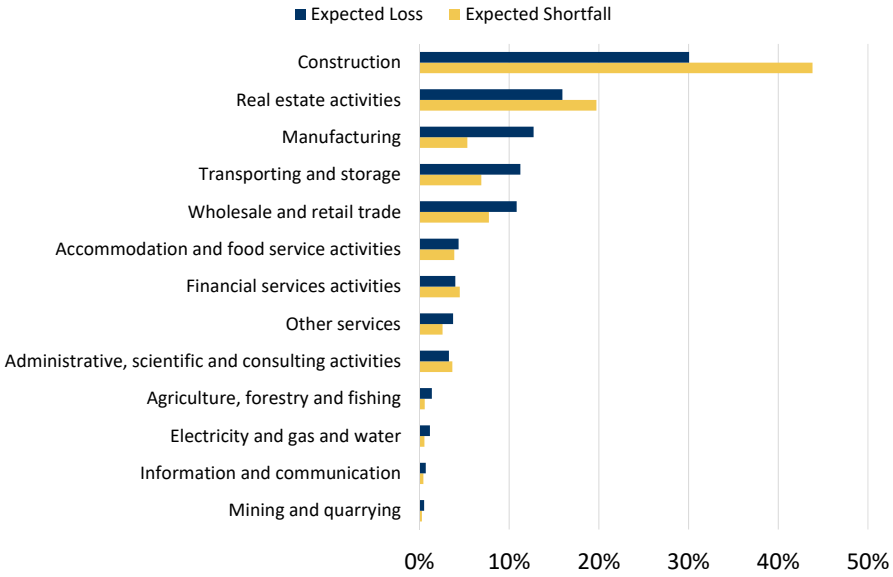


FIGURE 9: Average contributions to EL and $ES_{99.9\%}$.
For each measure contributions must sum up 100%.

Conclusion

The Basel capital framework has opted for a simple and transparent model that do not to explicitly account for portfolio concentration risk. This fact is then compensated in several ways. Still, the objective of this study is not to evaluate whether the Basel capital requirements is sufficiently conservative or not. As already argued, the fact that all the usual tail risk measures are largely dependent on the factor loading assumption, whose estimation is particularly challenging, significantly affects the value of this type of exercise. Instead, this study has three objectives. The first objective is to track the evolution of tail risk in banks' portfolio of performing loans. Under the model proposed in this article, tail risk increased significantly until 2013 and then started decreasing. The decline in tail risk measures such as the value-at-risk and the expected shortfall has been considerably more pronounced than the reduction in the expected loss. The second objective of this study is to analyze the determinants behind tail risk evolution. In particular, we are interested in the ratio between the unexpected loss and the expected loss, which is especially affected by interdependency in borrowers' defaults. Under our multi-factor model, where borrowers' correlations result mostly from sector concentration and inter-sector relations, the progressive reduction in banks' exposure to the construction sector causes the ratio between the unexpected loss and the expected loss to decrease gradually. The last objective

of this article is to call the reader's attention for the discrepancy between the marginal contribution of each loan to the expected loss and to the expected shortfall, depending on the borrowers' sector of activity. In particular, it is shown that the ratio between these two contributions is significantly above unity in the construction and real estate sectors while it is considerably below unity in sectors like manufacturing. This difference suggests the existence of potential diversification gains.

References

Accornero, Matteo, Giuseppe Cascarino, Roberto Felici, Fabio Parlapiano, and Alberto Maria Sorrentino (2017). "Credit risk in banks' exposures to non-financial firms." *European Financial Management*, pp. 1–17.

Antunes, António, Homero Gonçalves, Pedro Prego, *et al.* (2016). "Firm default probabilities revisited." *Economic Bulletin and Financial Stability Report Articles*.

BIS (2001). "The Internal ratings-based approach." *Bank for International Settlements*.

BIS (2005). "An explanatory note on the Basel II IRB risk weight functions." *Bank for International Settlements*.

Das, Sanjiv R, Darrell Duffie, Nikunj Kapadia, and Leandro Saita (2007). "Common failings: How corporate defaults are correlated." *The Journal of Finance*, 62(1), 93–117.

Düllmann, Klaus and Nancy Masschelein (2006). "Sector concentration in loan portfolios and economic capital." Tech. rep., Discussion Paper, Series 2: Banking and Financial Supervision.

Grippa, Pierpaolo and Lucyna Gornicka (2016). *Measuring Concentration Risk - A Partial Portfolio Approach*. International Monetary Fund.

Merton, Robert C (1974). "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance*, 29(2), 449–470.

Puzanova, Natalia and Klaus Düllmann (2013). "Systemic Risk Contributions: A Credit Portfolio Approach." *Journal of Banking and Finance*, 37.

Saldías, Martín (2013). "A market-based approach to sector risk determinants and transmission in the euro area." *Journal of Banking & Finance*, 37(11), 4534–4555.

Vasicek, Oldrich (2002). "The distribution of loan portfolio value." *Risk*, 15(12), 160–162.

Appendix A

The correlation between the systematic sector risk factors, Y_s, is referred as factor correlation and denoted by ρ_{ij}. Consider that Y_s (known as a composite factor) can be expressed as a linear combination of iid standard normal factors, Z, that impose the factor correlation structure between sectors:

$$Y_s = \sum_{k=1}^{S} \alpha_{s,k} Z_k, \quad \text{with} \quad \sum_{k=1}^{S} \alpha_{s,k}^2 = 1 \tag{A.1}$$

The matrix $(\alpha_{s,k})$ is obtained from the Cholesky decomposition of the sector correlation matrix, ρ_{ij} – Table B.2 Appendix B. To ensure that Y_s has unit variance it must hold that $\sum_{k=1}^{S} \alpha_{s,k}^2 = 1$.

The correlation between asset returns of two firms in sectors i and j is then obtained as:

$$\omega_{ij} = r_i r_j \rho_{ij} = r_i r_j \sum_{k=1}^{S} \alpha_{i,k} \alpha_{j,k}. \tag{A.2}$$

The correlation between the systematic sector factors and the sensitivity of the asset return to the composite factor determine the dependencies between firms. The intra-sector asset return correlation for each pair of firms is given by considering that $\rho_{ij} = 1$. In this case, $\omega_{ij} = r_s^2$.

Appendix B

Sector of activity	r_s
01 - Agriculture, forestry and fishing	0.229
02 - Mining and quarrying	0.303
03 - Manufacturing	0.098
04 - Electricity and gas and water	0.162
05 - Construction	0.457
06 - Wholesale and retail trade	0.199
07 - Transporting and storage	0.244
08 - Accommodation and food service activities	0.304
09 - Information and communication	0.258
10 - Real estate activities	0.363
11 - Financial services activities	0.472
12 - Administrative, scientific and consulting activities	0.422
13 - Other services	0.313

TABLE B.1. Factor Loadings.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	**1**	-0.03	0.28	0.03	0.29	0.36	0.02	-0.02	0.07	0.09	0.23	-0.12	0.16
2	-0.03	**1**	0.45	0.24	0.27	0.46	0.29	0.45	0.01	0.35	0.11	0.34	0.13
3	0.28	0.45	**1**	0.28	0.56	0.69	0.39	0.55	0.16	0.52	0.42	0.42	0.39
4	0.03	0.24	0.28	**1**	0.46	0.36	0.2	0.3	0.33	0.32	0.35	0.32	0.13
5	0.29	0.27	0.56	0.46	**1**	0.64	0.3	0.42	0.45	0.76	0.51	0.45	0.39
6	0.36	0.46	0.69	0.36	0.64	**1**	0.42	0.54	0.49	0.65	0.44	0.56	0.25
7	0.02	0.29	0.39	0.2	0.3	0.42	**1**	0.53	0.18	0.38	0.27	0.56	0.21
8	-0.02	0.45	0.55	0.3	0.42	0.54	0.53	**1**	0.05	0.42	0.5	0.45	0.51
9	0.07	0.01	0.16	0.33	0.45	0.49	0.18	0.05	**1**	0.5	0.4	0.33	0.06
10	0.09	0.35	0.52	0.32	0.76	0.65	0.38	0.42	0.5	**1**	0.32	0.6	0.28
11	0.23	0.11	0.42	0.35	0.51	0.44	0.27	0.5	0.4	0.32	**1**	0.28	0.6
12	-0.12	0.34	0.42	0.32	0.45	0.56	0.56	0.45	0.33	0.6	0.28	**1**	0.3
13	0.16	0.13	0.39	0.13	0.39	0.25	0.21	0.51	0.06	0.28	0.6	0.3	**1**

TABLE B.2. Sectoral Correlations.

Cyclically-adjusted current account balances in Portugal

João Amador
Banco de Portugal
Nova School of Business and Economics

João Falcão Silva
Banco de Portugal
Nova School of Business and Economics

January 2019

Abstract
This article uses the methodology suggested by Fabiani *et al.* (2016) to compute cyclically-adjusted current account balances for the Portuguese economy in the period 1995-2017. The methodology makes use of domestic and foreign output gaps, export elasticities and the import content of domestic demand, distinguishing between cyclically-adjusted exports and imports. In addition, we compute the cyclically-adjusted bilateral exports and imports relative to the main Portuguese trade partners. We conclude that the strong current account adjustment observed in the Portuguese economy after 2010 was mainly structural, though a positive effect resulting from cyclical developments was also observed. (JEL: E32, F32, F40)

Introduction

The increase of the current account balance after 2010 is one of the major features of the macroeconomic rebalancing of the Portuguese economy, which took place in the context of the Portuguese Economic and Financial Assistance Program, implemented in the aftermath of the sovereign debt crisis in the euro area. According to the statistics of the Balance of Payments, the Portuguese current account balance evolved from a deficit of approximately 10 per cent of GDP in 2010 to a surplus of 0.5 per cent of GDP in 2017. Sizable current account adjustments have also taken place in other European Union (EU) countries. In this context, an important question is whether such developments resulted from a structural adjustment or simply from cyclical developments. This article tries to answer this question for the Portuguese economy.

Acknowledgements: The authors are thankful to Nuno Alves, António Antunes, Sónia Cabral, Sónia Félix, Miguel Gouveia, José Ramos Maria, Filipa Lima and António Rua for useful comments and suggestions. All errors and omissions are the sole responsability of the authors. The opinions expressed in this article are those of the authors and do not necessarily coincide with those of Banco de Portugal or the Eurosystem.

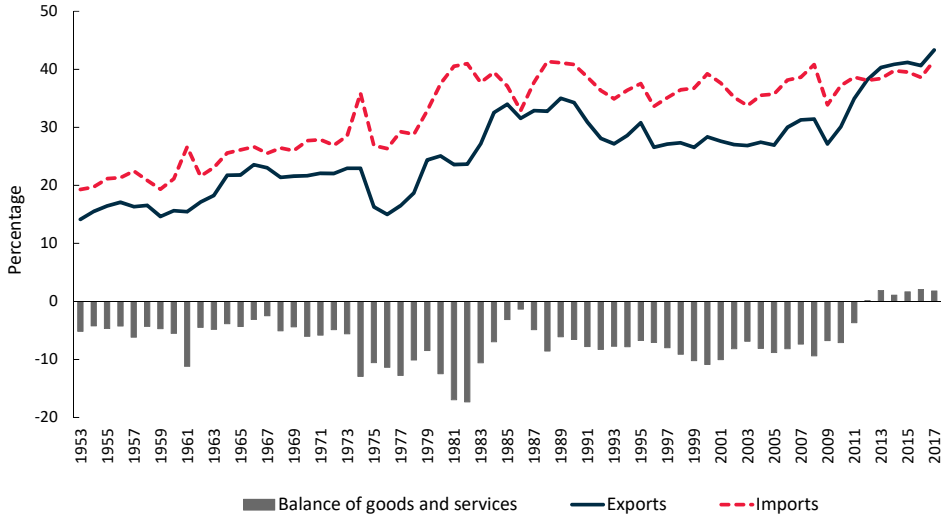
E-mail: jamador@bportugal.pt; jmfsilva@bportugal.pt

Current account imbalances and subsequent external financing difficulties have been recurrent in Portugal over the last six decades. In 1977-78 and 1983-84 Portugal underwent economic stabilization programs with the International Monetary Fund (IMF). Low private savings, important investment needs and fiscal imbalances repeatedly boiled down to deficits in the external accounts and sizable external financing requirements.

Figure 1 plots the share of exports, imports and the balance of goods and services as a percentage of GDP in a historical perspective. Economic developments in the Portuguese economy in the nineties and in the first decade of this century were characterized by large current account deficits that led to a strong deterioration of the net international investment position, which reached -108 per cent of GDP in 2009. The decreasing interest rates associated to the transition to a low inflation regime, on the way to the accession to the monetary union, greatly expanded domestic demand and this was aggravated by a pro-cyclical fiscal stance. The higher imports associated with the growing domestic demand coincided with a reshuffling of comparative advantages that led to a sizable loss of export market. This was motivated by the EU enlargement to Central and Eastern European countries and strong Asian competition. Moreover, the sluggish adjustment to the macroeconomic imbalances and the slow shift of resources from the non-tradable into the tradable sector implied a prolonged exposure to external risks, which materialized with the 2008 economic and financial crisis. The sudden-stop of external financing in some euro area countries and the self-reinforcing loop between bank and sovereign debt risks threatened the monetary union (see, for example, Salto and Turrini (2010)). In Portugal, the strong difficulties to access external financing led to an external assistance program in 2011 involving the European Commission, European Central Bank and the IMF, which included conditionality in several areas.

The period after 2011 has been characterized by improvements in the Portuguese external balance. As visible in Figure 1, these developments have been quite significant in historical terms. The small surpluses recently recorded in the balance of goods and services are in striking contrast with the large deficits of the last decades. Nevertheless, the adjustment of the Portuguese external balance took place in a context of contraction of economic activity, thus raising concerns about its sustainability in the recovery phase of the cycle. A complementary issue is the impact on the balance of goods and services of economic developments in the main trade partners, for example, to what extent the domestic adjustment in external accounts was made harder by parallel improvements in the current account balance of trade partners.

The literature comparing structural and cyclical current account balances has been growing in the last years. Initial methodological contributions were those of Sachs (1981) and Buiter (1981), while Obstfeld and Rogoff (1995) approached this topic from an intertemporal perspective. Several empirical applications, mostly basing on the relationship between external balances and



Source: Banco de Portugal (Séries Longas and BPStat; Statistics of the Balance of Payments)

FIGURE 1: Balance of goods and services as a percentage of GDP in Portugal 1952-2017

the savings-investment gap, discuss the fundamental determinants of current account balances (e.g. Faruqee and Debelle 1996; Milesi-Ferretti and Blanchard 2011; Chinn and Prasad 2003; Gruber and Kamin 2005; CáZorzi and A. Chudik 2009).

The literature presents two main methods of adjusting the current account balance for the impact of the cycle. The first method bases on the estimation of regressions where the current account balance is correlated with a set of demographic, macroeconomic, financial and institutional variables. The structural current account is obtained by applying the estimated coefficients to the (medium-term) trend values of the explanatory variables. This approach typically considers a panel of countries over a long period of time. Alternatively, it is possible to obtain the cyclical adjustment by estimating a short-run equation with the lagged current account balance and a set of variables that do not affect structural positions but have a short-run influence on the current account.

International organizations have been using and developing this type of methods. The IMF Consultative Group on Exchange Rates (CGER) and its most recent External Balance Assessment (EBA) method are a good example (see Phillips *et al.* (2013)). The European Commission has been using a method broadly similar to that of the IMF EBA, producing specific policy indicators. The OECD has also been using this type of methodology. In particular, Cheung and Rusticelli (2010) assess the link between structural and cyclical determinants of current account balances using panel data on dimensions like differences in demographics, fiscal positions, oil dependency, oil intensity and stage of economic development, amongst others. Tamara (2016) refers the

caveats of this type of methodology, pointing out that current account balances are estimated directly, considering both fundamental and shorter-term factors. Although the EBA framework is considered a strongly integrated and robust current account predictor, it is sensitive to data sources and endogeneity problems between current account balances and output gaps may arise. Moreover, this methodology does not consider the heterogeneity between countries neither, as mentioned by Sastre and Viani (2014), competitiveness factors.

As for Portugal, Afonso and Silva (2017) studied the decomposition of the current account between cyclical and structural components, using Germany as a benchmark to assess its determinants. More recently, Afonso and Jalles (2018) distinguished between cyclical and non-cyclical current account determinants, while providing a refinement and a counter check of the methodologies used when conducting policy decisions.

The second method of computing structural current account balances, focuses on the goods and services account and bases on international trade elasticities. A strong advantage of this approach is the possibility of adjusting separately the export and import components of the current account. Haltmaier (2014) quantifies the cyclical part of the current account balance for several countries by estimating a long-run (or trend) elasticity from a co-integration relationship between trade and income, as well as a short-run (or cyclical) elasticity.[1] The caveats of this approach lie on the uncertainty and revisions associated to output gaps and trade elasticities. In addition, it should be highlighted that the adjustments resulting from the methodology relate exclusively to the output gaps, i.e., all other changes in exports or imports attributable to temporary aspects are included in the structural component. This partly explains the moderate deviations between observed and cyclically-adjusted current account balances. Overall, the two methodological approaches should be taken as complementary and not as substitutes.

An important contribution to the latter strand of literature is that of Fabiani *et al.* (2016), which suggests a model that relies on trade elasticities for exports and imports. The authors focus on the Italian case but also apply the methodology to France, Germany and Spain. According to the results, the overall balancing of the Italian external accounts has largely been of a non-cyclical nature, with a positive contribution coming from the decline in the prices of energy commodities. For the other countries considered, they find that current account imbalances over the recent period are amplified when assessed in cyclically-adjusted terms. One important feature of Fabiani *et al.* (2016) is the explicit consideration of the composition effects associated with

1. The effects of foreign and domestic output gaps on real exchange rate deviations are used in other models, such as Wu (2008) and Kara and Sarikaya (2013).

the different components of domestic demand, as suggested by Bussière *et al.* (2013).

In this article we apply the methodology suggested by Fabiani *et al.* (2016) to the Portuguese economy in the period 1996-2017. We consider the cyclical adjustment of the current account, both for exports and imports. However, we do not discuss elements associated with energy prices nor with the income account. Nevertheless, we go beyond Fabiani *et al.* (2016) by calculating the adjusted exports and imports relatively to the main Portuguese trade partners, making use of estimated bilateral trade elasticities.

The rest of the article is organized as follows. In the next section, we briefly describe the methodology used for the cyclical adjustment of exports and imports, as suggested by Fabiani *et al.* (2016). Section *Data* identifies the data sources. The following section presents the results obtained in aggregate terms, details relatively to the main trade partners and discusses their robustness by using different output gaps and trade elasticities. The last section offers some concluding remarks.

Methodology

Aggregate adjustment

This section closely draws on Fabiani *et al.* (2016) to explain the main features of the model that generates the expressions used for the elasticity of exports and imports to foreign and domestic output gaps, respectively. We start from the basic definition of the current account balance (CAB):

$$CAB = Exports - Imports + BPI + BSI \tag{1}$$

where BPI and BSI stand for "Balance of Primary Income" and "Balance of Secondary Income", respectively. Nevertheless, our adjustment focuses exclusively on the goods and services account. In terms of notation, the home and foreign economies are presented as H and F, respectively. Moreover, current and potential GDP in the home country, in real terms, are identified as Y and Y^*, respectively. In the same way X^* and M^* stand for potential exports and imports in the home economy, in real terms. In addition, nominal variables are denoted as the product of the real counterpart and the corresponding price index.

As in Fabiani *et al.* (2016), home imports and exports are taken to be *isoelastic*, which means that an exogenously given constant long-run elasticity is assumed. Therefore, if the foreign (home) GDP increases by one percent, exports (imports) increase by $\Delta X(\Delta M)$ percent. Starting with the export side, potential exports in real terms are obtained as:

$$X^* = X + \Delta X = \\ = X\left(1 + \frac{\Delta X}{X}\right) = X\left(1 + \theta_x \times \frac{\Delta Y^F}{Y^F}\right) = X\left(1 + \theta_x \times \frac{-y^F}{1+y^F}\right) \quad (2)$$

where ΔX and ΔY^F are the differences between observed and prevailing levels of real exports and real foreign output at the potential (i.e., distances to the potential and not changes between consecutive periods), respectively, and θ_x represents the long-run elasticity of exports to foreign real GDP. In addition, the definition of the foreign output gap $y^F = (Y^F - Y^{*F})/Y^{*F}$ establishes the last term in equation (2):

$$\frac{\Delta Y^F}{Y^F} = \frac{-y^F}{1+y^F} \quad (3)$$

Next, assuming that prices (P_X and P_Y) are unchanged, the cyclically adjusted nominal exports (x^{adj}) is obtained by multiplying the unadjusted export share on GDP (x, computed in nominal terms) by the ratio of potential to actual real exports:

$$x^{adj} = \frac{P_X X^*}{P_Y Y} = \frac{P_X X}{P_Y Y} \times \frac{X^*}{X} = x\frac{X^*}{X} \quad (4)$$

Finally, combining equations (2) and (4), we write cyclically adjusted exports as:

$$x^{adj} = x\left(1 - \theta_x \frac{y^F}{1+y^F}\right) \quad (5)$$

The key exogenous variable is the foreign output gap y^F and the intuition is straightforward: the cyclical adjustment of exports depends negatively on the foreign output gap. If Portuguese trade partners' output is higher than their potential, they will import more and consequently domestic exports benefit from the cycle. The crucial export elasticity is based on the cross-country panel regression in Bussière *et al.* (2013).[2] In the Appendix A we present the methodology and results for the elasticities of home exports to foreign GDP ($\theta_x = 2.6$).

If home imports are assumed to be isoelastic to home GDP, an expression similar to that used for exports could be applied to determine cyclically-adjusted imports. However, as stated by Fabiani *et al.* (2016), this would be a very strong simplification for the import side. Imports are activated

2. In the panel regression we considered the following OECD countries: Australia; Belgium; Canada; Finland; France; Germany; Italy; Japan; Korea; Netherlands; New Zealand; Norway; Spain; Sweden; United Kingdom; United States. These were also the countries considered by Bussière *et al.* (2013), except for Denmark, for which the information was not available. The foreign output gap is the weighted average of individual output gaps with weights proportional to the share of these countries in Portuguese exports.

by demand, rather than GDP, thus it may be misleading not to distinguish between components of demand in order to allow for their different import intensities.

Bussière *et al.* (2013) suggests a new measure that reflects the import intensity of the different components of domestic expenditure and the import content of exports. This import intensity-adjusted measure of demand is labeled as IAD, and it is constructed for each country as:

$$IAD_t = C_t^{\omega_{C,t}} G_t^{\omega_{G,t}} I_t^{\omega_{I,t}} X_t^{\omega_{X,t}} \tag{6}$$

where C stands for private consumption, G for government consumption, I for investment, and X for exports. The weights, $\omega_{k,t}$, with $k = C, G, I, X$ are the total import contents of these final demand components. These weights are time-varying and normalized in each period such that their sum equals one.

Bussière *et al.* (2013) model imports as being activated by a geometric weighted average of the various demand components, with weights reflecting their relative import contents. The authors present rolling-window estimates confirming that the assumption of a stationary, time-invariant long-run elasticity of imports is reasonable only in the case of the IAD variable, whereas the long-run elasticity of imports to GDP shows an increasing trend. In this article, the IAD approach is implemented in a reduced-form approach, as in Fabiani *et al.* (2016). While the original version separately considers four components of demand (private consumption, public consumption, investment, exports), we just isolate the component that typically shows the highest import intensity: exports. This approach has also been used by Christodoulopoulou and Tkacevs (2016).

As in the case of exports, real imports are assumed to be isoelastic relatively to the reduced form IAD variable, which is a convex combination of exports and domestic demand (in log terms). Therefore, the growth rate of imports is given by:

$$\frac{\Delta M}{M} = \theta_M^{IAD} \frac{\Delta IAD}{IAD} = \theta_M^{IAD} \left[\omega_x \frac{\Delta X}{X} + (1 - \omega_x) \frac{\Delta DD}{DD} \right] \tag{7}$$

where θ_M^{IAD} is the constant long-run elasticity relatively to imports, which is calibrated using the regressions suggested in Bussière *et al.* (2013), ω_x is the weight of exports in building the IAD variable, and DD stands for domestic demand. As in Bussière *et al.* (2013) we compute the import intensity of each IAD component with global input-output tables, using a linear interpolation to construct quarterly series and normalizing so that they sum to unity.

Taking Δ as the difference between potential and current levels of the variables, potential imports are defined as:

$$M^* = M + \Delta M = M + \theta_M^{IAD} \omega_x \left(\frac{M}{X} \right) \Delta X + \theta_M^{IAD} (1 - \omega_x) \left(\frac{M}{DD} \right) \Delta DD \tag{8}$$

where $\theta_M^{IAD} = (\Delta IAD/IAD)/(\Delta Y/Y)$.

Similarly to what was done for export elasticities, the methodology and panel regression results for the elasticity of IAD are presented in Appendix A ($\theta_M^{IAD} = 1.48$). Next, equation 8 can be simplified to:

$$M^* = M + \eta_X(X^* - X) + \eta_D(DD^* - DD) \tag{9}$$

where $\eta_X = \theta_M^{IAD}\omega_x \frac{M}{X}$ and $\eta_D = \theta_M^{IAD}(1-\omega_x)\frac{M}{DD}$.

Considering the national accounts identity $Y^* = DD^* + X^* - M^*$ and including equation (9) we obtain:

$$Y^* = DD^* + X^* - [M + \eta_X(X^* - X) + \eta_D(DD^* - DD)] \tag{10}$$

then, solving with respect to DD it is possible to write equation (9) as:

$$M^* = M + \frac{\eta_D(Y^* - Y)}{1-\eta_D} + \frac{(X^* - X)(\eta_X - \eta_D)}{1-\eta_D} \tag{11}$$

Equation (11) expresses the level of imports that would prevail if domestic and foreign output were jointly taken at their potential level, thus simultaneously determining (home) exports and domestic demand. These are the two components of aggregate demand that activate imports, each with a specific intensity. Moreover, the relative share of potential domestic demand and potential exports determine potential imports and are coherent with potential output.

As in the case of exports, the ratio between potential and actual imports in real terms is sufficient to pin down cyclically-adjusted nominal imports (nominal potential imports as a percentage of nominal unadjusted GDP):

$$m^{adj} = \frac{p_M M^*}{p_Y Y} = \frac{p_M M}{p_Y Y}\frac{M^*}{M} = m\frac{M^*}{M} \tag{12}$$

where m denotes the unadjusted import share on GDP (computed in nominal terms). Finally, the adjusted current account, which is the ultimate object of interest, is given by:

$$ca^{adj} = x^{adj} - m^{adj} + bpi + bsi, \tag{13}$$

where bpi and bsi denote the unadjusted balance of primary income and secondary income, as percentage of GDP.

Bilateral adjustment

In this article, we go beyond the methodology previously presented and take a bilateral perspective. Conceptually, this is not different from what was described above, though it involves explicitly considering the output gap of

the different trading partners and the structure of imports originating from them. Therefore, there is a larger number of (bilateral) import elasticities to be estimated.

On the export side, the cyclically adjusted exports of country i (home) to country j are obtained as:

$$x_{ij}^{adj} = x_{ij}\left(1 - \theta_x \frac{y_j}{1+y_j}\right) \tag{14}$$

where x_{ij} represents the unadjusted bilateral exports of country i to country j on home GDP. As before, we assume that the long-run elasticity of exports is the same for all countries: $\theta_x = 2.6$. The main difference is that the adjustment of bilateral exports relies on the foreign output gap which, in this case, is considered to be the individual output gap of country j and not a weighted average of those of the main trade partners.

The cyclical adjustment of imports of country i from country j is given by:

$$m_{ij}^{adj} = m_{ij}\frac{M_{ij}^*}{M_{ij}} \tag{15}$$

where m_{ij} represents the unadjusted bilateral imports of country i from country j on GDP of country i and M_{ij}^* measures the bilateral potential imports, which are defined as:

$$M_{ij}^* = M_{ij} + \frac{\eta_{ij}^D(Y^* - Y)}{1-\eta_{ij}^D} + \frac{(X_{ij}^* - X_{ij})(\eta_{ij}^X - \eta_{ij}^D)}{1-\eta_{ij}^D} \tag{16}$$

In addition, bilateral elasticities are given by:

$$\eta_{ij}^X = \theta_{Mij}^{IAD}\omega_x\frac{M_i}{X_i} \tag{17}$$

and

$$\eta_{ij}^D = \theta_{Mij}^{IAD}(1-\omega_x)\frac{M_i}{DD_i} \tag{18}$$

where θ_{Mij}^{IAD} represents the bilateral elasticity of the IAD variable.

Data

The implementation of the methodologies described in the previous section required a large amount of statistical information and some hypotheses. Firstly, the source of comparable cross-country data was the OECD Economic Outlook (November 2018). In particular, we used quarterly data from Q4 1995 until Q4 2017 for the volumes of GDP and its components: government consumption, private consumption, gross total fixed capital formation,

imports and exports of goods and services. Moreover, we collected the corresponding deflators of GDP and total imports of goods and services.

Secondly, the information on the domestic and foreign output gaps, which are key elements in the methodology, was collected from the IMF World Economic Outlook (April 2018). It is widely acknowledged that estimates of output gaps depend on the method used for computation (statistical or structural methods) and are sensitive to revisions of data.[3] For this reason in subsection *Robustness* we evaluate the results obtained with different output gaps for the Portuguese economy. Nevertheless, in order to ensure the consistency of results we take the a common statistical source for domestic and foreign output gaps: the IMF World Economic Outlook.

Thirdly, the estimation of the long-run elasticity of the IAD requires information contained in global about input-output matrices. For this purpose we used the 2016^{th} edition of the OECD Inter-Country Input-Output database (ICIO), which includes information for a total of 71 countries and 34 industries (according to a classification based on ISIC Rev3) on an annual basis from 1995 until 2011.

Finally, bilateral trade flows are not available in existing databases. Therefore, in order to break down the aggregate of total real imports in the OECD database, we assume that the share of each country on nominal and real Portuguese total imports is equal. The shares of the different partners in nominal trade flows are taken from Portuguese National Statistics.

Results

In this section, we present the results for the cyclically-adjusted current account balance of the Portuguese economy between 1995 and 2017. Firstly, we present the results for trade elasticities estimations. Secondly, we separately examine the adjustment for exports and imports. Thirdly, we compute the cyclical adjustment of exports relatively to the main Portuguese trade partners. Moreover, we present the cyclically adjusted current account balance for different series of the Portuguese output gap. Finally, we test the impact on the cyclical adjustment that results from using different elasticities. These two exercises make it possible to evaluate the robustness of the main results, while highlighting the uncertainty underlying this methodological approach.

We estimated trade elasticities both for exports and imports according to the methodology previously described. The Appendix A presents the results of the elasticity of home exports to foreign GDP (Table A.2). As in Bussière

3. For a discussion on output gap methodologies with an emphasis on Portugal see Banco de Portugal (2017).

et al. (2013), the exports elasticity is obtained through a panel regression and is assumed to be the same for all countries. We considered only the coefficients statistically significant at a 10 percent level and obtain $\theta_x = 2.6$.[4] The elasticity of imports to the *IAD* is also described in Appendix A and, using the statistically significant parameters, it is equal to $\theta_M^{IAD} = 1.48$.

Cyclically-adjusted exports and imports

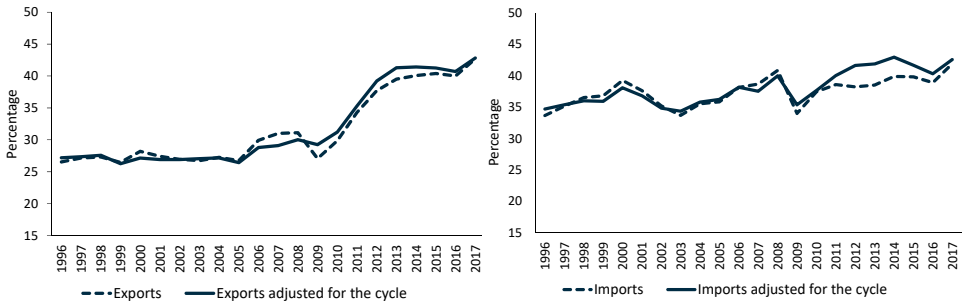
Panel A of Figure 2 presents the series for the observed and cyclically-adjusted Portuguese exports as a percentage of GDP, basing on equation (5). The element that stands out is the sharp increase in the share of exports as a percentage of GDP since the turn of the century. This corresponds to the adjustment of the Portuguese productive structure to the new pattern of comparative advantages that followed the enlargement of the EU to Central and Eastern European countries and the rise of Asian competition in the mid-nineties. Those were negative shocks to Portuguese exports and the recovery that followed started well before the economic and financial crisis of 2008 and the subsequent sovereign debt crisis in the euro area.

The cyclical developments in foreign clients did not strongly affect the path of domestic exports. In the years before the 2008 crisis, the positive foreign output gaps drove Portuguese exports above their structural level. Conversely, the problems that emerged in the aftermath of the sovereign debt crisis led the ratio of exports on GDP to increase less than potential. More recently, the dynamics of exports moderated and they have remained close to the structural level as a percentage of GDP. Overall, the gap between observed and structural export to GDP ratios has been relatively small, never exceeding 2.2 percentage points (p.p.) in absolute terms (Appendix B).

In panel B of Figure 2 we show the results for the adjustment of Portuguese imports to the domestic cycle, taking into account the structure of domestic demand, as presented in equation (12). The results show that from 1996 to 2008 the changes in imports of goods and services as a percentage of GDP were largely of a structural nature. Nevertheless, after this period the observed import ratio stood systematically below the structural level, meaning that the contraction of domestic demand that was associated to a negative output gap brought down imports significantly. In this period, the strongest cyclical adjustment of imports represented 3.4 p.p. of GDP in 2012 and 2013, while the smallest adjustment stood close to zero in 2006 (Appendix B).

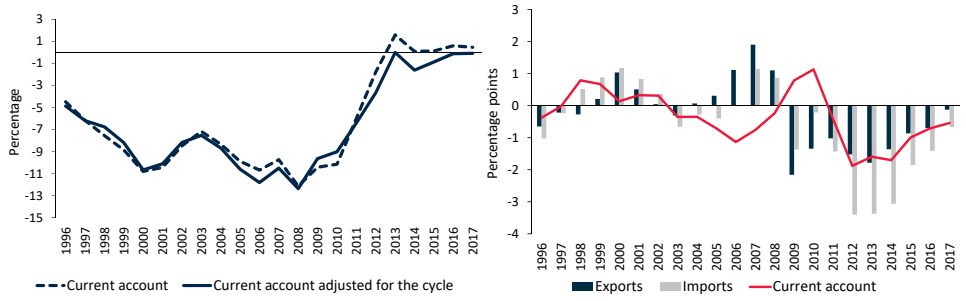
When the cyclical adjustment of exports and imports is combined, we obtain the proxy of the structural current account balance as a percentage of GDP for the Portuguese economy (Figure 3). In panel A we present the balance

4. In the robustness section we assess the impact of considering exactly the same export elasticity as in Bussière *et al.* (2013).



(A) Exports (B) Imports

FIGURE 2: Cyclically-adjusted exports and imports (percentage of GDP), national accounts statistics



(A) Observed and cyclically-adjusted current account balances (B) Contributions to cyclical adjustments

FIGURE 3: Cyclically-adjusted current account balance (percentage of GDP), national accounts statistics

and in panel B the contributions of exports and imports to the difference between the adjusted and observed values. According to our results, the observed external balance stood about 0.5 p.p. of GDP lower than structural in the period 1998-2001, mostly due to the impact of the cycle on imports. From 2003 onwards the adjustment reversed (except in 2009 and 2010), amounting to 1.5 p.p. of GDP in the average of the period 2012-2015 period, due to the effect of imports, which was not compensated by the fact that exports also stood below their structural level. Finally, in the most recent years the gap between adjusted and non-adjusted current account balances progressively diminished to 0.5 p.p. in 2017.

Overall, the adjustment of the Portuguese current account balance to the economic cycle is not very large. Nevertheless, a clear message is that most of the correction observed in the Portuguese current account balance in the latest years has a structural nature. Although the structural balance remains

negative in the period studied, 2017 stands as the year with the second lowest deficit in the sample (-0.1 per cent of GDP).

Detail for the main trade partners

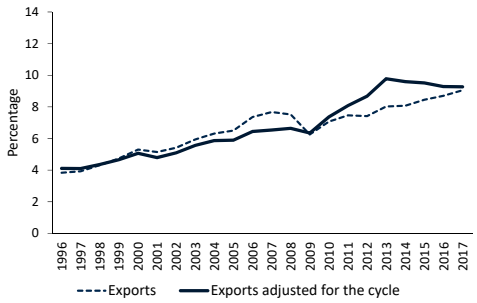
The developments in the Portuguese current account balance are affected by cyclical developments in the main trade partners, notably in terms of demand for Portuguese exports. Moreover, Portuguese imports adjusted for demand differ for each trade partner. Therefore, by using the estimated bilateral elasticities, changes in the domestic output gap have a different impact on imports from each trade partner. In this subsection we take Spain, Germany and France and assess the cyclical adjustment on bilateral exports and imports.[5]

These three countries represent a large share of Portuguese international trade in the period considered. Spain, Germany and France are the three top export destinations and import origins, representing together 60 and 70 per cent of these aggregates in 2017, respectively.

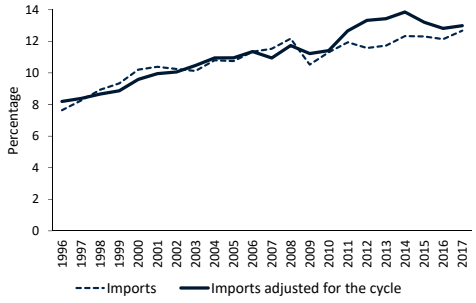
Figure 4 presents the results for the three countries and shows some differences. Spain (panels A and B), which has been reinforcing its role as the main trade partner, is the country where the the distance between the observed and structural exports a percentage of GDP is higher. The structural exports stood above the observed ratio in the years before the sovereign debt crisis but turned significantly below trend afterwards due to the downturn in the Spanish economy, while correcting its own macroeconomic imbalances. Nevertheless, this gap has diminished in 2017. As for Portuguese structural imports from Spain, they stood slightly above the observed ratio up to the sovereign debt crisis but the severe downturn of the Portuguese economy reversed this situation. Overall, the adjustment in exports and imports partly offset each other, which should be seen as a normal situation among strongly integrated economies, whose business cycles are synchronized.

Relatively to Germany, which has broadly stabilized its importance as a Portuguese trade partner, the adjustments in exports are very small (panel C). This is partly explained by the fact that this country was not significantly affected by the sovereign debt crisis in the euro area. As for imports, the adjustment is important and results from the high bilateral elasticity estimated for the import content of domestic demand components (panel D). As for France (panels E and F), whose share in Portuguese exports has increased very significantly in the latest years, structural exports and, mostly, imports stood above what was observed.

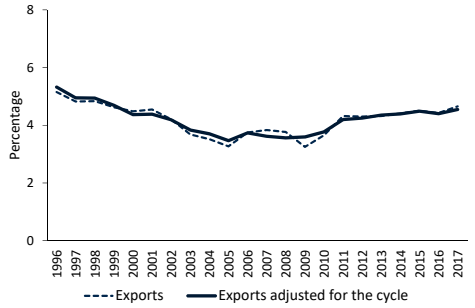
5. Bilateral IAD coefficients for Spain, Germany and France vis-à-vis Portugal are: $\theta^{IAD}_{M_{ESP}} = 0.94$; $\theta^{IAD}_{M_{DEU}} = 1.57$; $\theta^{IAD}_{M_{FRA}} = 0.84$.



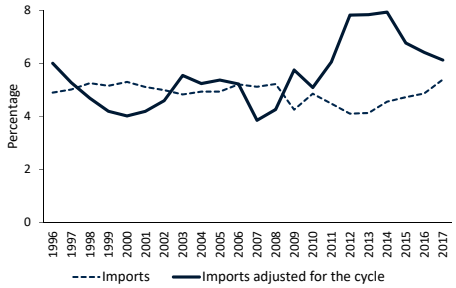
(A) Spain - exports



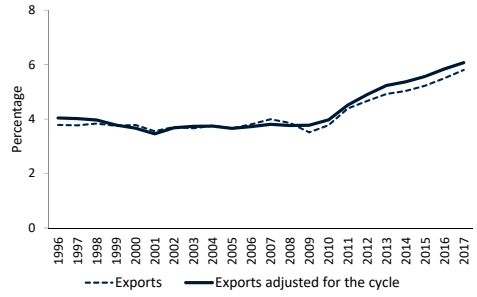
(B) Spain - imports



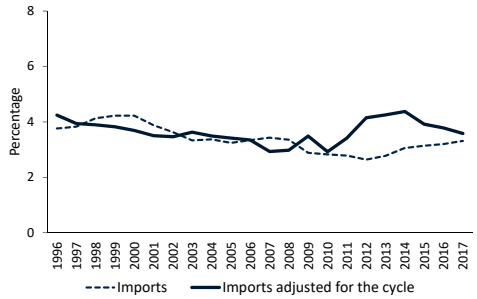
(C) Germany - exports



(D) Germany - imports



(E) France - exports

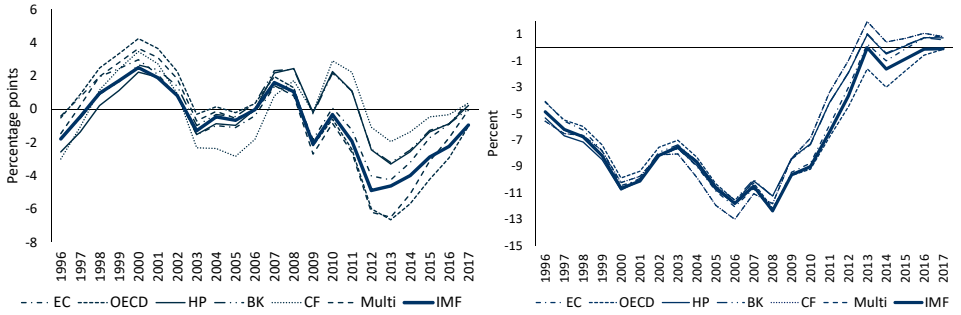


(F) France - imports

FIGURE 4: Cyclically-adjusted exports/imports vis-à-vis Spain, Germany and France (percentage of GDP)

Robustness

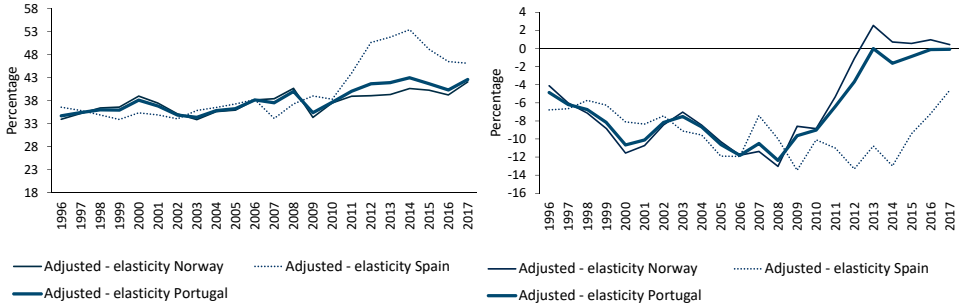
There is uncertainty regarding some parameters in the methodology, which may affect the results obtained for the Portuguese cyclically-adjusted current account balance as a percentage of GDP. In order to assess the robustness of results, we recomputed the adjusted current account balances with different series for the Portuguese output gap and for a range of import elasticity estimates.



(A) Output gaps (B) Adjusted current account

FIGURE 5: Robustness of results - Output gap

Notes: HP- Hodrick–Prescott filter; BK- Baxter-King filter; CF- Christiano-Fitzgerald filter, Multi- Multivariate filter.



(A) Adjusted imports (B) Adjusted current account

FIGURE 6: Robustness of results - Elasticity of imported adjusted demand

Panel A of Figure 5 plots several series for the Portuguese output gap from 1996 to 2017. Beyond our baseline output gap (of the IMF) we show estimates by the OECD, European Commission and calculations by Banco de Portugal with different statistical filters. The range of output gap estimations is considerable, reaching more than 4 p.p. of GDP in some periods. The panel B of Figure 5 plots the cyclically adjusted balances with the different output gap series. This exercise only affects the adjusted imports and it is visible that the main features of the results are not altered. Foreign output gaps are part of the calculations for cyclically-adjusted exports but the consideration of different estimates for all these variables is beyond the scope of this article.

In addition, we computed the cyclically adjusted imports and the subsequent current account balance using the highest and lowest import elasticities that would emerge from adopting the methodology for the set of countries considered to compute the Portuguese external demand,

particularly the ones for Spain ($\theta_{IAD}^{ESP} = 2.68$) and Norway ($\theta_{IAD}^{NOR} = 0.51$), respectively (Figure 6). The difference relatively to the benchmark situation is strong if we use the Spanish elasticity as the structural adjustment only takes place in the recent years. In any case the structural correction of the Portuguese current account balance is visible. It should be noted that these alternative elasticities affect the parameters η_X and η_D in equation (11) and have a non-linear impact on adjusted imports.

Another robustness exercise consists of computing the cyclically-adjusted current account balance with the export elasticity used by Fabiani *et al.* (2016), that is $\theta_x = 1.9$ instead of our $\theta_x = 2.6$. We observe that this change does not affect the structural current account balance in any significant way, thus we do not plot it. Finally, we replicated the overall exercise excluding exports and imports of energy products and the results remain qualitatively unchanged.

Final remarks

The current account balance is a key macroeconomic indicator. Although in the nineties and early years of the new century its importance was somewhat downplayed for the case of countries taking part in a monetary union, the global economic and financial crisis of 2008 and the euro area sovereign debt crisis that followed have shown that countries cannot run prolonged current account deficits and strongly deteriorate the net external position.

As in the case of other macroeconomic variables, exports and imports are affected by cyclical developments. Therefore, it is important to disentangle structural and cyclical developments. In this article, we adopt the methodology presented by Fabiani *et al.* (2016) and apply it to the Portuguese economy in the period 1995-2017. In addition, we extend the analysis to the bilateral dimension and identify specific adjustments for the Portuguese exports and imports with its main trade partners.

We conclude that the strong current account adjustment observed in the Portuguese economy after 2010 was mainly structural, though a positive effect from cyclical developments is also observed. Taking the average of the period 2012-2017, the cyclically adjusted current account balance lies 1.2 p.p. below the observed balance. In 2017, the structural current account balance stood at -0.1 percent of GDP. The results are robust for different series of the Portuguese output gap and import elasticities. As for the bilateral analysis, we conclude that the recession in the main Portuguese trade partner (Spain) deteriorated Portuguese exports. However, for Germany and France the adjustments to exports are small but relevant for imports.

The Portuguese current account balance has strongly improved after the euro area sovereign debt crisis and the subsequent Portuguese economic and financial assistance program. Although the methodology only adjusts the current account balance for domestic and foreign output gaps, thus

leaving other all other fluctuations unaffected, the structural nature of the Portuguese adjustment is visible. Nevertheless, this trend should be reinforced and a continuing screening of current account developments is necessary. Only through near balance or positive current account balances will the Portuguese external indebtedness decrease, reducing exposure to future external economic and financial risks.

References

Afonso, António and João Jalles (2018). "Decomposing and Analysing the Determinants of Current Accounts' Cyclicality: Evidence From the Euro Area." Working Papers 2018-42, ISEG - Lisbon School of Economics and Management.

Afonso, António and Jorge Silva (2017). "Current account balance cyclicality." *Applied Economics Letters*, 24(13), 911–917.

Banco de Portugal (2017). "Potential output: Challenges and uncertainties."

Buiter, Willem (1981). "Time Preference and International Lending and Borrowing in an Overlapping-Generations Model." *Journal of Political Economy*, 89(4), 769–97.

Bussière, Matthieu, Giovanni Callegari, Fabio Ghironi, Giulia Sestieri, and Norihiko Yamano (2013). "Estimating Trade Elasticities: Demand Composition and the Trade Collapse of 2008-2009." *American Economic Journal: Macroeconomics*, 5(3), 118–151.

CáZorzi, M. and A. Dieppe A. Chudik (2009). "Current account benchmarks for Central and Eastern Europe: A desperate search?" ECB Working Papers 995, European Central Bank.

Cheung, D. Furceri, C. and E. Rusticelli (2010). "Structural and Cyclical Factors behind Current-Account Balances." Working paper 775, OECD Economics Department.

Chinn, Menzie D. and Eswar S. Prasad (2003). "Medium-term determinants of current accounts in industrial and developing countries: An empirical exploration." *Journal of International Economics*, 59(1), 47–76.

Christodoulopoulou, Styliani and Olegs Tkacevs (2016). "Measuring the effectiveness of cost and price competitiveness in external rebalancing of euro area countries: What do alternative HCIs tell us?" *Empirica*, 43(3), 487–531.

Fabiani, Silvia, Stefano Federico, and Alberto Felettigh (2016). "Adjusting the external adjustment: Cyclical factors and the Italian current account." Questioni di Economia e Finanza (Occasional Papers) 346, Bank of Italy, Economic Research and International Relations Area.

Faruqee, Hamid and Guy Debelle (1996). "What Determines the Current Account? A Cross-Sectional and Panel Approach." IMF Working Papers 346, International Monetary Fund.

Gruber, Joseph W. and Steven B. Kamin (2005). "Explaining the global pattern of current account imbalances." International Finance Discussion Papers 846, Board of Governors of the Federal Reserve System.

Haltmaier, Jane (2014). "Cyclically Adjusted Current Account Balances." International Finance Discussion Papers 1126, Board of Governors of the Federal Reserve System.

Kara, A. Hakan and Cagri Sarikaya (2013). "Cyclically Adjusted Current Account Balance." Research Notes 18, Research and Monetary Policy Department, Central Bank of the Republic of Turkey.

Milesi-Ferretti, Gian M and Olivier J Blanchard (2011). "(Why) Should Current Account Balances Be Reduced?" IMF Staff Discussion Notes 3, International Monetary Fund.

Obstfeld, Maurice and Kenneth Rogoff (1995). "The intertemporal approach to the current account." In *Handbook of International Economics*, *Handbook of International Economics*, vol. 3, edited by G. M. Grossman and K. Rogoff, chap. 34, pp. 1731–1799. Elsevier.

Phillips, Steven T, Luis Catão, Luca A Ricci, Rudolfs Bems, Mitali Das, Julian Di Giovanni, Filiz D Unsal, Marola Castillo, Jungjin Lee, Jair Rodriguez, and Mauricio Vargas (2013). "The External Balance Assessment (EBA) Methodology." IMF Working Papers 272, International Monetary Fund.

Sachs, Jeffrey D. (1981). "The Current Account and Macroeconomic Adjustment in the 1970s." *Brookings Papers on Economic Activity*, 12(1), 201–282.

Salto, Matteo and Alessandro Turrini (2010). "Comparing alternative methodologies for real exchange rate assessment." European Economy - Economic Papers 427, European Commission.

Sastre, Teresa and Francesca Viani (2014). "Countries' safety and competitiveness, and the estimation of current account misalignments." Working Papers 1401, Banco de España.

Tamara, Tinti (2016). *External sustainability analysis*. Master's thesis, Nova School of Business and Economics, Portugal.

Wu, Yi (2008). "Growth, Expansion of Markets, and Income Elasticities in World Trade." *Review of International Economics*, 16(4), 654–671.

Appendix A

A.1. Elasticity of imports to the Imported Adjusted Demand (IAD)

Bussière *et al.* (2013) show that the total import content of an expenditure component, assuming S sectors and v final demand components in the economy and that the output of each sector is used both as an intermediate and to satisfy final demand, can be defined as:

$$\omega_v = \frac{uM_v^{ind} + uM_v^{dir}}{uF_v^d + uF_v^m} = \frac{uA^m(1-A^d)^{-1}F_v^d + uF_v^m}{uF_v^d + uF_v^m} \tag{A.1}$$

where u is a $1 \times S$ vector with all elements equal to one and the subscript v selects the v^th column of each matrix corresponding to the expenditure components of interest. $(1-A^d)^{-1}$ stands for the usual Leontief inverse, A^d is an $S \times S$ matrix of domestic input coefficients, A^m is the $S \times S$ matrix of imported input coefficients, F^d is the matrix of final demands of domestic goods and services and the direct imports are given by the $S \times V$ matrix, $F^m = M^{dir}$. Therefore, ω_v allows us to capture the IAD aggregate to be used in the regressions.

The estimation of the IAD elasticity follows the theoretical underpinnings of some empirical trade literature, notably the CES demand system. Under CES preferences, the logarithm of import demand is determined by:

$$lnM_t = lnD_t + \beta_p lnP_{M,t} \tag{A.2}$$

where D_t is aggregate demand (a CES aggregation of domestic and imported goods) and $P_{M,t}$ is the relative import price. This equation is estimated in first differences either for a panel of countries or for each country separately to obtain the elasticities of imports. However, standard measures of aggregate demand are replaced with IAD. Therefore:

$$\Delta lnM_{k,t} = \sum_{l=0}^{L} \beta_{IAD,l} \Delta lnIAD_{k,t-l} + \sum_{l=0}^{L} \beta_{P,l} \Delta lnP_{M,k,t-l} + \sum_{l=1}^{L} \beta_{M,l} \Delta lnM_{k,t-l} + \varepsilon_{k,t} \tag{A.3}$$

where k is a country, Δ denotes first differences and $\varepsilon_{k,t}$ is the error term. Applying the steady-state condition for a maximum of one lag we obtain:

$$\Delta lnM_{k,T} = \frac{\widehat{\beta}_{IAD,0} + \widehat{\beta}_{IAD,1}}{(1-\widehat{\beta}_{M,1})} \Delta lnIAD_{k,T} + \frac{\widehat{\beta}_{P,0} + \widehat{\beta}_{P,1}}{(1-\widehat{\beta}_{M,1})} \Delta lnP_{M,k,T} \tag{A.4}$$

Table A.1 presents the results of the regression estimated for Portugal, which leads to $\theta_M^{IAD} = 1.48$. It should be noted that coefficients for prices are not statistically significant at a level of 10 percent.

$M_{k,t}$	Coef.	Std. Error	t	P-value
$M_{k,t-1}$	-0.343	0.123	-2.79	0.007
$IAD_{k,t}$	1.381	0.122	11.32	0.00
$IAD_{k,t-1}$	0.61	0.209	2.92	0.00
$P_{M,k,t}$	0.003	0.108	0.26	0.798
$P_{M,k,t-1}$	0.123	0.107	1.15	0.254
R^2=71.1	Number of periods=63		F(5,57)=31.52	

TABLE A.1. Import elasticity estimates for Portugal

A.2. Elasticity of home exports to foreign GDP

The long-run elasticity of home exports to foreign GDP is assumed to be equal to the long-run elasticity of imports to GDP in the cross-country panel regression. It requires running the following panel regression:

$$\Delta lnM_{k,t} = \delta_k + \sum_{l=0}^{L} \beta_{GDP,l} \Delta lnGDP_{k,t-l} + \sum_{l=0}^{L} \beta_{P,l} \Delta lnP_{M,k,t-l} + \sum_{l=l}^{L} \beta_{M,l} \Delta lnM_{k,t-l} + \varepsilon_{k,t} \quad \text{(A.5)}$$

where k is a country, Δ denotes first differences, δ_k is the country fixed effects and $\varepsilon_{k,t}$ is the error term. Applying the steady-state condition for a maximum of one lag we obtain:

$$\Delta lnM_{k,T} = \frac{(\widehat{\beta}_{GDP,0} + \widehat{\beta}_{GDP,1})}{(1 - \widehat{\beta}_{M,1})} \Delta lnGDP_{k,T} + \frac{(\widehat{\beta}_{P,0} + \widehat{\beta}_{P,1})}{(1 - \widehat{\beta}_{M,1})} \Delta lnP_{M,k,T} \quad \text{(A.6)}$$

Table A.2 presents the results of the regression estimated for Portugal, which leads to $\theta_x = 2.6$. It should be noted that coefficients for lagged imports, prices and the constant are not statistically significant, at a 10 percent level.

A final note regards the extension of the methodology to the bilateral dimension. In all stages of the IAD computation and in the regression that estimates elasticity of imports, the conceptual approach is similar. This implies taking sub-blocks of the global input-output matrix and bilateral export and import flows.

$M_{k,t}$	Coef.	Std. Error	t	P-value
$M_{k,t-1}$	-0.061	0.046	-1.34	0.201
$GDP_{k,t}$	1.606	0.294	5.46	0.00
$GDP_{k,t-1}$	0.994	0.102	9.74	0.00
$P_{M,k,t}$	-0.190	0.078	-2.44	0.027
$P_{M,k,t-1}$	0.005	0.059	0.09	0.928
R^2=0.36	Number of obs. (17 countries)=1,071			F(5,16)=102.32

TABLE A.2. Exports elasticity estimates for Portugal

Appendix B: Observed and cyclically adjusted exports and imports

	Exports			Imports			Current account		
	Observed	Adjusted	Difference	Observed	Adjusted	Difference	Observed	Adjusted	Difference
1996	26.5	27.2	-0.6	33.7	34.7	-1.0	-4.5	-4.9	0.4
1997	27.1	27.4	-0.2	35.1	35.4	-0.2	-6.2	-6.2	0.0
1998	27.3	27.6	-0.3	36.5	36.0	0.5	-7.5	-6.8	-0.8
1999	26.5	26.3	0.2	36.8	35.9	0.9	-8.9	-8.2	-0.7
2000	28.2	27.2	1.0	39.3	38.1	1.2	-10.8	-10.7	-0.1
2001	27.4	26.9	0.5	37.6	36.8	0.8	-10.4	-10.1	-0.3
2002	26.9	26.9	0.0	35.2	34.8	0.4	-8.5	-8.2	-0.3
2003	26.8	27.1	-0.3	33.7	34.3	-0.7	-7.2	-7.5	0.4
2004	27.3	27.2	0.1	35.5	35.8	-0.3	-8.3	-8.7	0.3
2005	26.7	26.4	0.3	35.8	36.2	-0.4	-9.9	-10.6	0.7
2006	29.9	28.8	1.1	38.1	38.2	0.0	-10.7	-11.8	1.1
2007	31.0	29.1	1.9	38.7	37.5	1.1	-9.7	-10.5	0.8
2008	31.1	30.0	1.1	40.8	40.0	0.9	-12.1	-12.4	0.2
2009	27.1	29.2	-2.2	34.0	35.4	-1.4	-10.4	-9.6	-0.8
2010	29.9	31.2	-1.3	37.4	37.6	-0.2	-10.1	-9.0	-1.1
2011	34.3	35.3	-1.0	38.6	40.0	-1.4	-6.0	-6.4	0.4
2012	37.7	39.2	-1.5	38.2	41.6	-3.4	-1.8	-3.7	1.9
2013	39.5	41.3	-1.8	38.5	41.9	-3.4	1.6	0.0	1.6
2014	40.1	41.4	-1.4	39.9	43.0	-3.1	0.1	-1.6	1.7
2015	40.4	41.3	-0.9	39.8	41.7	-1.9	0.1	-0.9	1.0
2016	40.0	40.7	-0.7	38.9	40.3	-1.4	0.6	-0.1	0.7
2017	42.7	42.8	-0.1	41.9	42.6	-0.7	0.5	-0.1	0.5

TABLE B.1. Yearly observed and cyclically adjusted exports and imports as a percentage of GDP

Aggregate educational mismatches in the Portuguese labour market

Ana Catarina Pimenta
Banco de Portugal

Manuel Coutinho Pereira
Banco de Portugal

January 2019

Abstract
This article assesses the evolution of educational mismatches in Portugal over the last two decades. There has been a consistent reduction in undereducation, as younger and more educated generations replace the older ones in the labour market. Higher undereducation in Portugal vis-à-vis European Union countries is currently mostly a problem of older employees. Overeducation remains contained even in recent years: the figures for Portugal stand below those for the majority of the European countries. Furthermore, there has been a large increase in the number of college graduates during this period, most of whom have been able to find highly-skilled occupations. (JEL: I21, J21, J24)

Introduction

The Portuguese labour force has been structurally characterized by low levels of education relatively to other European countries. Despite still lagging behind, over the last decades the labour market in Portugal has undergone important transformations, with a considerable rise in employees' education levels. Alongside this trend, production technologies have changed in most industries, increasing the demand by firms for more educated workers. What has been the result of this interplay between a larger demand and supply of more educated employees?

Acknowledgements: The authors acknowledge the outstanding work of Mariana Tavares in the initial treatment of *Quadros de Pessoal*, in particular as regards the creation of the harmonized classification of occupations used in the study. The authors are also grateful for the help with the data provided by Marta Silva and Carlos Martins. Comments from Carlos Melo Gouveia, Nuno Alves, António Antunes e Miguel Gouveia are also acknowledged. The opinions expressed in this article are those of the authors and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the authors.

E-mail: apimenta@bportugal.pt; manuel.coutinho.pereira@bportugal.pt

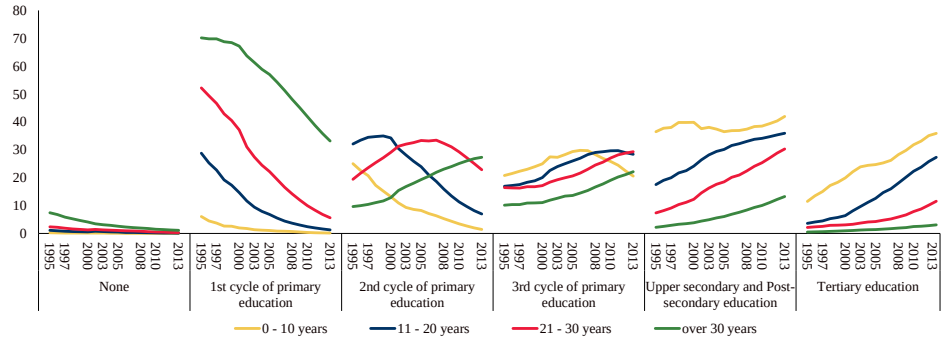


FIGURE 1: Employment structure by educational attainment within experience cohorts (in percentage).

Note: Lines represent the proportion of employees who, within a given experience cohort, have the indicated education level in that year. Details about the computation of breakdowns of education and experience are given in the text.
Source: Authors' calculations based on *Quadros de Pessoal*.

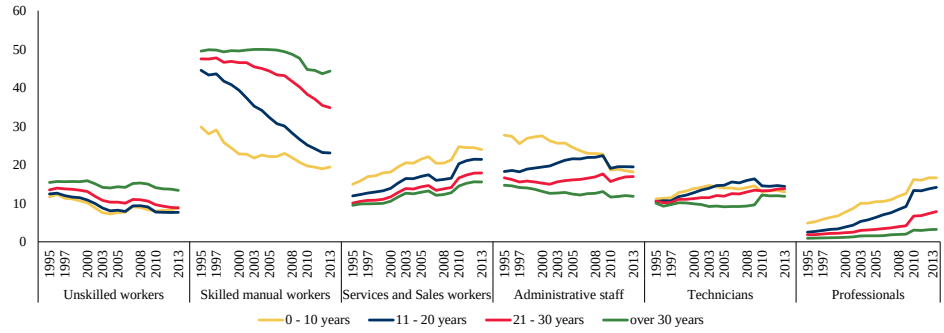


FIGURE 2: Employment structure by occupation within experience cohorts (in percentage).

Note: Lines represent the proportion of employees who, within a given experience cohort, have the indicated occupation in that year. Details about the computation of breakdowns of occupations and experience are given in the text.
Source: Authors' calculations based on *Quadros de Pessoal*.

Figures 1 and 2 present the evolution of employment structure, respectively, by education and occupation, 1995-2013, computed using microdata for the Portuguese labour market.[1] The upward trend in

1. The bulk of our results are based on *Quadros de Pessoal*, for which the last year available at the time of writing is 2013. Moreover, as further explained in the text, this article deals with secondary and tertiary market activity sectors.

educational attainment, along with a shift towards occupations requiring more skills, is rather evident in the charts for younger generations of workers. The purpose of this article is to explore the impact of such developments on educational mismatches, i.e. the lack or excess of education relatively to that required for workers to perform their jobs – labelled, respectively, as undereducation and overeducation. More specifically, we look at two main questions. First, to what extent undereducation is still an issue, given the catching up of the Portuguese workforce to higher education levels. Second, whether the growing number of highly educated workers, particularly college graduates, coming to the labour market has given rise to overeducation. This topic has been raised for developed economies by authors such as Hartog (2000) who considers that the strong expansion of participation in education has outpaced the increase in the demanded levels of education. In contrast to previous studies for Portugal – see below – that focused on the relationship between educational mismatches and earnings, our analysis focuses on the mismatches themselves (at an aggregate level), considering aspects such as breakdowns by occupations and experience cohorts, the relationship between overeducated college graduates and their fields of study and comparisons between Portugal and European Union countries.

Educational mismatches have been considered costly for the economies, justifying a long-standing researchers' concern about them (Freeman 1976; Thurow 1975). In the case of overeducation, output is lower than it would be if the workers' qualifications were fully used, while undereducated workers are likely to have a deficit of skills impacting negatively on their performance. Most studies of educational mismatches have focused on wage returns, both in the literature for Portugal (see, for instance, Araújo and Carneiro 2017; Cerejeira *et al.* 2007; Santos and de Oliveira 2002; Kiker *et al.* 1997) and in the international literature (see, for instance, Di Pietro and Urwin 2006; Duncan and Hoffman 1981; Iriondo and Pérez-Amaral 2016; Bauer 2002; Frenette 2004). In general, findings indicate that the returns to required education exceed those to overeducation, while the returns to undereducation are negative. Therefore, educational mismatches may also indirectly impact on firm productivity through their effects on wages and, more generally, job satisfaction (Hartog 2000). Mahy *et al.* (2015) provide evidence on the relationship between educational mismatches and firm productivity across working environments. They concluded that higher levels of required education or overeducation (undereducation) impact positively (negatively) on productivity. Furthermore, the effect of overeducation on productivity is higher among firms with highly-skilled jobs, belonging to high-tech industries and operating in a more uncertain economic environment.

There is no single best definition of required education for a given occupation in the literature while, at the same time, the measured levels of under- and overeducation turn out to be rather sensitive to such a definition. As described below, our study uses two common measures of

required education. The first one is a standard correspondence between the International Standard Classification of Occupations (ISCO) and the International Standard Classification of Education (ISCED). This measure shows a consistent reduction of undereducation over time in Portugal, from around two thirds of the employees in 1995 to approximately one third in 2013. Moreover, undereducation shows a marked decreasing profile from upper to lower experience cohorts. At the end of the sample, undereducation was below 10% for employees with up to 10 years of experience, but still at 60% for their peers with more than 30 years of experience. This pattern essentially reflects the replacement in the labour market of older generations by newer, more educated ones. Overeducation assumes negligible values at the beginning of the sample, and remains confined to around 5% of the workforce at the end of it.

We compute an alternative indicator, internal to the characteristics of Portuguese labour force, which measures required education directly from the data, as the mode of the employees' educational attainment within occupations. Undereducation is much lower and overeducation is higher in this second indicator, which is less demanding in terms of workers' education than the ISCO-ISCED one, particularly at the beginning of the sample. Moreover, because modal educational attainment categories may change over time, this second indicator shows no clear trends, thus being less appropriate for assessing the evolution of mismatches.

Finally, we present cross-country comparisons, on the basis of the EU-SILC database (2007-2016) and using the ISCO-ISCED indicator, which is internationally comparable. As far as undereducation is concerned, Portugal remains at a disadvantage vis-à-vis the European Union countries, but this is largely confined to more experienced workers. Moreover, Portugal is among the countries with the lowest incidence of overeducation.

An important caveat for the methodologies as the ones followed in this article is that required qualifications are evaluated solely on the basis of formal education, while other skills coming from broad-based knowledge, on-the-job training and experience might be equally important for workers to adequately perform their jobs.

This article is organized as follows. The following section briefly summarizes the methodologies in the literature to measure required education for a given occupation. The third section describes the data used and the computation of mismatches. The fourth section presents the key findings, dealing particularly with overeducated college graduates. The fifth section compares educational mismatches in Portugal and European Union countries. The last section concludes.

Methods for identifying educational mismatches

The starting point for identifying educational mismatches is a measure of required education for each occupation. In the educational mismatches literature two major approaches have been used. In the first one, studies have considered mismatches between individuals available for work and available jobs (macroeconomic mismatch). In the second one, the literature has considered discrepancies between workers' qualifications and requirements for their job only for employed workers, at a micro level. Moreover, educational mismatches may encompass both vertical mismatches, measured in terms of over- and undereducation, and horizontal mismatches. The latter measure the extent to which workers, normally college graduates, are employed in occupations unrelated to their main field of study, on the basis of a subjective question or comparing fields of study with occupation codes. Our study takes a micro approach and focuses on vertical mismatches.

There are three methodologies to measuring the required education for a certain occupation: the job evaluation method, the empirical method and the subjective method. Each of them has advantages and shortcomings while findings tend to differ depending on the one being used (McGuinness *et al.* 2017; Mysíková 2016).

The job evaluation method relates educational attainment and job qualifications based on an external definition of education requirements by job analysts (see, for instance, Ortiz and Kucel 2008). This methodology is perceived as quite accurate because it is based on field expertise, but it does not take into account that occupational requirements can change rapidly over time. The first measure of required education we use – a correspondence between major occupations of ISCO and skill levels as defined in ISCED – belongs to this category (see the next section).

The empirical – or realized matches – method estimates the level of required education on the basis of a statistical indicator computed by occupation, such as average years of schooling within a range of one standard deviation, or the modal educational achievement category (see, for instance, Bauer 2002; Cerejeira *et al.* 2007; Verdugo and Verdugo 1989; Kiker *et al.* 1997; Rahona-López and Pérez-Esparrells 2013; Iriondo and Pérez-Amaral 2016). In practice, the mode is used more frequently than the mean, as it is less sensitive to the presence of outliers and provides a more accurate measure of adequate education (Santos and de Oliveira 2002). Furthermore, the number of years of schooling used to calculate the mean is frequently upwardly biased, and the one-standard-deviation range is arbitrary (see, for instance, Ortiz and Kucel 2008; Mysíková 2016).

The key advantages of the empirical method are that it can be applied to any micro datasets containing information about both educational attainment and occupations and it is sensitive to technological changes and labour market characteristics. Nevertheless, the fact that such indicator is tied

to a given point in time and country should be taken into account, in particular, when workers' education levels are changing substantially or heterogeneous countries are being compared. We also present results for educational mismatches in Portugal on the basis of this methodology in order to complement those from the job evaluation method.

The two methodologies just described are the so-called objective ones. It is worth mentioning the subjective method (not used in this article) that is based on the worker's self-assessment of the level of education required to perform her/his job, which is then compared to the highest education level completed by the worker (see, for instance, Allen and Van der Velden 2001; Capsada-Munsech 2015; Di Pietro and Urwin 2006; Duncan and Hoffman 1981; Hersch 1991). The main advantages of this self-declared approach is that it is relatively easy to apply and it is job-specific. Its drawbacks include the subjective bias arising from the workers' tendency to overestimate their own qualifications (Groot and Van Den Brink 2000).

Data and computation of educational mismatches

Data

Our analysis is mainly based on *Quadros de Pessoal*, a longitudinal matched employer-employee database collected every year by the Ministry of Labour and Social Solidarity. The design of *Quadros de Pessoal* allows the identification of required education only through the objective approaches: job evaluation and realized matches. This database provides detailed information about firms with at least one employee, establishments and their workers. The two key variables for this article are employee's occupation and education in a given year. Thus only employees for whom such an information was available were considered. We used data covering the 1995-2013 period (except 2001 for which no data exist), for the subsample of full-time employees aged between 16 and 65 years.[2] As far as activity sectors are concerned, we confined our sample to secondary and tertiary market activities: manufacturing industry,

2. According to BPLim (2017), full-time work corresponds to a weekly work period over 75% of the normal work period at the establishment or firm.

energy, utilities, construction and market services.[3] Our database contains 23,415,079 observations in total.[4]

Workers' occupations were reported in *Quadros de Pessoal* during our sample period according to two classifications: the 1994 National Classification of Occupations and the 2010 Portuguese Classification of Occupations. In order to analyse the evolution of educational mismatches over time, we had to harmonize and aggregate these two classifications. We developed a harmonized breakdown comprising 26 occupations, as well as a more aggregated one that comprises 6 occupations (Table 1). As explained below, this latter breakdown matches the major groups of ISCO. We dropped workers in managerial occupations from the analysis on the account that for this group formal education is a less suitable approximation of required qualifications, as other factors such as experience and managerial skills play a very important role along with purely technical skills.

In *Quadros de Pessoal*, employees' educational attainment is a categorical variable reporting the highest level completed. The categories of this variable changed in 1994, 2000 and 2006 to accommodate the evolution of the Portuguese education system. The 2006 classification incorporates, in particular, the implementation of Bologna Process.[5] Moreover, until 2006 the education level was censored at the level of *Bacharelato*, i.e. workers with *Bacharelato*, *Licenciatura* or above were allocated to the same category. In order to make all the information comparable, we considered 6 levels of education throughout the sample: (i) none; (ii) 1st cycle of primary education; (iii) 2nd cycle of primary education; (iv) 3rd cycle of primary education; (v) upper secondary and post-secondary education and (vi) tertiary education. Furthermore, the data provided by *Quadros de Pessoal* include the field of study for employees with tertiary education, an information also used in this study.

Finally, using a common procedure in the literature, we calculated experience as the difference between age (derived from date of birth) of the employee and the number of years of schooling[6], minus 6. This variable

3. More specifically, our analysis includes firms whose main economic activity (NACE) is: (i) manufacturing; (ii) electricity, gas, steam, cold and hot water and cold air; (iii) water collection, treatment and distribution, sewerage, waste management and remediation activities; (iv) construction; (v) wholesale and retail trade, repair of motor vehicles and motorcycles; (vi) transportation and storage; (vii) accommodation and food service activities; (viii) information and communication activities; (ix) financial and insurance activities; (x) real estate activities; (xi) consultancy, scientific and technical activities and (xii) administrative and support service.

4. Given the aggregate nature of the analysis, we considered a sequence of cross-sections. Nevertheless, we took advantage of the longitudinal nature of the dataset to correct some inconsistencies in the data, such as in employees' educational attainment and date of birth.

5. For more information, see Decree-Law n. 74/2006 of March 24.

6. We assumed the minimum number of school years required to complete the highest educational level reported.

6-occupation breakdown	**26-occupation breakdown**
Unskilled workers	Agricultural and fishing workers Unskilled workers in other sectors
Skilled manual workers	Machinery and transport operators Skilled manual workers (except agriculture)
Services and Sales workers	Personal service workers Sellers
Administrative staff	Administrative staff
Technicians	Intermediate technicians for research and industry Intermediate technicians of electronics and computer science Intermediate life and health technicians Intermediate management and administration technicians Intermediate technicians for other services
Professionals	Experts in physics, chemistry and similar Experts in mathematics and statistics Computer experts Engineering experts Experts in life sciences Doctors Nurses University professors Teachers and childhood educators Management and administration Jurists Economists Journalists Specialists in social and human sciences

TABLE 1. Breakdowns of occupations used in the analysis.

corresponds therefore to *potential* labour market experience. We excluded from the sample workers having more than 55 years of potential experience.

We also carry out a comparison of educational mismatches for Portugal vis-à-vis other European Union countries and, in this instance, we use data from the Survey of Income and Living Conditions (EU-SILC). This last dataset is described in the corresponding section.

Computation of educational mismatches

ISCO provides a framework for the production of internationally comparable occupational data. The first method we employ for the calculation of under- and overeducation indicators is based on a standard correspondence between the eight ISCO-08 major groups relevant for our analysis and formal education levels of UNESCO's ISCED-97, given in Table 2, which was developed by the International Labour Organization (ILO 2012). Note that ISCO-08 has two additional major groups that are not relevant for us: Managers, excluded from the analysis for the reasons given above, and Armed forces, as we only deal with market activities. The 6-occupation breakdown presented in Table 1 above is a slightly aggregated version of those eight ISCO-08 major groups, as detailed in Table 3.

ISCO-08	Required education (ISCED-97)
Elementary Occupations Skilled Agricultural, Forestry and Fishery Workers	1st, 2nd or 3rd cycle of primary education or Upper secondary and Post-secondary education
Services and Sales Workers Craft and Related Trades Workers Plant and Machine Operators and Assemblers Clerical Support Workers	3rd cycle of primary education or Upper secondary and Post-secondary education
Technicians and Associate Professionals Professionals	Tertiary Education

TABLE 2. Mapping of ISCO-08 major groups to levels of education of ISCED-97.

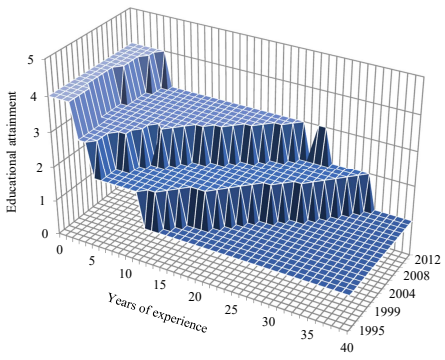
6-occupation breakdown	ISCO-08
Unskilled workers	Elementary Occupations Skilled Agricultural, Forestry and Fishery Workers
Services and Sales Workers	Services and Sales Workers
Skilled manual workers	Craft and Related Trades Workers Plant and Machine Operators and Assemblers
Administrative staff Technicians Professionals	Clerical Support Workers Technicians and Associate Professionals Professionals

TABLE 3. Mapping of the 6-occupation breakdown to ISCO-08 major groups.

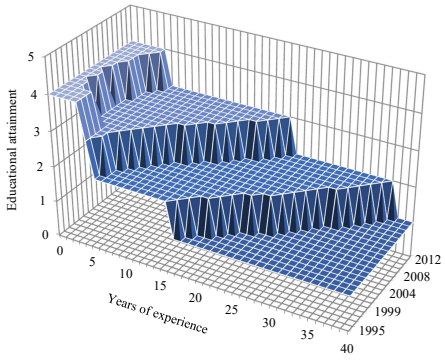
We thought it appropriate to present an alternative measure of educational mismatches, calculated from a country-specific indicator capturing the dynamics in the qualifications of the Portuguese workforce. We determined required education directly from *Quadros de Pessoal*, using the mode as the relevant statistic and taking as a reference both breakdowns of occupations in Table 1.

When one considers the evolution of modal educational attainment by occupation, as a rule this has changed over time, and often the difference in the proportion of employees who have the «first» and the «second» modes is reduced. Such changes in the modal educational attainment mostly occur as younger generations of workers replace older ones. Figure 3 illustrates this issue, presenting the evolution of the modal educational attainment category (1995-2013) for each of the 6 aggregated occupations, *by year of potential experience*. For most occupational categories, there are instances where two levels of educational attainment assume an important role. The exception is Professionals who uniformly possess tertiary education throughout the sample.

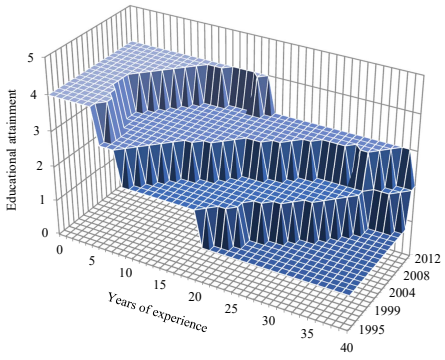
Therefore, in the alternative indicator, we defined required education as the two levels of education with the highest number of employees, i.e. the first and the second mode, calculated within each occupation for a given year, except for Professionals, for whom only the first mode was taken. Furthermore, when the modal qualifications were not contiguous, the in-between level of education was also considered as required education. This procedure also makes the computation of the mode indicator closer to that of the ISCO-ISCED indicator which comprises more than one educational category for most occupations (except for Technicians and Professionals).



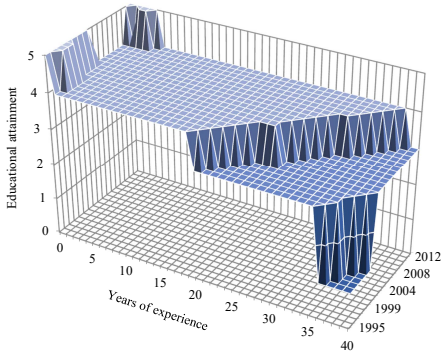
(A) Unskilled workers



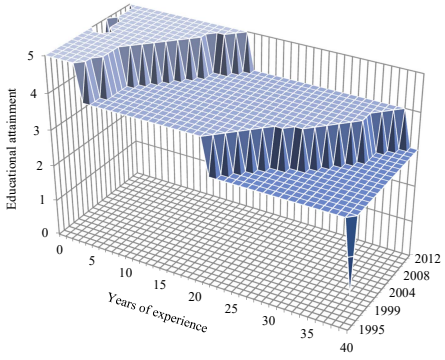
(B) Skilled manual workers



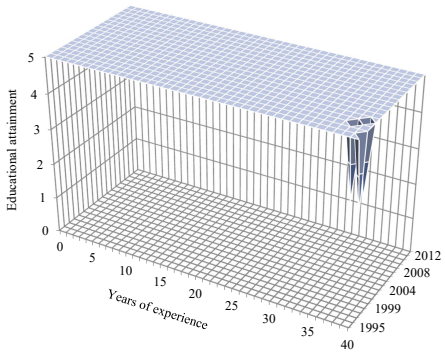
(C) Services and Sales workers



(D) Administrative staff



(E) Technicians



(F) Professionals

FIGURE 3: Evolution of employees' modal education by occupation and years of experience.

Note: Educational attainment: 0=None, 1=1st cycle of primary education, 2=2nd cycle of primary education, 3=3rd cycle of primary education, 4=Upper secondary and Post-secondary education and 5=Tertiary education.
Source: Authors' calculations based on *Quadros de Pessoal* (1995-2013).

Evolution of educational mismatches in the Portuguese labour market

Table 4 shows the evolution (1995-2013) of overall educational mismatches according to both ISCO-ISCED and mode indicators. The mode indicator was computed taking as a reference the two breakdowns of occupations in Table 1 – the more aggregated one being also used for the computation of the ISCO-ISCED indicator.

	1995	1997	1999	2002	2005	2007	2009	2011	2013
ISCO-ISCED indicator									
Required education	34.6	37.3	39.3	43.2	47.7	51.6	54.9	57.7	60.0
Overeducation	0.9	1.3	1.6	2.1	2.8	3.2	3.8	4.3	5.1
Undereducation	64.6	61.5	59.1	54.7	49.6	45.2	41.3	38.0	35.0
Mode indicator									
6-occupation breakdown									
Required education	76.8	75.4	74.2	71.6	71.7	69.8	71.4	69.8	70.7
Overeducation	11.3	14.6	14.8	19.0	17.3	20.8	20.1	10.5	9.0
Undereducation	11.9	10.0	11.0	9.4	11.0	9.5	8.6	19.7	20.3
26-occupation breakdown									
Required education	76.4	74.9	73.5	73.9	67.9	66.6	68.7	67.8	68.7
Overeducation	11.3	14.6	16.8	17.8	15.6	19.0	17.9	9.8	8.3
Undereducation	12.3	10.5	9.7	8.3	16.5	14.4	13.4	22.4	23.0

TABLE 4. Evolution of overall educational mismatches according to each indicator.

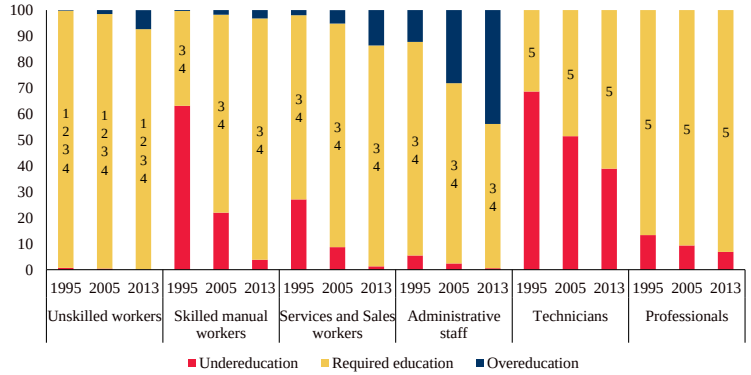
Source: Authors' calculations based on *Quadros de Pessoal* (1995-2013).

ISCO-ISCED indicator

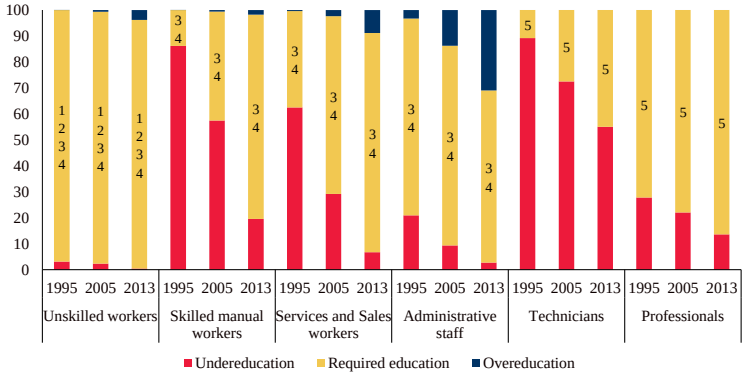
Figure 4 presents the breakdown by occupations and potential experience cohorts for the ISCO-ISCED indicator. We selected the first, an intermediate and the last year of the sample to simplify the presentation of results. The indicator shows a consistent reduction of undereducation over time, from around 2/3 of the employees in 1995 to approximately 1/3 in 2013 (Table 4). This trend is common to all four potential experience cohorts. However, the *level* of undereducation varies substantially across such cohorts (Figure 4). It stood at around 80% for the cohort of employees with more than 30 years of experience, in the mid-90s, remaining at 60% at the end of the sample (affecting in 2013 particularly the skilled manual workers and technicians). In contrast, in the lowest experience cohort, it came down from approximately 33% at the beginning of the sample to 7% at the end of it (being in 2013 largely confined to technicians). The decreasing profile of undereducation over time within potential experience cohorts reflects the gradual replacement of older, less educated, generations in the labour market. Such a reduction reflects the approximation of Portuguese workforce education to European

standards. Nevertheless, other factors may play a role in the results across experience cohorts: for instance, undereducated workers at the beginning of their employment spell may get further education and thus achieve the required attainment level. This would tend to reduce overall undereducation within a given generation of workers as they move up in the distribution of potential experience. However, these factors are better studied by following individual workers' behaviour over time rather than in an aggregate analysis such as here.

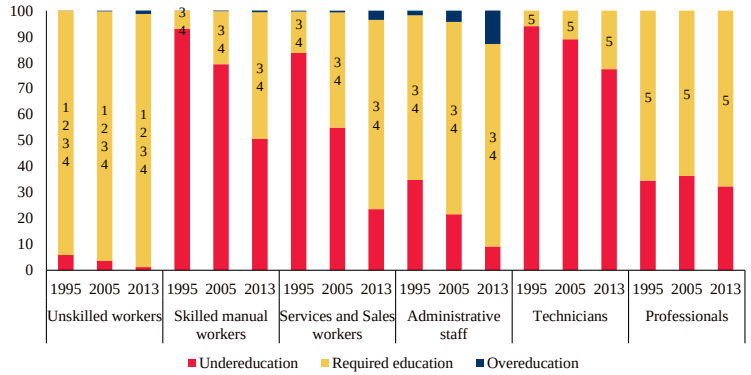
In contrast, overeducation remains a rather unimportant phenomenon throughout, rising from negligible values at the beginning of the sample to around 5% at the end. Overeducation is higher for employees with up to 10 years of experience, increasing from 4% in 1995 to 12% in 2013. For more experienced cohorts, the proportion of overeducated employees is still negligible at the end of the sample. Such a tendency is in line with the coming to the labour market of more educated workers, particularly with Tertiary education, who are not able to find an occupation matching their formal education level. We will come back to this issue later on. Similarly, other factors may influence the differences across experience cohorts towards an attenuation of mismatches as workers become more experienced. Some authors pointed out that overeducation may arise from a trade-off between education and other components of human capital (such as Araújo and Carneiro 2017; Cerejeira *et al.* 2007; Kiker *et al.* 1997; Sicherman 1991). Thus, overeducated employees may substitute education by the lack of previous job experience, accepting jobs requiring less education than they actually have in order to acquire the required experience for job mobility.



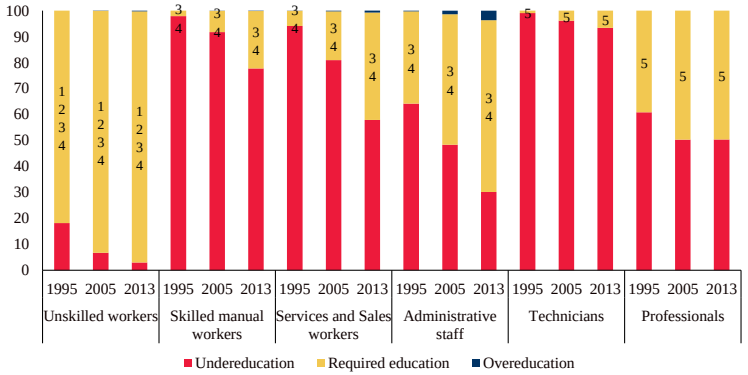
(A) Cohort 0-10 years of experience



(B) Cohort 11-20 years of experience



(C) Cohort 21-30 years of experience



(D) Cohort over 30 years of experience

FIGURE 4: Evolution of educational mismatches by occupation and experience cohorts using ISCO-ISCED indicator (in percentage).

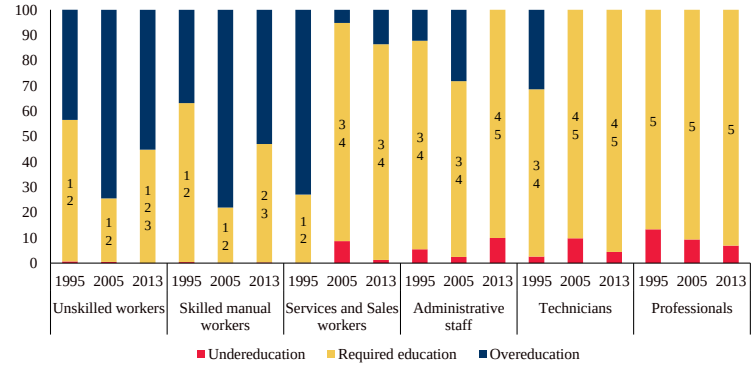
Note: Numbers on the bars refer to the education required for each occupation in each year: 0=None, 1=1st cycle of primary education, 2=2nd cycle of primary education, 3=3rd cycle of primary education, 4=Upper secondary and Post-secondary education and 5=Tertiary education.
Source: Authors' calculations based on *Quadros de Pessoal* (1995-2013).

Mode indicator

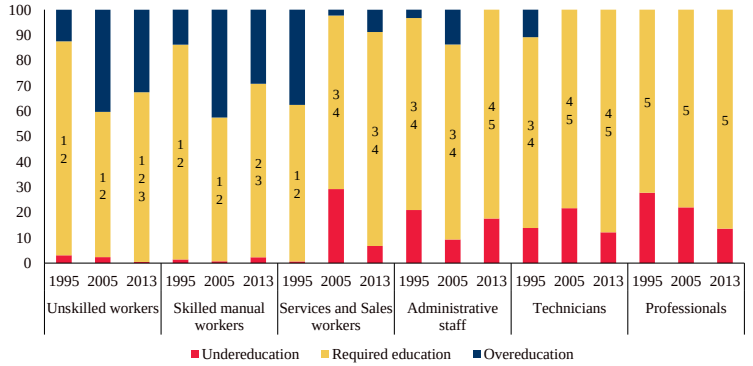
The measured levels of under- and overeducation for the mode indicator on the basis of each of the two occupational breakdowns considered are very close, while differing substantially from those on the basis of the ISCO-ISCED correspondence both as to the level and trend (Table 4). The ISCO-ISCED indicator is generally more demanding in terms of workers' qualifications, given that it takes as a reference up-to-date international standards (ILO 2012). In contrast, the mode indicator reflects the characteristics of the Portuguese workforce, and partly accommodates the gap to international standards by a lower level of required education for some occupations. As a consequence, there is a striking difference in the percentage of undereducation, lower in mode indicator particularly at the beginning of the sample, while overeducation is higher (Araújo and Carneiro 2017, also using *Quadros de Pessoal*, obtained results in line with ours).

Figure 5 presents the breakdowns by occupations and experience cohorts for the mode indicator, considering the more aggregated occupational breakdown, also showing the respective required education levels (matching Figure 4 above). Required education is lower than in the ISCO-ISCED correspondence at the beginning of the sample for Skilled manual workers and Technicians. For other occupations, such as Administrative staff and Professionals, required education either coincides in both methodologies or, in the case of Administrative staff, is higher at the end of the sample, reflecting the increasing weight of college graduates in that occupation. It is worth noting that the larger amount of undereducation, and smaller amount of overeducation, in the upper vis-à-vis the lower cohorts of potential experience holds for both methodologies.

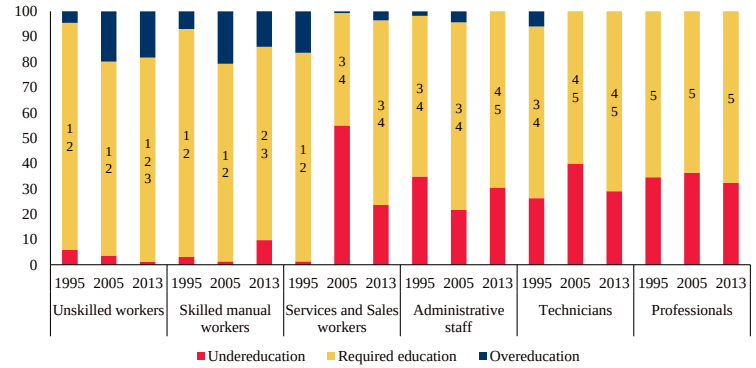
Unlike the ISCO-ISCED indicator, for the mode indicator the overall under- and overeducation figures do not show marked trends, because the modal educational attainment categories change over time for some occupations. This renders the indicator less appropriate for assessing the evolution of these phenomena over time. In particular, undereducation decreases slightly in the initial period after 1995, denoting improvements in the workforce qualifications, but this eventually leads to an upward revision in the modal categories, and to a hike in undereducation. Such was the case of Service and sale workers, from 2002 onwards, and Skilled manual workers, from 2010 onwards. The trajectory of overeducation is affected in an analogous way. It shows a rising trend at the beginning of the sample, as younger workers possessing intermediate education enter into low-skilled occupations, but it then jumps down when such workers become the group with modal education level.



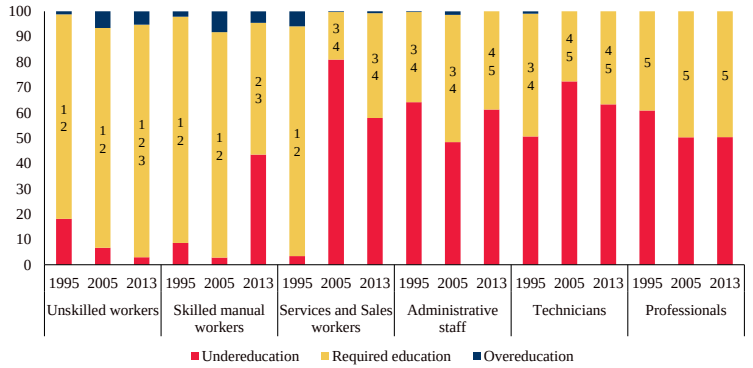
(A) Cohort 0-10 years of experience



(B) Cohort 11-20 years of experience



(C) Cohort 21-30 years of experience



(D) Cohort over 30 years of experience

FIGURE 5: Evolution of educational mismatches by occupation and experience cohorts using mode indicator (in percentage).

Note: Numbers on the bars refer to the education required for each occupation in each year: 0=None, 1=1st cycle of primary education, 2=2nd cycle of primary education, 3=3rd cycle of primary education, 4=Upper secondary and Post-secondary education and 5=Tertiary education.
Source: Authors' calculations based on *Quadros de Pessoal* (1995-2013).

Overeducated college graduates

The rise in the proportion of graduates in the labour force has been a prominent development in the last decades, as seen in Figure 1. This section looks more in detail into how the occupational structure of the Portuguese labour market has accompanied such a development. Figure 6 shows that the proportion of college graduates increased substantially between 1995 and 2013, in the considered activities, from about 3 to 16%. These employees have been mainly hired by the service sector, whose weight in employment of college graduates has gone up, matching a decrease in the weight of manufacturing industry.

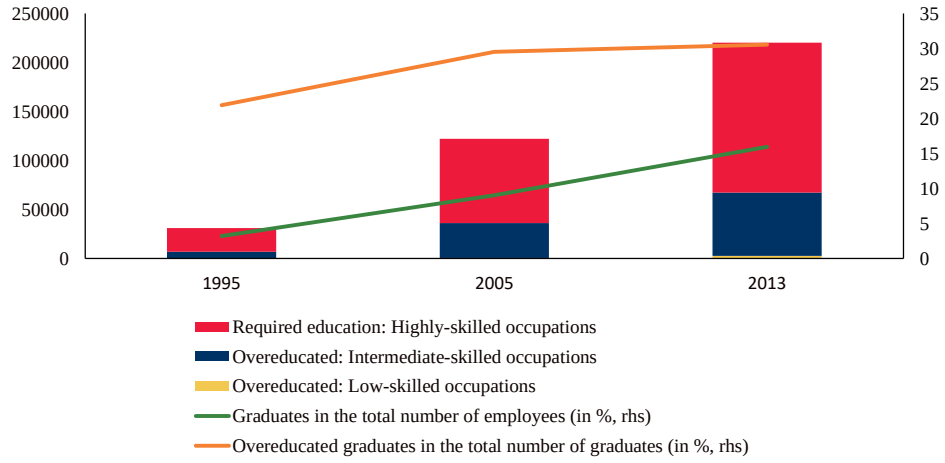


FIGURE 6: Occupations and educational mismatches of graduates.

Notes: Low-skilled occupations include Unskilled workers; intermediate-skilled occupations include Skilled manual workers, Services and Sales workers and Administrative staff; highly-skilled occupations include Technicians and Professionals. According to ISCO-ISCED correspondence (Tables 2 and 3), overeducated college graduates are those in low- and intermediate-skilled occupations.
Source: Authors' calculations based on *Quadros de Pessoal* (1995-2013).

The share of overeducated college graduates in the total number of college graduates has gone up in the first decade of the sample (from 20 to 30%) and then approximately stabilized. Additionally, the rising profile of overeducation within potential experience cohorts (Figure 7) signals that this has become more prevalent among the new generations of college graduates. Moreover, overeducated college graduates are predominantly in the service sector, performing administrative jobs.

Figure 7 also presents the breakdown by field of study of overeducated college graduates, who are mainly from Economics, Social Sciences and Law. These results are consistent with the literature. Capsada-Munsech

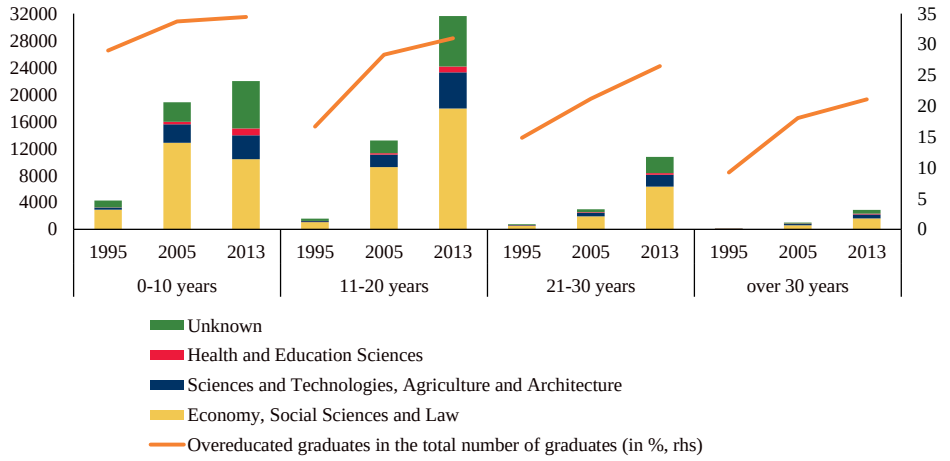


FIGURE 7: Evolution of overeducated graduates by field of study and experience cohort.

Notes: The field of study Health and Education Sciences includes Health, Education Sciences and Teacher Training. The field of study Sciences and Technologies, Agriculture and Architecture includes Sciences and Technologies, Agriculture and Natural Resources, Architecture, Design and Arts. The field of study Economy, Social Sciences and Law includes Economy, Management, Accounting, Social Sciences, Humanities, Services and Law.
Source: Authors' calculations based on *Quadros de Pessoal* (1995-2013).

(2015) found that fields of study where work performance is evaluated less objectively (such as, Humanities, Social Sciences and Economics) are associated with a higher probability of overeducation. In line with previous evidence, Dolton and Vignoles (2000) concluded that Social Sciences and Arts graduates are more likely to be overeducated relative to Engineers.

In spite of the increase in overeducation, the bulk of the incoming college graduates to the Portuguese labour market in the last two decades were able to find highly-skilled occupations, in which the economy should profit most from their qualifications. A caveat in this analysis is that we are looking at on-the-job mismatches and we do not observe the graduates who could not find a job (and perhaps were unemployed or emigrated), something that may be particularly important in the last years of the sample coinciding with the crisis.

Educational mismatches: Portugal vis-à-vis the European Union

We finalize this article by presenting an international comparison of educational mismatches based on harmonized microdata from the Survey of Income and Living Conditions (EU-SILC), covering a wide range of countries belonging to the EU, on an annual basis. We applied to this dataset the same selection criteria as for *Quadros de Pessoal*, focusing on full-time employees, aged between 16 and 65 years and with information about occupation and educational attainment. Moreover, we considered the same economic activities as described above. Our database contains data for Portugal plus 25 European Union countries, and we present results for 2007 (102,660 employees) and for the last year available, 2016 (94,617 employees).[7] All the results were calculated using the sample weights available (cross-sectional databases), which allow an extrapolation to the entire population.

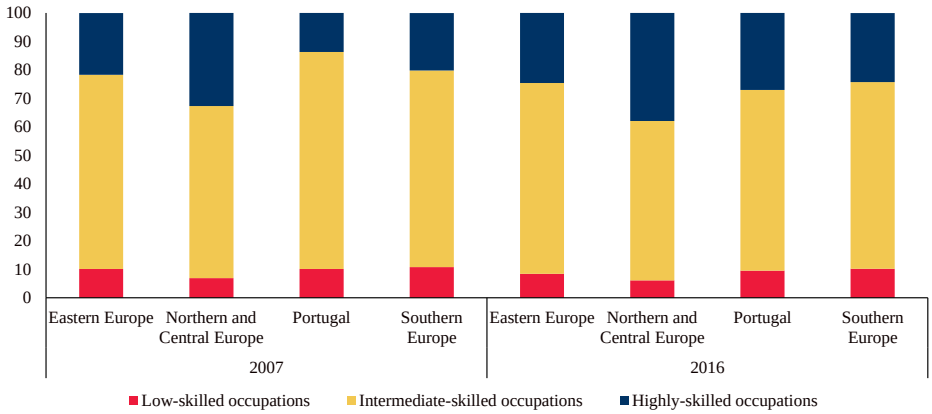


FIGURE 8: Evolution of employment structure by occupation for Portugal and three groups of EU countries (in percentage).

Notes: All results were weighted with sample weights. Northern and Central Europe includes Austria, Belgium, Finland, France, Germany, Ireland, Luxembourg, Netherlands, Denmark and Sweden; Southern Europe includes Cyprus, Greece, Italy, Malta and Spain; Eastern Europe includes Bulgaria, Czech Republic, Estonia, Latvia, Lithuania, Hungary, Poland, Romania, Slovenia and Slovakia.
Source: Authors' calculations based on EU-SILC.

Measured educational mismatches in a given country depend on the structure of occupations in the economy. More specifically, the overall figures are the average of mismatches by occupation weighted by that structure. Therefore we start with by showing the occupational structure in Portugal

7. For Ireland, Luxembourg, Italy and Malta, due to the unavailability of more recent data, 2015 data were used.

and in the remaining countries aggregated, for the sake of presentation, into three groups: Northern and Central Europe, Southern Europe and Eastern Europe (Figure 8). In general, in the last year of the sample, Portugal compares unfavourably only with Northern and Central Europe. There was a gap in the proportion of highly-skilled occupations in 2007 between Portugal and each of the groups, but this shrank in the ensuing decade. Within intermediate-skilled occupations, the skilled manual workers (not shown) were over-represented in Portugal vis-à-vis the three groups in 2007, making up half of the work force, both this proportion has come down to about 1/3 in 2016, a figure close to the ones for Southern and Eastern Europe.

Figure 9 presents the educational mismatches for all countries in the database, using the ISCO-ISCED methodology. Portugal was the country with the highest incidence of undereducation in both years considered, despite the reduction from 2007 to 2016. As regards overeducation, Portugal has an incidence below the EU average and, as shown below, this holds throughout experience cohorts.

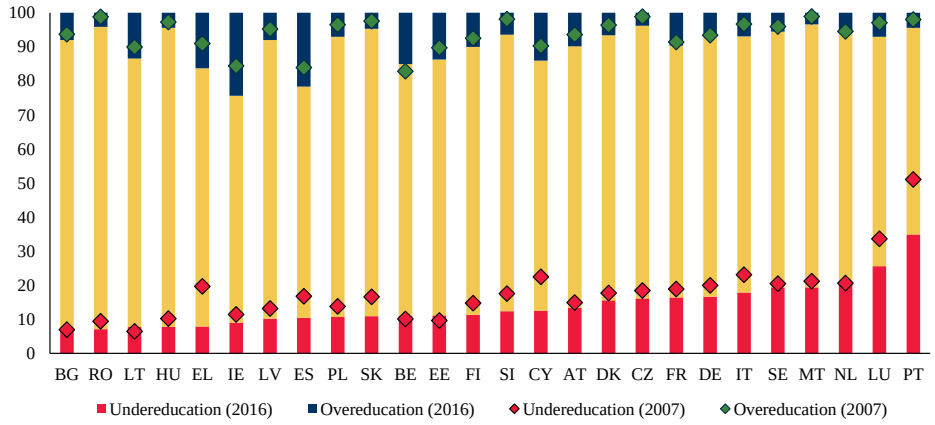


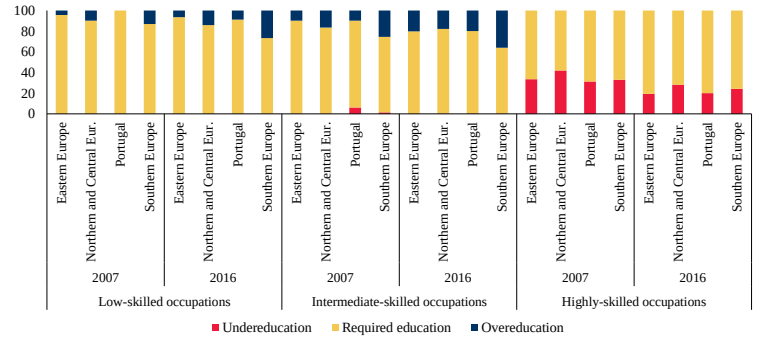
FIGURE 9: Evolution of educational mismatches by European Union country between 2007 and 2016 (in percentage).

Notes: All results were weighted with sample weights. The 26 countries represented are: Belgium (BE), Bulgaria (BG), Czech Republic (CZ), Denmark (DK), Germany (DE), Estonia (EE), Ireland (IE), Greece (EL), Spain (ES), France (FR), Italy (IT), Cyprus (CY), Latvia (LV), Lithuania (LT), Luxembourg (LU), Hungary (HU), Malta (MT), Netherlands (NL), Austria (AT), Poland (PL), Portugal (PT), Romania (RO), Slovenia (SI), Slovakia (SK), Finland (FI) and Sweden (SE).
Source: Authors' calculations based on EU-SILC.

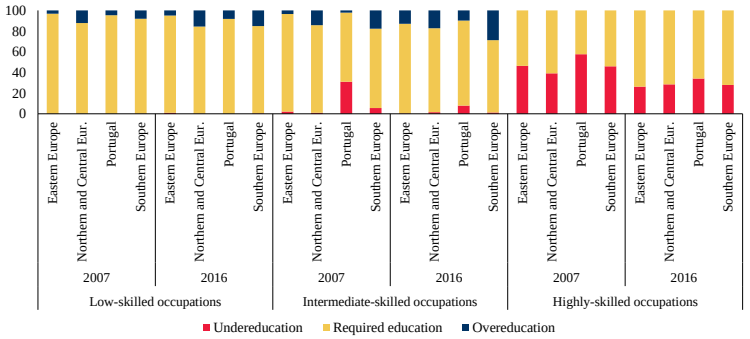
We carried out an additional exercise, at the country level, in order to separate out the contributions to the overall mismatches coming, respectively, from mismatches within occupations and from the occupational structure. To this end, we took a standard occupational structure given by the average for the countries considered, in 2007 and 2016. We replicated Figure 9, keeping the

occupational structure constant throughout countries, in each year. However, results do not significantly change, in particular for Portugal. This reflects the fact that the economies in the EU are relatively homogeneous in terms of occupations, as suggested by Figure 8.

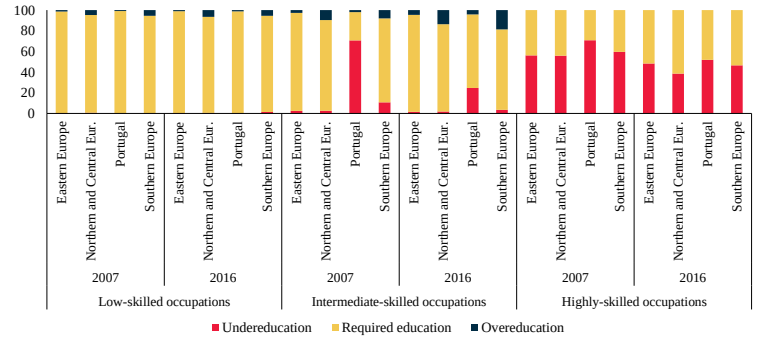
Figure 10 presents educational mismatches broken down by potential experience cohorts and occupations. The results show that the disadvantage of Portugal vis-à-vis the European Union mainly reflects the low levels of education of older generations in intermediate-skilled occupations. Such a disadvantage has, however, faded away considerably in the last two decades. It is worth noting that the Portuguese workers in low-skilled occupations also had lower levels of education than their European peers, but this does not show up as undereducation because the ISCO-ISCED correspondence considers as required education elementary attainment levels. Undereducation remains relatively high for younger workers in highly-skilled occupations (more specifically, for Technicians) in Portugal, but this is common to other European countries.



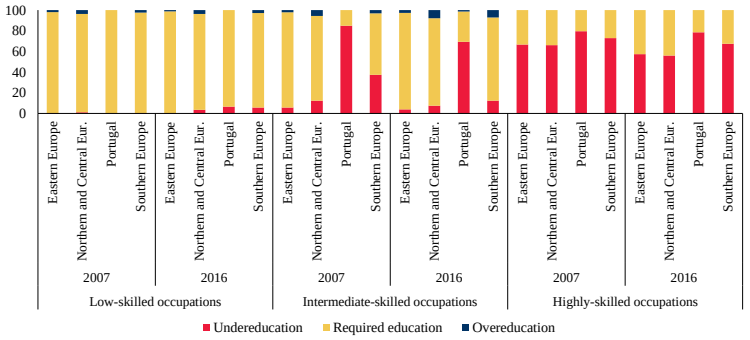
(A) Cohort 0-10 years of experience



(B) Cohort 11-20 years of experience



(C) Cohort 21-30 years of experience



(D) Cohort over 30 years of experience

FIGURE 10: Evolution of educational mismatches by occupation and experience cohorts: Portugal vs. European Union countries groups (in percentage).

Notes: All results are weighted by sample weights. See the note to Figure 6 for the breakdown of occupations and the note to Figure 8 for the composition of countries groups.
Source: Authors' calculations based on EU-SILC.

Concluding remarks

This article assessed the evolution of educational mismatches in Portugal over the last two decades, on the basis of two datasets: *Quadros de Pessoal*, for Portugal (1995-2013), and EU-SILC, for the European Union (2007 and 2016). In doing so, we used two methodologies, namely, a correspondence between ISCO and ISCED (belonging to the job evaluation methods) and the mode of worker's education within occupations (belonging to the empirical methods). Measured levels and trends of under- and overeducation are quite sensitive to the methodology used.

The ISCO-ISCED indicator lends itself better to a comparison across countries and over time. The catching up of education of the Portuguese workforce to higher levels has meant a considerable reduction of undereducation, according to this indicator. The approximation of the Portuguese workforce education level to international standards has implied that the disadvantage vis-à-vis the European Union – mainly associated with low levels of education in intermediate-skilled occupations – has largely faded away for younger generations. Some undereducation remains in highly-skilled occupations, an issue common to other European countries.

Overeducation is still of limited importance. In particular, the bulk of incoming college graduates coming to the Portuguese labour market in last two decades were able to find highly-skilled occupations, in which their qualifications in principle can be best put to use.

The figures for undereducation are lower, and those for overeducation higher, in the mode indicator. This indicator partly accommodates the gap to international standards by a lower level of required education for some occupations. Moreover, it does not show marked trends, as the modal educational attainment categories within some occupations change over time.

An important caveat for all this analysis is that formal education is an imperfect approximation of the qualifications needed for workers to adequately perform their jobs.

In terms of further research, two lines could be in particular pursued, at a micro level. A first one would be to ascertain the «typical» reaction of undereducated workers, in terms of getting further education, or overeducated workers, in terms of changing job, vis-à-vis comparable workers with required education. A second line would be to study the relationship between educational mismatches and productivity, at the firm level, which to the best of our knowledge has not yet been analysed for Portugal.

References

Allen, Jim and Rolf Van der Velden (2001). "Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search." *Oxford economic papers*, 53(3), 434–452.

Araújo, Isabel and Anabela Carneiro (2017). "Educational mismatches and wages: evidence from a matched employer/employee dataset." *In Portuguese Stata UGM (Sept 15)*.

Bauer, Thomas K (2002). "Educational mismatch and wages: a panel analysis." *Economics of education review*, 21(3), 221–229.

BPLim (2017). "Quadros de Pessoal (Personnel Records) 1982-2013: Data Manual." *Banco de Portugal*.

Capsada-Munsech, Queralt (2015). "The role of social origin and field of study on graduates' overeducation: the case of Italy." *Higher Education*, 69(5), 779–807.

Cerejeira, João, Miguel Portela, Carla Angélica da Silva Pinto Sá, Fernando Alexandre, *et al.* (2007). "Os fluxos e as remunerações de sub e sobre-escolarizados em Portugal no período 1995-2005." Tech. rep., Universidade do Minho. Núcleo de Investigação em Políticas Económicas (NIPE).

Di Pietro, Giorgio and Peter Urwin (2006). "Education and skills mismatch in the Italian graduate labour market." *Applied Economics*, 38(1), 79–93.

Dolton, Peter and Anna Vignoles (2000). "The incidence and effects of overeducation in the UK graduate labour market." *Economics of education review*, 19(2), 179–198.

Duncan, Greg J and Saul D Hoffman (1981). "The incidence and wage effects of overeducation." *Economics of education review*, 1(1), 75–86.

Freeman, Richard (1976). "The overeducated american." *Academic press*.

Frenette, Marc (2004). "The overqualified Canadian graduate: the role of the academic program in the incidence, persistence, and economic returns to overqualification." *Economics of Education Review*, 23(1), 29–45.

Groot, Wim and Henriette Maassen Van Den Brink (2000). "Overeducation in the labor market: a meta-analysis." *Economics of education review*, 19(2), 149–158.

Hartog, Joop (2000). "Over-education and earnings: where are we, where should we go?" *Economics of education review*, 19(2), 131–147.

Hersch, Joni (1991). "Education match and job match." *The Review of Economics and Statistics*, pp. 140–144.

ILO (2012). "International Standard Classification of Occupations: Structure, group definitions and correspondence tables." *International Labour Organization*.

Iriondo, Iñaki and Teodosio Pérez-Amaral (2016). "The effect of educational mismatch on wages in Europe." *Journal of Policy Modeling*, 38(2), 304–323.

Kiker, Billy Frazier, Maria C Santos, and M Mendes De Oliveira (1997). "Overeducation and undereducation: evidence for Portugal." *Economics of*

Education Review, 16(2), 111–125.

Mahy, Benoît, François Rycx, and Guillaume Vermeylen (2015). "Educational mismatch and firm productivity: Do skills, technology and uncertainty matter?" *De Economist*, 163(2), 233–262.

McGuinness, Seamus, Konstantinos Pouliakas, and Paul Redmond (2017). "How Useful is the Concept of Skills Mismatch?"

Mysíková, Martina (2016). "Educational mismatch in the Czech Labour Market." *Review of Economic Perspectives*, 16(2), 103–120.

Ortiz, Luis and Aleksander Kucel (2008). "Do fields of study matter for over-education? The cases of Spain and Germany." *International Journal of Comparative Sociology*, 49(4-5), 305–327.

Rahona-López, Marta and Carmen Pérez-Esparrells (2013). "Educational attainment and educational mismatch in the first employment in Spain." *ISRN Education*, 2013.

Santos, Maria Clementina and M Mendes de Oliveira (2002). *Qualification Requirements and Educational Attainment in Portugal, 1985-1997*. Comunicação apresentada em: Desenvolvimento Económico Português no Espaço Europeu: Determinantes e Políticas, Lisboa.

Sicherman, Nachum (1991). "Overeducation in the Labor Market." *Journal of labor Economics*, 9(2), 101–122.

Thurow, Lester C (1975). *Generating inequality*. Basic books.

Verdugo, Richard R and Naomi Turner Verdugo (1989). "The impact of surplus schooling on earnings: Some additional findings." *Journal of Human Resources*, pp. 629–643.

Economics Synopsis

Why is price stability a key goal of central banks?

Bernardino Adão
Banco de Portugal

January 2019

Central banks place great emphasis on maintaining low inflation. A primary objective of central banks in advanced economies is to keep inflation low and stable. This is often defined as inflation near 2% in the medium-run. Here we investigate if this primary objective is consistent with the theoretical literature[1]. We find that the literature confirms that a low and stable level for inflation is efficient.

When money is demanded for transaction purposes, and there is a complete set of tax instruments the optimal nominal interest rate is zero. This policy, also known as the Friedman rule, implies an optimal rate of inflation that is negative and equal in absolute value to the real rate of interest. A negative inflation has the extra advantage of promoting equity, as the inflation tax is a regressive tax. Moreover, a negative inflation can correspond to the growth that maximizes society's welfare. However, if the set of tax instruments is not complete then the Friedman rule may not be optimal. The optimal inflation level depends on which tax instrument is not available and on the particular model (and calibration considered). In many of these cases inflation is still low or negative. But there are reasons to have a positive inflation too. For instance, effective stabilization may require a safety margin, or positive inflation level, because nominal interest rates have a lower bound.

The remainder of the survey proceeds as follows. First, we describe the arguments for a low level of anticipated inflation and later we review the reasons for a stable inflation. We start by providing an explanation of the

Acknowledgements: This paper uses previous work done in co-authorships with André Silva, Isabel Correia, Sandra Gomes and Sofia Saldanha. I thank Nuno Alves, Isabel Correia, António Antunes and Miguel Gouveia, for valuable comments and suggestions. The opinions expressed in this article are those of the author and do not necessarily coincide with those of Banco de Portugal or the Eurosystem.

E-mail: badao@bportugal.pt

1. We leave outside of this survey the less conventional literature on search models of money. Typically, the optimal inflation rate in these type of models is low too and the costs of inflation even higher than in the standard models. See Lagos and Wright (2005).

Friedman rule. After, we show that the inflation prescribed by the Friedman rule in addition to being efficient promotes equity too. Poor households are better off when inflation is low. Next, we describe why restrictions in the tax system can invalidate the Friedman rule. These restrictions can take various forms and imply higher optimal inflation rates than the one associated with the Friedman rule. We look at a great number of them and conclude that they in general do not justify high inflation targets for the central banks. Later, we review the relationship between inflation and growth. Finally, we argue that surprises in inflation should be avoided.

Money demand

The most celebrated result in this literature is due to Milton Friedman (1969). Milton Friedman provided a simple rule for determining the optimal rate of inflation in the long run. He started with the observation that money provides valuable services, as it makes it easier and more convenient for consumers to do transactions. In many transactions the sellers accept both money and credit cards as payment for goods and services, but some only accept cash. A consumer could probably get by with credit cards alone, but this would be more cumbersome because it would imply spending more time seeking out sellers that accept them. Having some money in one's pocket saves the time and inconvenience of doing so. Moreover, some consumers only have access to money as they are not eligible to have a credit card.

Although money is useful for carrying out transactions, it is costly to hold. Monetary instruments, like currency or checking accounts, in general earn less interest than riskless short-term securities such as Treasury bills. The decision to hold more money means investing less in securities that pay more interest, and the opportunity cost of holding money depends on how much interest income is lost. In order to decide how much money to hold, consumers must trade off the benefits of the ease and convenience in carrying out transactions against the cost in terms of forgone interest earnings. In their choices economic agents balance these two factors, holding more money when the opportunity cost is low and less when it is high. But as long as monetary instruments pay less interest than other securities, money will be costly to hold and consumers will have an incentive to economize on its use. In other words the money demand is lower the higher the difference between the interest paid on the other securities and the one paid on money.

But being frugal on the use of money is not socially optimal. Money is costly to hold, but it is essentially costless for central banks to produce. A central bank could make everyone that uses its money better off at no cost by increasing the quantity of real balances (i.e., the nominal quantity of money divided by the price level). The consumers would benefit from additional real balances, because money is more convenient to carry out transactions,

and it is costless for the central bank to provide. Thus, the optimal policy involves eliminating incentives to economize on the use of money. To do so, the central bank should eliminate the difference between interest rates on monetary instruments and on other securities, because then money would be costless to hold.

Since most types of money pay little or no interest, the optimal policy calls for setting nominal interest rates on short-term riskless bonds equal to zero. Setting the nominal interest rate to zero implies that the inflation rate should be equal in absolute value to the real rate of interest. This follows from the fact that in the long run, to a first approximation, the nominal interest rate equals the real interest rate plus inflation. Thus, if the real interest rate were around 2 to 3%, Friedman's arguments suggest that the central bank should follow a negative inflation rate for the economy at a rate of -2 to -3%.

Inflation tax and other taxes

Edmund Phelps (1973) criticized the Friedman rule on the grounds that it ignores considerations related to taxation and guessed that the inflation tax should be part of an overall optimal tax scheme. Phelps argued that inflation is a source of tax revenue for the government and that if inflation were reduced other taxes would have to be increased in order to replace the lost revenue. He also conjectured that some inflation would be desirable. That would be the case if the distortions associated with the inflation tax were less costly than the distortions associated with other taxes that the government might resort to.

Phelps raised very important questions. Why is inflation a source of tax revenue? Do the other taxes produce less distortions than the inflation tax? We start by addressing the first question. The government can borrow either by issuing debt or issuing money. In general, borrowing by issuing money is cheaper, as the government pays no interest on money. The revenue from the inflation tax, also known as seigniorage, is the amount the government saves by issuing money instead of debt. Formally, it is the product of the outstanding money times the interest rate (on government bonds). The tax base of the inflation tax is the stock of money and the tax rate is the nominal interest rate. Similarly to the other taxes, there is a Laffer curve for seigniorage. For low levels of interest rate, the real seigniorage increases with the interest rate, but for very high levels of the interest rate, it decreases with the interest rate as the real stock of money decreases more than proportionally with the interest rate.

Phelps argued that the other taxes introduce distortions of their own, which may outweigh the benefits of deflation. What is the nature of these distortions? Taxes distort private economic decisions because they create incentives on the economic agents to alter their behavior in order to avoid the tax. For example, a tax on bread increases the after-tax price that consumers

pay but decreases the pre-tax price that firms receive. An increase in the after-tax price reduces the quantity consumers want to buy, and a fall in the pre-tax price reduces the quantity firms are willing to supply. The tax makes firms worse off because they receive less per unit and produce less units. The tax also makes consumers worse off because they pay more per unit and consume less units. Typically, the government collects revenue from the tax and uses it to provide public goods and services, but the losses of those who pay the tax exceed the revenue collected. The difference between the losses of those who pay the tax and the revenue raised is known as the "deadweight loss" of the tax, and one principle of public finance is that taxes should be chosen in a way that minimizes these losses.

The Friedman rule would certainly not be optimal if the inflation tax was replaced by other tax increases that were even more distortionary. On the other hand, the fact that governments must choose among distortionary taxes does not necessarily invalidate the Friedman rule. The optimal mix depends on how distortionary the various taxes are. Phelps conjectured that at low rates of inflation, distortions associated with the inflation tax might be minor and that replacing the inflation tax with other taxes might result in greater deadweight losses. It turns out that Phelps's conjecture was inaccurate. The inflation tax is more distortionary than the other taxes.

Two set of results from public finance justify the Friedman rule even when all taxes are distortionary: the Diamond and Mirrlees (1971) optimal taxation rules of intermediate goods and the taxation rules of final goods developed by Atkinson and Stiglitz (1972). Intermediate goods are those goods which are used in the production of other goods and services. Taxes on these commodities are inefficient, according to Diamond and Mirrlees (1971), because they introduce two sets of distortions. First, they reduce production efficiency and increase the cost of producing final goods. Second, as this increase in cost affects the final goods prices, they distort final goods markets as well. As an alternative, the same revenue can be obtained by taxing final goods directly, and while this would distort final goods markets, it would not distort production efficiency. The optimal commodity taxation of final goods was established by Atkinson and Stiglitz (1972). According to them, under certain conditions, the optimal commodity taxation of final goods should be uniform. All final goods should be taxed at the same *ad valorem* rate.[2]

Economists have been modelling money in different ways. There are economists that classify money as an intermediate good and there are other economists that classify it as a final good. The first set of economists say that money, unlike a consumption good, is intrinsically useless. Money is valuable because it facilitates transactions, and as such it should be considered an intermediate good. Thus, according to Diamond and Mirrlees (1971) rule, the

2. An *ad valorem* tax is a tax whose amount is based on the value of the transaction.

inflation tax is really an indirect tax on other goods and taxing those goods directly is more efficient. The Friedman rule is a corollary of this public finance rule.

Other economists consider monetary models with less reasonable micro foundations, in which money is a final good. In this case the rules on optimal taxation, described by Atkinson and Stiglitz (1972) could justify Phelps' insight. These rules apply to *ad valorem* taxes on costly goods, and involve the comparison of the marginal excess burdens of alternative taxes that give the same revenue. However, the application of these rules to money is not straightforward for two reasons. Money has a negligible production cost and the inflation tax is a unitary tax, not an ad valorem tax.[3] Correia and Teles (1999) explain why the Friedman rule holds even when money is a final good. Here is their intuition. Assume a world with two consumption goods. One unit of good 1 can be produced with one unit of time while one unit of good 2 can be produced with α units of time. According to Atkinson and Stiglitz (1972) the optimal ad-valorem taxes on good 1 and 2 are equal to some positive tax τ. Measured in units of good 1 the unitary tax rate of good 2 is $\alpha\tau$. If the production costs of good 2 converges to zero, i.e. α goes to 0, then the unitary tax rate on good 2 when measured in units of good 1 is zero. Thus, also in a world where the alternative taxes are all distortionary, it is the zero marginal cost of producing money that implies a zero opportunity cost of holding money, that is a zero inflation tax.[4]

Incomplete set of tax instruments

One reason why the Friedman rule might not be optimal is when the tax system is incomplete. Phelps (1973) conjecture for inflation is valid when there are restrictions on taxation, that is, when there are factors of production, goods, monopoly profits or rents that cannot be taxed optimally. The optimality of the zero inflation tax is denied because the general conditions under which the two rules of public finance hold are violated. The production efficiency is undermined by the impossibility of setting taxation at an optimal level for some final goods or services. Various examples of this type of situation have been studied in the literature.

For instance Schmitt-Grohe and Uribe (2004) show that if the government is unable to fully tax pure monopoly profits, then deviating from the Friedman rule may be desirable. Taxing fully profits is optimal because it is a lump-sum

3. An unitary tax rate is based on the physical units transacted.

4. Burstein and Hellwig (2008) considered a model with money in the utility function and estimated the welfare costs of a 10% inflation to be about 1.3% to 2% of consumption. Lucas (2000) considers a shopping time model of money and reports welfare costs around 1%. Lagos and Wright (2005) in a search model of money report welfare costs between 3 and 4%.

tax, i.e. it does not have distortionary effects. When the government is unable to optimally tax profits, positive inflation may be a desirable instrument to tax the part of income that is sub-optimally taxed. The reason is that because at some point all types of private income are devoted to consumption, and because inflation acts as a tax on consumption, a positive nominal interest rate represents an indirect way to tax all sources of income. The higher the profits the higher will be the optimal inflation level. Schmitt-Grohe and Uribe (2004) determine that an optimal rate of inflation of 2% would require a mark-up, between cost and price, exceeding 30 percent, which is on the high end of the empirical estimates for most developed countries.

Adão and Silva (2018) provide another example, they show that when there are government transfers the Friedman rule fails too. Some government transfers, which represent pure rents, should be taxed by inflation. The optimal inflation rate increases significantly with the amount of these transfers. When transfers (net of income taxes) as a percentage of GDP are 5%, the optimal inflation rate is about 80 basis points but when they are 10% of GDP the optimal inflation rate is around 6%.

The failure of the Friedman rule due to tax evasion was established by Nicolini (1998). He considered a model with an underground economy sector in which firms evade income and consumption taxes. This underground sector generally consists of unregistered companies and small businesses that are usually owner-operated and that typically do not engage in illegal activities. They are just not regulated or taxed by the government. The firms operating in the underground economy enjoy a pure rent given by the amount of taxes that they manage to evade. Given that the underground economy uses cash intensively and is difficult to reach with other tax instruments, the government could indirectly tax these pure rents by imposing a positive inflation tax.

To understand the intuition it is helpful to consider the extreme case where cash is only used to purchase goods in the underground economy while purchases in the rest of the economy are made with credit. Then a consumption tax would only affect credit transactions. In this case uniform commodity taxation could be achieved by imposing, in addition to the consumption tax, a positive inflation tax. However, in the real world, where cash and credit are used for transactions in the above-ground economy, there is a trade-off between the consumption tax and the inflation tax. A positive nominal interest rate can partially balance tax rates across commodities by redistributing the tax burden from the consumption tax to the inflation tax on cash purchases in the underground economy.

Schmitt-Grohe and Uribe (2011) compute that for a 10% share of underground economy, which is regarded as reasonable for a developed economy, the optimal rate of inflation is only 50 basis points above the one associated with the Friedman rule. An inflation of 2% would require an underground economy around 1/3 of the total economy, which is

unrealistically too high. On the other hand, Cavalcanti and Villamil (2003), using a different monetary model with an informal sector[5], find that the optimal inflation is around 0% when the share of the underground economy is about 10%, and the optimal inflation is 6% when the share of the underground economy is 30%.

These three examples have in common that the monetary authority finds it optimal to use inflation as an indirect levy on pure rents that would otherwise remain untaxed. The literature has quantitatively evaluated them and found that depending on the fiscal instrument that is missing, the size of source of income that cannot be taxed and the monetary model that is used the departure from the Friedman rule can be timid or substantial.

The inflation tax is regressive

The evidence indicates that the income elasticity of the demand for money is less than one, which means that the poorer households hold a larger fraction of their income in money and conduct a bigger share of their transactions with money than the richer households. Based on this evidence Erosa and Ventura (2002) and Adao and Correia (2012) show that anticipated inflation has redistributive effects, which can justify a negative inflation. In Adao and Correia (2012) it is assumed that households perform transactions using either cash or costly credit. As transactions with credit exhibit economies of scale, optimality implies that a larger consumption level uses a larger share of credit. The transaction costs per unit of consumption decrease with the volume of consumption as they use a higher share of credit. Since high income households consume more than low income households they pay a higher fraction of their purchases with credit, pay lower transactions costs per unit of consumption, and hold less money as a fraction of total assets than low income households. Because high income households are better at avoiding the inflation tax than those with low incomes it follows that inflation is a regressive tax. This channel, therefore, is likely to put downward pressure on the optimal rate of inflation, insofar as the objective function of the policymaker is egalitarian.

Adão and Correia (2018) calibrate their flexible price model to the U.S. wealth and income distribution quintiles. They find that the effects on equity tend to reinforce the effects on efficiency. There is no trade-off: a lower inflation increases efficiency and equity. Moreover, they confirm that the impact of a moderate inflation varies across the quintiles. For a 10% inflation, after taking into account all costs with transactions and consumption taxes, the poorest

5. The two models differ with respect to the shopping time transaction technologies and production functions. The calibrations considered in the two papers are different too.

quintile of the population pays 3% more per each unit of consumption than the richest quintile. For low levels of inflation the difference is small. For instance, for an inflation of 2% this difference is about 0.3%.

Menu costs, sticky prices and relative price dispersion

There is a large and more recent literature, the neo-Keynesian literature, which in general abstracts from the role that money has in facilitating transactions and gives more importance to the interaction between inflation and nominal rigidities in the form of sluggish price adjustments. The models in this literature assume that some prices and wages cannot move so as to clear markets. Instead, firms change prices infrequently because it is costly to change them, even if these costs are small. These costs are referred to as menu costs and the models that incorporate these costs explicitly (or implicitly) are known as sticky price models. The typical example given is that of a restaurant, which must print new menus whenever it changes its prices. Printing menus is costly and this causes the restaurant to change prices only once in a while.

The majority of the models in this literature does not incorporate explicitly the menu costs in the pricing decision of the firms. Instead, they simplify the problem of the firms. They assume that in each period only a fraction of the firms can adjust their price, which will be invariant for some time in the future, until they have the opportunity to change it again. This implies that prices fail to adjust promptly and uniformly to changing market conditions. Such nominal rigidities in general lead to relative price dispersion. For instance, if in the presence of positive inflation some firms do not change their prices while others do. The firms that do not change their prices will have prices too low relative to the average price. This price dispersion, not caused by changes in preferences or technology, leads to a misallocation of resources. The best policy in this case is to minimize price dispersion by setting inflation to zero.

The welfare costs of a positive steady-state inflation in a sticky price model without explicit menu costs are somewhat different from the costs in those models that incorporate them explicitly. In the first type of models, higher inflation leads to greater price dispersion which leads to a more inefficient allocation of resources among firms, thereby lowering aggregate welfare. In the second type of models, there is an additional channel affecting welfare. Firms change their prices more frequently the higher is inflation, and this decreases welfare because firms incur in menu costs more often.[6]

6. Burstein and Hellwig (2008) determine that the welfare costs of inflation are much higher in an economy with Calvo style price staggering, than in a menu cost economy. In their benchmark model with money in the utility function and menu costs the welfare costs of a 10% inflation

More realistic monetary models incorporate both frictions, the price stickiness and the transactional demand for money. In such models the optimal rate of inflation falls in between the one determined by the money demand friction, deflation at the real rate of interest, and the one determined the sticky price friction, zero inflation. The intuition behind this result is simple. The benevolent government faces a trade-off between minimizing price adjustment costs and minimizing the opportunity cost of holding money. Khan, King and Wolman (2003) were the first to quantify the optimal policy. For their benchmark calibration the inflation under the optimal policy is -76 basis points. Thus, the quantitative analysis suggests that the trade-off is decided in favour of price stability.

Correia, Nicolini, and Teles (2002) point out that these neo-Keynesian models are also models with an incomplete set of tax instruments. They consider a neo-Keynesian model with money as a facilitator of transactions and prove that if the government can set state-contingent consumption taxes, then it is optimal to set the nominal interest rate to zero at every date and state. The basic intuition is the following. If there were no costs of changing prices then all firms would change their price after a shock and there would not be any relative price distortion. However, if prices are sticky, some firms would change their prices (incurring in menu costs) while others would not, which originates a relative price distortion. They show that this price distortion can be avoided by the government if the prices gross of consumption taxes change in response to the shock in such a way that no firm has the incentive to change its own price.

Downward nominal rigidities in factor prices

One reason to target a positive inflation is the presence of downward nominal rigidities. Allocative efficiency requires that relative prices should reflect the relative marginal costs of production. If these relative costs change then the relative prices should change too. If nominal prices are downward rigid, and contingent consumption or labor income taxes are not available, then any relative price change can only happen with an increase in the aggregate price level. Since variations in relative prices are efficient, a positive rate of inflation, aimed at accommodating such variations is welfare improving. An important example of a downward rigid price is the nominal wage. There is evidence for several developed economies of downward nominal wage rigidity,[7] which

---

are 1.3% in terms of consumption-equivalent variation. Calvo-style staggered pricing raises the welfare costs to 3.4%.

7. The downward nominal wage rigidity is due to the stringency of the employment legislation, the high coverage of collective agreements, and the dominance of sector-level bargaining and widespread extension procedures. Additionally, the downward nominal wage rigidity might be

acts as an obstacle to labour market adjustments.[8] Positive inflation alleviates this problem by allowing real wages to adjust in face of negative shocks, even if nominal wages do not fall, and avoid an increase of unemployment. A natural question, therefore, is how much inflation is necessary to 'grease the wheel of the labor market.' The answer appears to be not much. An incipient literature using estimated macroeconomic models with downwardly rigid nominal wages finds optimal rates of inflation below 50 basis points.[9]

Inflation rates can differ across regions of large countries (or monetary unions), reflecting among others: normal adjustment processes (such as price convergence or the Balassa-Samuelson effect), different cyclical position of the region, different composition of consumption, economic distortions resulting from segmented markets and insufficient competition. In the presence of frictions that make market changes more difficult, like downward nominal wage rigidities, adjustments within a large country (or monetary union) are easier if the central bank has a higher target for inflation. It would be easier as it would avoid having regions with extremely low inflation or even deflation, when that is not optimal. To our knowledge no fully fledged model has been built to address this specific question.

Zero bound on nominal interest rates

Stabilization is another main goal of central banks. Stabilization increases welfare because it reduces the spurious fluctuations of the main macroeconomic variables and promotes economic growth.[10] Central banks use the nominal interest rate to stabilize economic activity. Since prices take time to adjust when a central bank changes the nominal interest rate it affects temporally the real interest rate and the aggregate demand. Thus, in response to a negative shock to economic activity, central banks reduce the nominal interest rate as a way to decrease the real interest rate and foster aggregate demand.

Common sense would indicate that the lower the inflation rate the higher the risk of the nominal interest rate hitting the zero lower bound. Thus, a central bank's ability to conduct successful stabilization policy can be restricted if the rate of inflation is very low. To determine rigorously the

---

related with the "morale effects" associated with wage cuts (see Bewley (2002)). Du Caju et al. (2008) measure the strength of employment protection laws in 23 EU countries and concludes that it ranges from the high level of Spain, Portugal and Greece to the very low level of Ireland.

8. Evidence based on micro data and firm surveys suggests significant nominal wage rigidity in developed countries. See ECB (2009) for a survey of the evidence for Europe.

9. For instance, Kim and Ruge-Murcia (2009) quantify this effect, and get an optimal inflation rate of 35 basis points.

10. Gadi Barlevy (2004) argues that fluctuations can affect welfare, by affecting the growth rate of consumption. He estimates that the welfare effects are likely to be substantial, about 8% of consumption, much larger than Lucas' original estimates.

inflation level that enables a successful stabilization policy it is necessary a macroeconomic model calibrated (or estimated) to fit a real economy and a central bank that follows a reasonable (or optimal) monetary policy. A few exercises of this sort have been done. Most of these studies recommend an inflation rate below 2%. Coibion et. al. (2012) calibrate a sticky price model to broadly match the moments of macroeconomic series and the historical incidence of hitting the zero lower bound in the U.S., and solve for the rate of inflation that maximizes welfare. For plausible calibrations of the structural parameters of the model and reasonable properties of the shocks driving the economy, the optimal inflation rate is less than 2%. The result is robust to changes in parameter values, as well as to the stabilization policy followed by the central bank.

There are other factors that should be taken into account when considering the impact of the zero lower bound on the optimal inflation level, which current models in general do not contain. One is that unconventional monetary policy can mitigate the constraint posed by the zero lower bound. The quantitative evaluation of the strength of the unconventional monetary policy is still work in progress, but the existing results point towards the effectiveness of the unconventional monetary policy, though the evidence regarding the degree of effectiveness is mixed.[11] [12] Another argument has to do with the idea that in a higher inflation environment agents adjust their holdings of money and prices more frequently, reducing the impact of changing the policy rate on the real rate and thus on the macro variables. This feature is usually not included in macro models used to evaluate the relevance of the zero lower bound constraint, which typically assume that the frequency of adjusting prices does not change and that there is no opportunity cost of holding money. If one considers this type of adjustment, a higher target may not necessarily provide a larger buffer against the zero lower bound.[13] Finally, when there are downward nominal wage rigidities firms that experience large negative shocks will have to adjust not by cutting wages but by laying off more workers. Thus, the nominal aggregate wage, or the unitary labour costs,

11. The existing studies vary on the degree of unconventional monetary policy success that is found. Some find that the effects are more moderate than conventional monetary policy while others find that they are comparable if not greater. For instance Chen, Cúrdia and Ferrero (2012) find that the unconventional monetary policy is less powerful than the conventional policy, but on the other hand Gilchrist, López-Salido and Zakrajsek (2014) find that the efficacy of unconventional monetary policy in lowering real borrowing costs is comparable and in some cases is twice as large as that of conventional policy (like for real corporate borrowing costs).

12. Dorich et al. (2015) results with the Bank of Canada's main macroeconomic model find that the possibility of using unconventional monetary policy basically offsets the need to increase the inflation target due to a fall in the real rate.

13. For instance Adão and Silva (2015) estimate for the US in 2000 that an unanticipated temporary decrease of 30 basis points in the nominal interest rate would decrease the real interest rate after one month by only 3 basis points, while in 2013 the same shock would decrease the real interest rate by 12 basis points.

do not fall by as much when there are downward nominal wage rigidities. As a result, by moderating declines in nominal wages, downward nominal wage rigidities moderates changes in prices and consequently in policy rates.[14] This dampening effect reduces the frequency of zero lower bound episodes for any given inflation target and, therefore, the case for a high inflation to avoid zero lower bound episodes is not as strong.[15]

Growth rate and inflation

There are many empirical studies that examine the relation between inflation and economic growth. Most of these studies report a negative correlation between inflation and economic growth during periods of high inflation (2 digits). For periods of low inflation the correlation tends to be statistically insignificant. On the other hand, there are few theoretical studies on this relationship. Recently, Oikawa and Ueda (2018) provided a model with a negative correlation between anticipated inflation and growth. In their model inflation (and deflation) has real growth effects. They consider an endogenous growth model with sticky prices due to menu costs. In the model growth is positively correlated with R&D investment. The higher the present value of the profits associated with R&D investment is, the higher its level will be.

In their model the growth rate of the nominal variables is equal to $n = g + \pi$, where g is the real growth rate of the economy and π the inflation rate. Everything else equal, the higher the nominal growth rate, the more often the firms have to change the price of their products and incur more menu costs. If the growth rate of the nominal variables is different from zero, firms would like to change their prices, which would reduce the reward for innovation and lower the level of R&D investment. As such the optimal solution in this model is to have a nominal growth of zero. In this case, firms abstain from changing their prices and avoid paying the menu costs.

In the model there is a relationship between the nominal growth rate and the real growth rate, i.e. g is a function of n, $g(n)$. On a balanced growth path for a given inflation rate, the real and nominal growth rates have the relationship $\pi = n - g(n)$. The pair of growth rates is pinned down by choosing an inflation

14. If the Taylor rule has a high weight on inflation's deviations from target and a low weight on output fluctuations then it is possible that with downward nominal wage rigidities the inflation target could be lower.

15. When both frictions are considered simultaneously (zero lower bound and downward nominal wage rigidities) the optimal inflation level might be smaller. Amano and Gnocchi (2015) find that by adding downward nominal wage rigidities into a sticky price model which already incorporates the zero lower bound, the optimal inflation target decreases. More specifically, without any of these two frictions the optimal inflation target is 0%, with the zero lower bound only is 4.5%, with the downward nominal wage rigidities only is 1% and with both frictions is 1.5%.

rate. The optimal inflation rate is the one that corresponds to the growth level that maximizes the households' welfare. Oikawa and Ueda (2018) calibrate their model to the U.S. economy and obtain interesting results. The optimal inflation rate is very close to the growth-maximizing inflation rate, which is around -2%. The cost of suboptimal inflation is substantial, on the balanced growth path, the growth rate is reduced by half at about 10% inflation or deflation.

Other arguments discussed in the literature:

Tax collection costs

Other arguments have been suggested in the literature to justify a positive inflation tax. One is the tax collection costs, which is an important difference between the traditional fiscal instruments and the inflation tax. While raising revenue with the inflation tax is costless, raising revenue with the other tax instruments implies higher costs. These costs include the burden of organizing the tax system and enforcing it. When these costs are taken into account then the Friedman rule is not optimal. De Fiore (2000) quantifies how important these are. She reports that under the most unfavorable (and unrealistic) case, where these costs are all variable costs and tax collection requires throwing away 20 percent of the government revenue, the optimal inflation tax remains below 1%. A different study, Yesin (2004), considers simultaneously tax collection costs and the presence of an informal sector and obtains that for the U.S. the optimal inflation is around 4%.

Foreign demand for domestic currency

A few countries, like the US and the E.U., have a currency with a positive foreign demand. For instance, it is estimated that more than half of U.S. currency circulates abroad. The Friedman rule is not optimal once there is a foreign demand for the domestic currency. The intuition is that the deflation implied by the Friedman rule would represent a transfer of real resources by the domestic economy to the rest of the world, as nominal money balances held abroad increase in real terms at the rate of deflation. Inversely, a positive inflation would entail collecting resources from foreign residents. Thus, the benefit of inflation is the resources collected abroad, while the cost would be a higher opportunity cost of holding money which would increase transactions costs for domestic agents. It turns out that the marginal benefit and marginal cost are equated for an inflation larger than the Friedman rule inflation.

The fraction of the seigniorage paid by foreigners is proportional to the fraction of domestic currency held abroad. The higher the fraction held by foreigners and the more inelastic is the foreign demand the larger is the benefit

of inflation and larger is the optimal rate of inflation. Schmitt-Grohe and Uribe (2011) quantify the optimal rate of inflation. Using a range of empirical estimates for the size of foreign demand for U.S. currency they come up with optimal rates of inflation between 2 and 10%. The value of 10% is obtained for a very high demand of foreign currency. Again, this argument for a relatively high inflation rate does not apply to the majority of countries because they do not have an international demand for their currency.

Quality improvements and measured inflation

The quality of the goods improves over time but, the price observed by the statistical agencies is the price of a physical unit of the good not the price of the good per unit of quality. This implies that if there is not a quality adjustment in the measured prices, the consumer price index will overstate the true inflation. The classic example used to illustrate this potential quality bias in inflation is the evolution of the price of the personal computer. The quality of personal computers, measured by characteristics such as memory, processor speed, and screen quality, increases every year. Suppose that the price of personal computers between 2017 and 2018 increased 2%. If the statistical office in charge of producing the consumer price index did not adjust the price index for quality improvements, then it would report a 2% inflation in personal computers. However, because a personal computer in 2018 provides more services than does a personal computer in 2017, the quality-adjusted rate of inflation in personal computers is lower than 2 percent.[16]

In the presence of improvements in the quality of goods, to guarantee price stability, and welfare maximization in a sticky price economy, the central bank might target either a positive, or a zero or a negative inflation. It will depend on two things: (i) whether the price stickiness is in the non-quality or in the quality adjusted prices and (ii) whether the statistical agency in charge of computing the price index used to determine inflation adjusts or not the prices for quality. The intuition is provided below.

Assuming there is a positive degree of price stickiness in the economy then the optimal policy should try to keep the prices of the goods that are sticky constant over time to avoid inefficient price dispersion. If the price stickiness is in the non-quality adjusted prices then the optimal policy should try to keep these prices constant over time. If the statistical agency responsible for constructing the consumer price index does not correct the prices for the quality of the goods then targeting a zero inflation rate is efficient. On the other

16. The difference between the reported rate of inflation and the quality-adjusted rate of inflation is called the quality bias in measured inflation. In 1996, the Boskin report (Boskin et al. (1996)) estimated the quality bias to be around 0.6 percentage points for the US. This bias is not constant over time as it depends on the economy's structure as well as on the index methodology used. Nowadays, this bias seems to be negligible for many developed countries.

hand, if the statistical agency adjusts the index to reflect quality improvements then to guarantee the non-quality adjusted price does not change, the price index should be falling at the rate of quality improvement. This means that the optimal deflation rate is equal to the rate of quality improvement. Thus, the optimal inflation is either zero (when the statistical agency does not correct the price index for quality improvements) or negative at the rate of quality improvement (when the statistical agency does correct the price index for quality improvements).

However, if instead it is the quality-adjusted prices that are sticky, then the optimal inflation is either zero (when the statistical agency does corrects the price index for quality improvements) or positive at the rate of quality improvement (when the statistical agency does not correct the price index for quality improvements). Ultimately, it is an empirical question whether it is the non-quality adjusted or it is quality adjusted prices that are stickier.[17] Nonetheless, given that at the present most statistical agencies correct for quality, the optimal inflation level should be either negative or zero.

Costs of unexpected inflation

Thus far we have discussed the optimal long-run level for inflation. We now address the importance of a stable inflation. An important recommendation in the literature on the optimal inflation is that inflation should be stable in order to avoid inflation surprises. Unexpected inflation has welfare costs. A stable rate of inflation is good for everyone, because it facilitates the use of prices in making decisions by all agents in the economy. A variable inflation rate makes it difficult to distinguish changes in the relative prices from changes in the aggregate price, which implies an efficiency loss in the allocation of the resources in the economy. For instance in Lucas (1972), when firms observe the price of the good they produce increasing more rapidly than expected, they might believe that there was an increase in the demand for their product. That will lead firms to increase aggregate supply which leads to a too high level of output in the economy. Also, if workers know in advance the rate of inflation, that helps them determining the purchasing power of their wages, and take better employment decisions. Similarly with homebuyers, knowing future inflation helps determining the real cost of a particular mortgage loan. Lenders and borrowers also benefit from knowing how much of a particular interest rate represents real interest. In short, stable inflation improves welfare by eliminating one source of uncertainty in economic life.

Unanticipated inflation may also have important distributional effects. Surprises in the inflation rate lead to redistributions of income and wealth

17. This question has not been fully addressed by the empirical literature on price rigidities.

between various groups of the population. Unanticipated higher inflation leads to a redistribution of wealth from lenders to borrowers, and unexpected lower inflation redistributes wealth in the opposite direction. Typically, the government sector is a nominal borrower while households are savers. As a result, an unexpected increase in the rate of inflation causes a redistribution from the latter to the former. In addition, there is also a redistribution of wealth inside the household sector, from the old generation (that usually holds higher amounts of nominal assets) to the new generation. This principle applies to other nominal contracts besides loan contracts. Two examples: an inflation above what was expected implies for the pensioners a deterioration in their real pension and for workers a deterioration in their real wage.

If inflation was observable and contracts could be indexed then the two problems described above could be eased substantially. With observability of inflation the economic agents could distinguish between relative price changes and aggregate price changes. Inflation indexation could alleviate the arbitrary redistributions of income. However, observability and indexation only offer a partial resolution to the problems caused by inflation.

There are two reasons for this. First, there is not perfect observability of inflation. Data on current inflation is not available in real time. Typically in developed countries, the CPI is estimated monthly and released in the middle of the following month, while the GDP deflator, which is more relevant for firms, is estimated quarterly and released in the middle of the following month. Second, there are many different possible measures of the price level, depending on the bundle of goods used in calculating the price index. A suitable measure for one economic agent might not be suitable for another. Unexpected changes in relative prices through time will favor some households at the expense of others, depending on the bundle of goods they consume. Indexing using the CPI for example may hurt a particular borrower if the CPI falls but the bundle of items that person consumes nevertheless rises in price. Also, in general, employers and workers will not agree on which basket of goods to target either. Employers will not be willing to provide full indexation to workers, because firms' costs and revenues will not rise in the same amount with the CPI inflation. Lender banks, of course, can lose in a similar way. Alternatively, maybe the borrower is interested in what he can afford to pay back; in that case he wants the contract indexed to the nominal wage.

Thus, indexing does not eliminate completely the risk. Both parties must expect that their risk of loss will be reduced with indexation to inflation, on average, but not completely eliminated. On the other hand, the costs of writing contracts indexed to inflation are higher than the costs of non-contingent contracts. These costs include: (i) bargaining costs over which indexes to adopt and the respective weights, and (ii) monitoring those indexes. In periods when the unanticipated movements in the price level are likely to be small, the risks from indexing and the risks from unanticipated inflation are of

the same order of magnitude. On the other hand, indexing is meaningful if unanticipated price level movements are large and frequent. In this case, the risks of unanticipated inflation exceed the risks associated with adopting a particular index. Indexing is therefore more likely to be observed in countries which tend to experience a lot of price level variability and less likely in countries whose inflation rates have tended historically to be quite stable.

Inflation interacts with the tax system in many ways, and as the various taxes are typically not fully indexed to inflation, it can cause additional distortions. We exemplify this with three examples. It might increase the effective marginal income tax rate of a progressive tax system, if the personal income tax brackets are not fully indexed to inflation. When nominal income increases, due to inflation, people may move to a higher tax bracket ending up paying higher taxes rates even though their real income is unchanged. Moreover, inflation decreases the net rate of return of the equity market and raise the cost of capital to firms, as it increases the effective capital gains tax rate. Even if the real value of a firm is unchanged, the owners of this firm's equity loose with inflation as capital gains are computed as the difference between the sale price and the buying price. Finally, depreciation allowances are based on book values and not on the current replacement cost of capital. As a result, inflation drives a wedge between the book value and the replacement cost of capital by understating the true cost of depreciation and overstating the profits for tax purposes, and thus reducing the incentive to invest in capital equipment.

Conclusion

This paper surveys whether the objective of central banks in maintaining a stable and low inflation is consistent with the literature on the optimal rate of inflation. It provides quantitative values for the optimal average inflation rate as well as the intuition. The literature in this topic is vast, and the papers can be divided into three groups: (i) those that find the inflation rate should be negative, (ii) those that find the average inflation should be zero, and (iii) those that find the inflation rate should be positive. Typically, in the first set of papers the Friedman result, that the optimal nominal interest rate is zero, holds. The optimal inflation rate is the negative of the real rate of interest. The optimal nominal interest rate is zero, so people feel no incentive to economize on money holdings. Included in this group are other papers that show that a negative inflation rate can promote growth and improve income and wealth equality.

In the second group of papers, the zero inflation is optimal. These papers consider sticky prices, which implies welfare costs either from inflation or deflation. In these models, price changes lead to price dispersion across firms and this results in demand being too high for some firms and demand being

too low for others. With zero inflation firms do not have to change prices. The third group of papers consider various reasons for a positive inflation rate. Reasons for a positive inflation can be: an incomplete set of tax instruments, downward nominal rigidities in factor prices, tax collection costs, foreign demand for domestic currency, and the zero bound on nominal interest rates.

Each of the papers surveyed provides a model for a particular optimal inflation rate. The policy makers should understand and know the model behind each optimal inflation rate. With that knowledge, the policy maker must then determine an optimal inflation rate that balances the various models. The 2% percent target for inflation that central banks follow can be thought as being determined in this way.

Based on the literature surveyed the 2% target cannot be ruled out as an optimal inflation level. We find that the 2% target followed by most central banks is in accordance with the literature as the papers surveyed estimate that the optimal rate of inflation ranges from minus the real rate of interest to 6%. Nevertheless, it can be argued that the 2% target is on the upper side of the recommendations, as the majority of the papers find the optimal inflation rate to be negative or zero.

One reason for a target is to "anchor expectations." According to this reasoning a justification for the target is not that the 2% is the precise optimal level of inflation. The justification is that a central bank must choose a number for the target and maintain it for a long period of time, even if it is no longer at the precise optimal level. The central bank makes this commitment, which should not be abandoned later, because otherwise it loses credibility. As such the rule should be that the inflation target should not be moving in reaction to changing macroeconomic conditions. Inflation targets should be changed infrequently, and only for very good reasons.

References

Adão, B. and A. Silva. 2018. "Real transfers and the Friedman rule". Economic Theory, https://doi.org/10.1007/s00199-018-1105-0

Adão, B. and A. Silva. 2015. "Increased Strength of Monetary Policy". Banco de Portugal, Economic Studies, 1(2).

Adão, B. and I. H. Correia. 2018. "Changes in the inflation regime". Banco de Portugal, Manuscript.

Amano, R. and S. Gnocchi. 2015. "2 + 2 < 4? : Monetary Policy in the Presence of Downward Nominal Wage Rigidity and the Zero Bound on Nominal Interest Rates". Bank of Canada, Manuscript.

Atkinson, A. B., and Stiglitz, J. E. 1972. "The Structure of Indirect Taxation and Economic Efficiency," Journal of Public Economics 1, 97-119.

Barlevy, Gadi. 2004. "The Cost of Business Cycles under Endogenous Growth" American Economic Review, September, vol. 94, nº 4, 964-990.

Bewley, Truman, 2002, "Why Wages Don't Fall During a Recession", Harvard University Press.

Cavalcanti, T. and A. Villamil. 2003 "The optimal inflation tax and structural reform." Macroeconomic Dynamics, 7(3):333–362, June

Boskin, M., E. Dulberger, R. Gordon, Z. Griliches and D. Jorgenson. 1996. "Toward a more accurate measure of the cost of living". Final Report to the Senate Finance Committee from the Advisory Commission to Study Consumer Price Index, Washington D.C., Senate Finance Committee.

Burstein, A. and C. Hellwig. 2008, "Welfare Costs of Inflation in a Menu Cost Model", American Economic Review: Papers & Proceedings, 98:2, 438–443

Chen, H., V. Cúrdia and A. Ferrero. 2012. "The Macroeconomic Effects of Large-scale Asset Purchase Programmes". The Economic Journal, 122(564): F289–F315.

Coibion, O., Y. Gorodnichenko and J. Wieland. 2012. "The Optimal Inflation Rate in New Keynesian Models: Should Central Banks Raise Their Inflation Targets in Light of the Zero Lower Bound?" Review of Economic Studies, 79: 1371–1406.

Correia, I.H., and P. Teles, 1999. "The optimal inflation tax." Review of Economic Dynamics 2 (2), 325–346.

Diamond, P. A., and Mirrlees, J. A. 1971. "Optimal Taxation and Public Production," American Economic Review 63, 8-27, 261-268.

Diercks, Anthony, 2007, "The Reader's Guide to Optimal Monetary Policy". Available at SSRN: https://ssrn.com/abstract=2989237 or http://dx.doi.org/10.2139/ssrn.2989237

Dorich, J., N. Labelle, V. Lepetyuk and R. Mendes. 2015. "Assessing the Benefits of a Higher Inflation Target". Bank of Canada, Manuscript.1-268.

Du Caju, P., E. Gautier, D. Momferatou, and M. Ward-Warmedinger. 2008. "Institutional features of wage bargaining in 23 European countries, the US and Japan". ECB Working Paper Series 974.

De Fiore, F. 2000. "The optimal inflation tax when taxes are costly to collect," Working Paper Series, 38, European Central Bank.

ECB. 2009. "Wage dynamics in Europe: final report of the Wage Dynamics Network (WDN)", available at http://www.ecb.europa.eu/home/pdf/wdn_finalreport_dec2009.pdf

Erosa, A., and G. Ventura. 2002. "On inflation as a regressive consumption tax." Journal of Monetary Economics 49 (4), 761–795.

Fagan, G., and J. Messina. 2009. "Downward Wage Rigidity and Optimal Steady State Inflation". ECB Working Paper Series 1048.

Fernández-Villaverde, J., G. Gordon, P. Guerrón-Quintana and J. Rubio-Ramírez. 2012. "Nonlinear Adventures at the Zero Lower Bound", Federal Reserve Bank of Philadelphia Working Paper 12-10.

Friedman, M. 1969. The Optimal Quantity of Money and Other Essays. Chicago: Aldine.

Gilchrist, S., D. López-Salido and E. Zakrajsek. 2014. "Monetary Policy and Real Borrowing Costs at the Zero Lower Bound". Federal Reserve Board Finance and Economics Discussion Series 2014-03.

Khan, A., R. G. King and A. L. Wolman, 2003. "Optimal Monetary Policy," Review of Economic Studies, Blackwell Publishing, vol. 70(4), pages 825-860, October

Kim, J. and F. Ruge-Murcia, 2009, "How much inflation is necessary to grease the wheels?" Journal of Monetary Economics, vol. 56, issue 3, 365-377

Kim, J. and F. J. Ruge-Murcia. 2011. "Monetary policy when wages are downwardly rigid: Friedman meets Tobin". Journal of Economic Dynamics and Control, 35(12): 2064-2077.

Lagos, R. and R. Wright. 2005. "A Unified Framework for Monetary Theory and Policy Analysis." Journal of Political Economy, 113(3): 463–84.

Lucas, R. E., Jr., 1972. "Expectations and the Neutrality of Money". Journal of Economic Theory, 4: 103-124.

Lucas, Robert E., Jr. 1987. Models of Business Cycles. Oxford: Basil Blackwell.

Lucas, Robert E., Jr. 2000. "Inflation and Welfare." Econometrica, 68(2): 247–74.

Nicolini, J.P., 1998, "Tax evasion and the optimal inflation tax." J. Dev. Econ. 55 (1), 215–232

Oikawa, K. and K. Ueda, 2018, "The optimal inflation rate under Schumpeterian growth," Journal of Monetary Economics 100 (2018) 114–125

Phelps, E. S. 1973. "Inflation in the Theory of Public Finance." Scandinavian Journal of Economics 75, pp. 67-82.

Schmitt-Grohe, S., and M. Uribe, 2004, "Optimal Fiscal and Monetary Policy under Imperfect Competition," Journal of Macroeconomics 26, 183-209.

Schmitt-Grohe, S. and M. Uribe, 2009, "Foreign Demand for Domestic Currency and the Optimal Rate of Inflation," Journal of Money Credit and Banking, 44, January.

Schmitt-Grohe, S. and M. Uribe, 2011, "The Optimal Rate of Inflation," in Handbook of Monetary Economics, editors Benjamin M. Friedman and Michael Woodford, Volume 3B, Elsevier, San Diego CA, 653-722.

Yesin, P. 2004. "Tax Collection Costs, Tax Evasion and Optimal Interest Rates," Working Papers 04.02, Swiss National Bank, Study Center Gerzensee.

www.bportugal.pt