Banco de Portugal Economic Studies

Volume II

BANCO DE PORTUGAL
EUROSYSTEM

2

2
volume II

Banco de Portugal Economic Studies

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
EUROSISTEMA

Lisbon, 2016 · www.bportugal.pt

BANCO DE PORTUGAL ECONOMIC STUDIES | Volume II - n.º 2 | Lisbon 2016 • Banco de Portugal Av. Almirante Reis, 71 | 1150-012 Lisboa • www.bportugal.pt • Edition Economics and Research Department • ISSN 2183-5217 (online)

Content

Editorial

April 2016

The second 2016 issue of the Banco de Portugal Economic Studies contains three very diverse contributions. The first paper deals with some of the most perennial problems of monetary policy, namely how to deal with low inflation. The perspective is global and the issues are as timely as they could be as they are now confronting monetary authorities all over the world, in particular those in the Euro area. The second paper looks at firms and examines how solid they are when seen through a credit market´s perspective. The paper uses Portuguese data and the results can be seen as yet another contribution deepening our knowledge of Portuguese firms. The third paper examines labor markets and the patterns they display concerning gender disparities in wages. The issues are universal but the contribution builds on a rich dataset with Portuguese longitudinal labor market information.

The first article of this issue of the Banco de Portugal Economic Studies was written by Pedro Teles and by Joana Garcia and it is titled "How can the Phillips Curve be used for today's policy?". Are there still Phillips curves showing a trade-off between inflation and unemployment? Teles and Garcia engage in an exercise showing that for the US, Germany, France and Japan the data clouds with annual unemployment and inflation rates can be seen as a sequence of shifting short run Phillips curves. From the 1950s up to the 1980s the curves shift up and become slightly closer to being vertical. Then, a reversal occurs, the disinflationary process of the last 30 years takes place and the curves shift down and become closer to being horizontal lines. The fact that these short run curves are becoming closer to being horizontal may be explained by the adoption of inflation targeting policies: in the limit a successful targeting policy would generate perfectly horizontal Phillips curves. Furthermore, when we examine how the locations of the curves change, we realize that in the European countries and in Japan there seems to be a shift to the right of the curves: this means that the natural rate of unemployment has gone up over time.

If we move from the short run to the long run we see that the evidence for monetary policy neutrality is still there. Teles and Garcia revisit the data, showing that in the long run inflation depends mostly on money growth and that nominal interest rates move one-on-one with inflation, exactly as predicted by Fisher's equation ($i = r + \pi^e$) with a stable real interest rate.

The facts seem uncontroversial but they raise a tough problem for monetary policy makers particularly in the Euro area. With inflation currently below target, it would seem it is time to climb up the short run Phillips curve by expansionary monetary policy actions pushing interest rates down. That is what policy makers have been doing, with instrumental rates close to or even below zero. However, since nominal interest rates have been close to zero for quite a few years, one can see the current situation as having some of the features of a long run equilibrium. The problem then is that the Fisher relation tells us that zero interest rates are only consistent with extremely low levels of inflation. There is a tension between the short run uses of instrumental interest rates, where expansionary policies drive down interest rates and keep them low, and a long run perspective where a higher rate of inflation, more in line with policy targets, requires higher nominal interest rates. We will only get to the target levels of inflation with higher interest rates, but how do we get there when the short run concerns seemingly point in the opposite direction? The answer is not obvious. Monetary policy is rarely easy but in the current situation it seems to be as tough as they come. . . .

The second paper, by António Antunes, Homero Gonçalves and Pedro Prego, presents us with a quite different set of issues as it deals with a set of questions a little closer to microeconomics. It has the title "Firm Default Probabilities Revisited". The paper´s objective is to model the probability that a firm will default in a given year based on data characterizing that firm and the business cycle for the previous year.

The data used for the estimation comes from two sources. The first is the Banco de Portugal Central de Balanços, a database with annual balance sheet and financial statements including most Portuguese firms. The second is the Portuguese central credit register. The data is split in 10 groups of firms, by two size types (micro and all others) and by five industry groups. Firms with no employees or turnover, negative assets, etc. were eliminated from the sample. A default is defined as having 2.5% or more of the total outstanding loans overdue for at least three consecutive months.

The large pool of variables used in the model includes measures of profitability, size, leverage, liquidity, capital structure, macro factors, etc. Variables were used in logs, ratios and as ranks in the firms group. The baseline model predicting probability of default was a logit. The methodology adopted used general criteria from the literature and specific criteria defined by the authors to select up to ten explanatory variables from this large pool for each logit equation for the ten groups of firms. The results were subject to several specification or robustness analysis with positive results. Variables that turned out to be important in the models included profitability and liquidity measures. One curious result obtained was that for micro firms the ratio of trade debt to total liabilities was always selected as a significant variable positively associated with the default probability. From the estimates described earlier the authors constructed a credit quality classes with eight credit quality steps following on ECB methodologies. Over the years the empirical default rates match the model's probabilities except for excessive defaults in 2009.

In a country where at the end of 2015 the debt of non-financial corporations was about 115% of GDP, improved decision making in the credit area seems quite crucial for achieving a better economic performance. The availability of this type of model will help in providing better credit analysis for financial institutions, improving the quality of the credit´s allocation process. Overall this type of instrument should also be relevant for assisting investors in their decisions.

The third and final article is authored by Ana Rute Cardoso, Paulo Guimarães, Pedro Portugal and Pedro Raposo. It carries the title "The Sources of the Gender Wage Gap". It is well known that all over the world men and women do not have the same wages and Portugal is no exception. However, to what extent is the wage gap between men and women a function of differences in labor market relevant characteristics of workers and of the industries and firms that employ them and the job titles to which they are assigned? What is the role of segregation into subsets of industries, firms or job titles in explaining the wage gap?

The authors address these questions using Quadros do Pessoal, a rich administrative employer-employee-job titles matched data set covering the years from 1986 up to 2013 and including almost 29 million observations of full time workers. The variable of interest is the real hourly labor earnings. The econometric analysis explains these wages using a set of variables measuring the characteristics of the workers and firms such as education, experience, tenure, and firm size. More to the point, the longitudinal nature of the dataset allows for the construction of fixed effects for workers, firms and job-titles, allowing for a good control of the many time-invariant characteristics that are the source of the sizeable heterogeneity found in labor market microdata.

The econometric analysis starts by estimating the wage gap after controlling for all the variables mentioned before by means of a Machado and Mata decomposition. At the level of median wages there was a reduction in the corrected wage gap between 1991 and 2013 as the gap went from (approximately) 35.1 percentage points to 20.5 points. However, this occurred despite an improvement in the relative positioning of women's characteristics over men´s in during the years studied (more education, more experience). This is explained by the "value" of these relevant characteristics being smaller for women than for men. For example, the return to schooling is smaller for women than for men with the log coefficients of years of education lower by almost 1% in absolute terms.

But the most interesting contribution of the paper is its use of a methodology that allows the estimation of a very large set of multiple fixed effects. The results of this econometric methodology, when subjected to a decomposition analysis proposed by Gelbach, show how the different sources of heterogeneity contributed to the change in the gender gap. The results show that women are disproportionately allocated to firms and job-titles that lead to lower wages. An elimination of the segregation across firms would decrease the wage gap by 5.8 percentage points. A similar elimination of job-title segregation would decrease the wage gap by 4.3 points. Taken together, segregation across firms and job-titles explains two fifths of the wage gap.

All in all, these results show that little progress has occurred concerning the gender equity of the Portuguese labor market, adding to a host of other problems that should be improved upon by well informed and designed policies.

How can the Phillips curve be used for today's policy?

Pedro Teles with Joana Garcia
Banco de Portugal

April 2016

Abstract
Simple observation seems to suggest a downward shift of the Phillips curve to low levels of inflation for countries such as the US, Germany, France and Japan. A cloud of inflation-unemployment data points can be read as a family of short run negatively sloped Phillips curves intersecting a vertical long run Phillips curve. How can the evidence on these families of Phillips curves be used for policy? How can it be used to induce higher inflation in today's low inflation context? (JEL: E31, E40,E52,E58, E62, E63)

Introduction

Why is inflation low in the Euro area? Is it because interest rates cannot be lowered further? Or is it because interest rates are too low? Can these two questions both make sense? Can inflation be low because interest rates are not low enough, as it can be low because interest rates are too low?

It is indeed a feature of monetary economics that apparently contradictory effects coexist. The key to finding answers to the questions above is to distinguish short run effects from long run ones that tend to work in opposite directions. While in the short run inflation may be raised by lowering nominal interest rates, in the long run high inflation can only be supported by high rates. In the short run, lower policy rates may induce both higher inflation, and lower unemployment. This is consistent with a negative empirical relationship between inflation and unemployment, the Phillips curve. Instead in the long run, lower rates do not seem to have first order effects on growth, and, instead of raising inflation, they lower it, one-to-one. This article is about this distinction, of the short run and long run effects of monetary policy, in an attempt at answering the questions of why inflation is low in the Euro area and what policy should do about it. In particular, we want to discuss ways in which the evidence on the Phillips curve can be used to achieve higher inflation.[1]

E-mail: pteles@ucp.pt; jomgarcia@bportugal.pt

1. While we would expect money to be neutral in the long run, money may also be neutral in the short run, meaning that the long run effects could happen fast, even instantaneously. When

Central bankers are confident that the way to keep inflation at target is to have nominal rates be lower than average when inflation threatens to deviate down from target, and to have nominal rates above average when inflation deviates upwards. Monetary models are not inconsistent with this view, provided average interest rates move positively, one-to-one with the target.

Short run deviations from average rates may keep inflation at target. These days nominal interest rates are much below average, since average nominal rates that are consistent with a target of 2% should be between 2 and 4%, and they are zero. So is this a way to induce inflation to go back to target? The key to answer this is in the time frame of the deviation from average. Policy rates have not been below average for the last one, two or even three years. They have been below average for the last eight years, and they are expected, and announced, to stay low for a few more years. This can hardly be seen as a short run deviation from average. It looks a lot more like a lower average. And lower average nominal rates mean lower average inflation rates, in the models and in the data.

Money in the long and short run

In his Nobel Lecture in 1996, Robert Lucas goes back to the data on the quantity theory of money and the Phillips curve to make the case for the neutrality of money in the long run and the absence of it in the short run. Lucas also goes back to David Hume' essays "Of Interest" and "Of Money" published in 1752. Two of the wonderful quotes from those essays are:

> It is indeed evident that money is nothing but the representation of labour and commodities, and serves only as a method of rating or estimating them. Where coin is in greater plenty, as a greater quantity of it is required to represent the same quantity of goods, it can have no effect, either good or bad ... any more than it would make an alteration on a merchant's books, if, instead of the Arabian method of notation, which requires few characters, he should make use of the Roman, which requires a great many. [Of Money, p. 32]

and

> There is always an interval before matters be adjusted to their new situation, and this interval is as pernicious to industry when gold and silver are diminishing as it is advantageous when these metals are encreasing. The workman has not the same employment from the manufacturer and merchant- chant, though he pays

---

the euro was introduced, money supply in Portugal was reduced 200 times (in units of money understood as the escudo and the euro), all prices were reduced also by 200 times, and there were no real effects. The neutral effects of money, which are a characteristic of the long run, happened instantaneously. What the long run and the policy of replacing escudos with euros have in common is that both in the long run and for simple policies like a change in monetary units, the policies are well anticipated and understood.

> the same price for everything in the market. The farmer cannot dispose of his corn and cattle, though he must pay the same rent to his landlord. The poverty, and beggary, and sloth which must ensue are easily foreseen. [p. 40]

Lucas relates these two apparently contradictory statements to the quantity theory evidence on the long run effects of money and to the evidence on short run effects from Phillips curves.

The central predictions of the quantity theory are that, in the long run, there is a one-to-one relationship between average growth rate of the money supply and average inflation and that there is no relation between the average growth rate of money and real output. We will add to this the long run evidence between nominal interest rates and inflation.

Figure 1 taken from McCandless and Weber (1995) plots 30 year (1960-1990) average annual growth rates of money against annual inflation rates (first panel) and average real output growth rates (second panel), for a total of 110 countries. For inflation and money growth, the dots lie roughly on a 45^{o} line, meaning that countries with higher average growth rate of money have higher inflation by the same magnitude.[2] Similarly countries with higher nominal interest rates also have higher inflation, also one-to-one as documented in Figure 2 (first panel), taken from Teles and Valle e Azevedo (2016). For real output growth and money growth, there seems to be no relationship between the variables.

For the short run, the evidence on the effects of monetary policy is mixed. Lucas (1996), using plots of annual inflation against unemployment rates for the United States in the period between 1950 and 1994 (from Stockman, A.C. (1996)) shows that at first sight the variables are unrelated. Then, he gives it its best chance by drawing in the cloud of points a family of short run Phillips curves that would be shifting up (Figure 3). The idea is that the downward sloping Phillips curve is evidence of short run effects of monetary policy. The curves would be shifting up as those short run effects would be exploited to reduce unemployment.[3] Higher surprise inflation would reduce unemployment in the short run, but it would eventually raise inflation expectations shifting the Phillips curve upwards. Higher, and higher surprise inflation would then be necessary to reduce unemployment further, and further, inducing further shifts of the Phillips curve. The use of the short run non-neutrality of money to systematically reduce unemployment would lead to shifts to higher short run Phillips curves, leading in the long run to higher inflation. In this sense one might be able to distinguish in the cloud of points a vertical long run Phillips curve and a family of short run Phillips

2. The evidence for countries with moderate to low inflation is much less striking. Teles *et al.* (2016) provide explanations for this that are still consistent with the quantity theory, long run neutrality of money. This is the content of Box 1.

3. See Sargent, T. J. (2001) for a formal analysis of this argument.

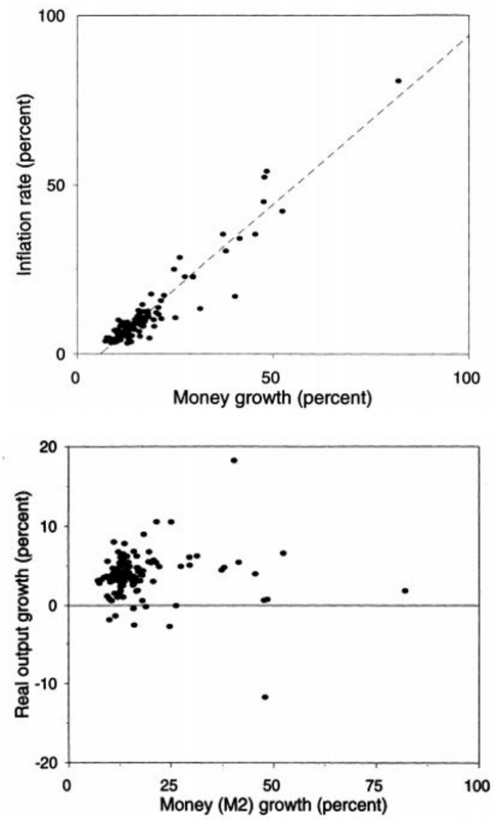

FIGURE 1: Long run money, prices and output

Source: McCandless and Weber (1995).

curves crossing it at points that over time are moving upwards towards higher inflation for some natural rate of unemployment.[4]

Extending the sample period to the more recent periods, and using the same approach where the short run Phillips curve is given its best chance[5], shows the reverse picture of shifting Phillips curves downwards (Figure 4). Not only the short run Phillips curves that appear out of the cloud of points seem to move downwards but the last three years could possibly suggest a new even lower curve.

4. The estimation of short run Phillips curves is difficult because of endogenous policy. See Fitzgerald and Nicolini (2014) for an econometric estimation of Phillips curves using regional data for the US.

5. The data breaks are hand picked to carefully try to make it work.

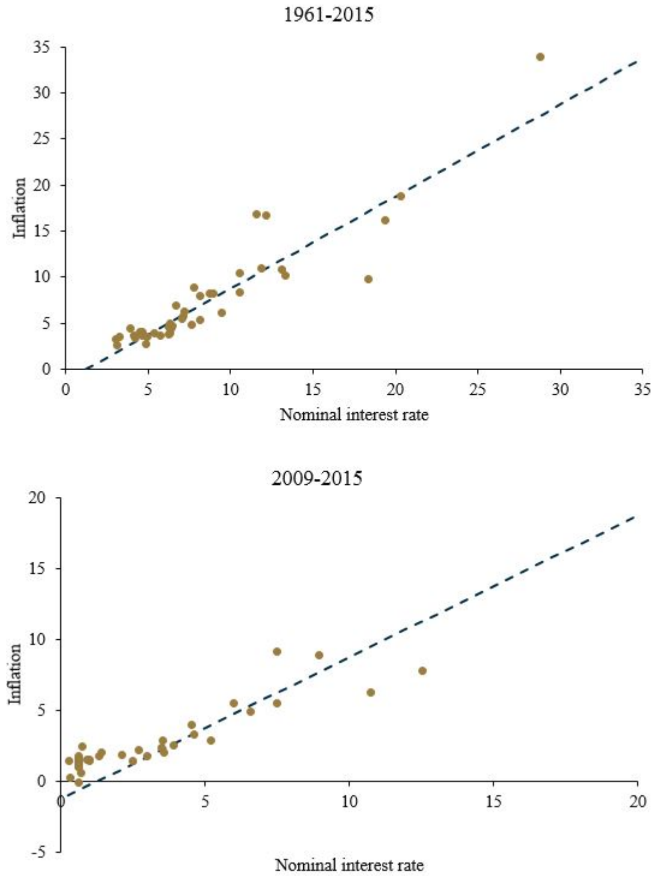

FIGURE 2: Nominal interest rates and inflation

Source: Teles and Valle e Azevedo (2016).

The story behind the movements along the short-run Phillips curve together with possible shifts of those Phillips curves, relies on a mechanism of expectations formation that adjusts to the economic context. Depending on the economic context those shifts of the short-run Phillips curves can happen at a very fast pace. Movements along the long run vertical Phillips curve can be almost instantaneous.

The picture is strikingly similar for other countries. For Germany the high inflation curves are lower than for the US but other than that they look alike (Figure 5). For Germany the last three years suggest a short run vertical Phillips curve, associated with a precipitate decline in inflation. For France there is clearly also a shift to the right towards more unemployment (Figure 6). What could explain that shift to the right? Stronger unemployment protection

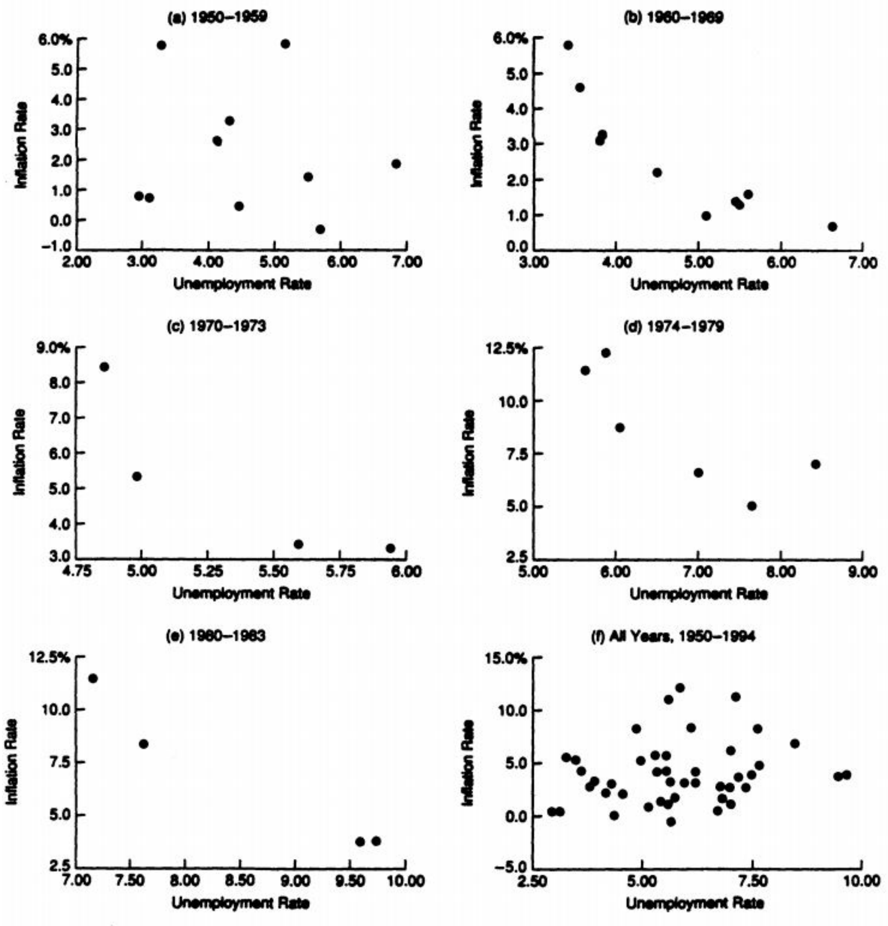

FIGURE 3: Lucas Phillips curves for the United States

Source: Lucas(1996).

and more effective minimum wages must be part of the explanation. Still the same shift downwards is clear.

Again, the picture for Japan is similar (Figure 7). Even if for Japan the whole curve looks like a Phillips curve, a more careful reading can still identify a family of curves, with similar shifts to the ones in France, where the curves seem to shift to the right and downwards, with resulting higher natural unemployment and lower inflation expectations.

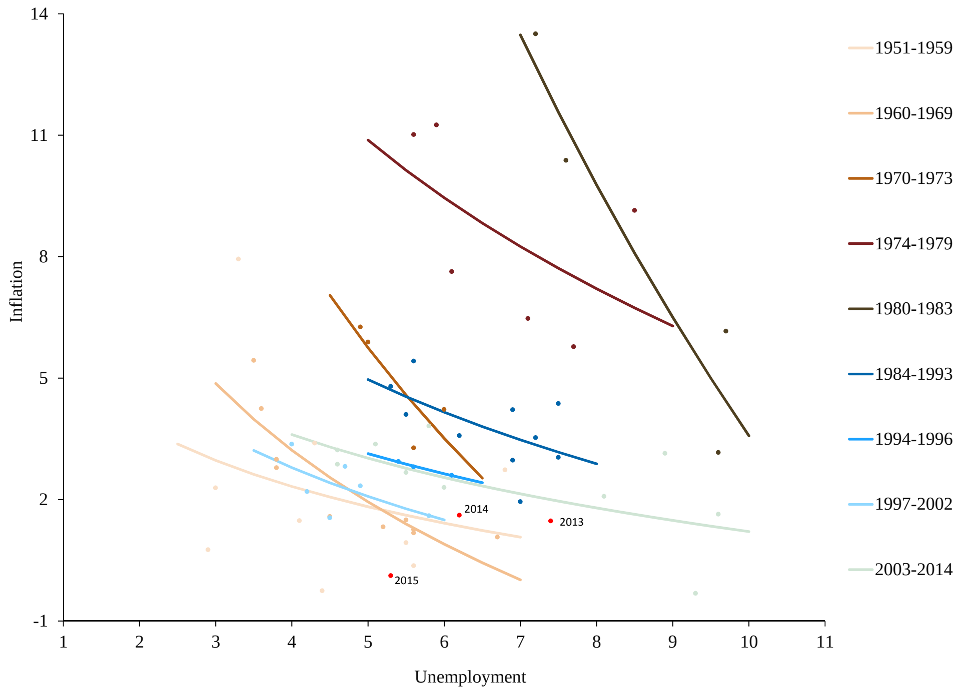

FIGURE 4: Phillips curves for the United States

Source: Bureau of Labour Statistics and own calculations.

Can the Phillips curve be used for policy?

The data on inflation and unemployment can be read as a family of downward sloping short run Phillips curves crossing a vertical long run curve. This reading is consistent with the apparently contradictory statements of David Hume. It is also the contribution of Friedman and Phelps that gave Phelps the Nobel Prize in 2006. Its formalization with rational expectations is one of the main contributions of Robert Lucas that also justified his Nobel prize. The reading is also consistent with all macro models with sticky prices or wages that are written today.

Even if there are certainly short run effects of monetary policy, and nominal frictions matter in the short run also in response to nonmonetary shocks, those effects are averaged out in the long run. In that sense, in the long run inflation is strictly a monetary phenomenon moving one-to-one with the growth rate of the money supply and with the nominal interest rate. In the long run the Phillips curve is vertical. There is some natural rate of unemployment because people take time to find jobs and firms take time to

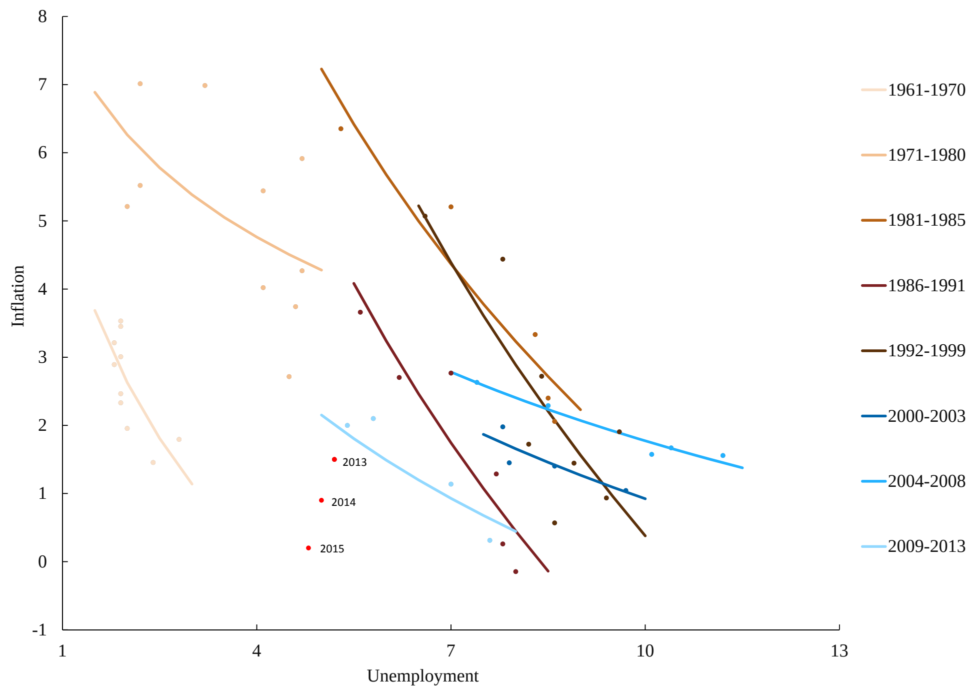

FIGURE 5: Phillips curves for Germany

Source: AMECO database and own calculations.

fill vacancies. That natural rate of unemployment is consistent with many possible levels of inflation. Inflation could be very low or very high, and only monetary policy would determine the level.

A simple quantity equation and the Fisher equation can be useful to formalize this. Because money must be used for transactions, some monetary aggregate, M, times velocity, v, equals the price level, P, times real output, Y:

$$Mv = PY$$

In growth rates, with stable velocity, this means that

$$\pi \approx \mu - \gamma,$$

where π is the inflation rate, μ is the growth rate of the money supply and γ is the long run real output growth rate. The Fisher equation will have the return on a nominal bond, i, be equal to the return on a real bond, r, plus expected inflation, π^e. This is an arbitrage condition between a nominal and a real bond, formally written as

$$i = r + \pi^e$$

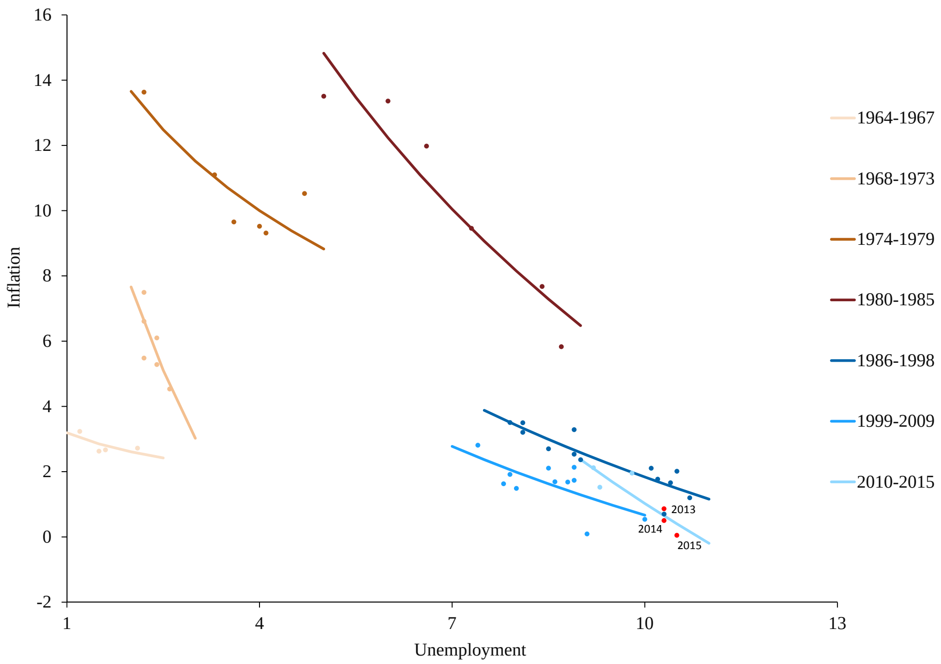

FIGURE 6: Phillips curves for France

Source: AMECO database and own calculations.

The simplest possible way to model the interaction between nominal and real variables will have the long run real growth rate, γ, and the real rate of interest, r, be invariant to monetary policy. A higher growth rate of money translates into higher inflation. A higher nominal interest rate also translates into higher inflation. Because the nominal interest rate cannot be very much below zero (otherwise only cash, that pays zero return, would be held), inflation is bounded below. But it is not bounded above.

This very simple model fits beautifully the long term data in Figures 1 and 2. A higher nominal interest rate translates into higher inflation, and growth rate of money, one-to-one.

The long run behavior of money and prices could be described by a more complete model without uncertainty and with fully flexible prices and wages. We now want to think of a world with aggregate uncertainty but without information frictions, with flexible prices and wages. In that world, the natural rate of unemployment would move over time, but monetary policy would not have short run effects. Inflation could be higher or lower, but that would have no bearing on real variables (other than through the

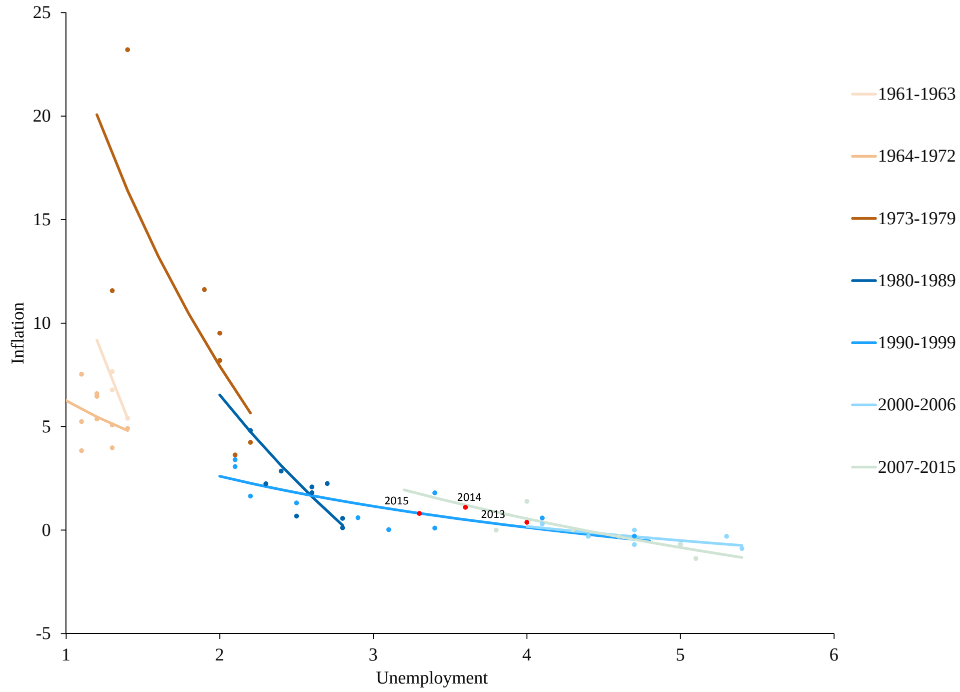

FIGURE 7: Phillips curves for Japan

Source: AMECO database and own calculations.

distortions imposed by volatile nominal interest rates). Notice that the raw data on inflation and unemployment is not inconsistent with this view. The natural rate of unemployment could be moving around in response to real shocks, and inflation could be moving around in response to both real and monetary shocks.

In particular the data could draw an horizontal Phillips curve even if prices are fully flexible. This is particularly relevant since more recent Phillips curves have very low slopes, very close to zero. The reason for an horizontal Phillips curve with flexible prices would be inflation targeting. If in a world with flexible prices monetary policy is successful in keeping inflation at a constant target, then we should see exactly an horizontal Phillips curve. Unemployment would be moving up and down, but inflation would be stable at target. As it turns out in such an environment, because it is a stable nominal environment, we have reasons to think that even if prices are sticky that price stickiness is irrelevant.

The long run Phillips curve in this context would average out the movements in unemployment and would be a vertical line at that average unemployment rate, for different targets for inflation.

Nominal rigidities and the use of the Phillips curve for policy

Now we want to give a chance to the Phillips curve as evidence for short run effects of monetary policy. One clear way to understand what these short run effects are, as well as the long run neutrality, is to read Lucas (1988) lecture "What economists do" given at a graduation ceremony at Chicago back in the 80's.[6] Basically, we are going to use Kennywood Park, the amusement park in Lucas lecture, as the model of short run effects of money.

In Kennywood Park a surprise appreciation of the currency internal to the park (or a decrease in the money supply) has negative real effects. Output goes below potential, and unemployment goes above its natural rate. But the experiment has no effect on inflation. One way there can be both a positive effect on unemployment and a negative one on inflation is by assuming that the model has two parks, one in which the appreciation takes everyone by surprise and the other where the appreciation is well anticipated. In the first park the effects would be negative on output, and positive on unemployment. In the second park the effects would be negative on prices. The joint effects would both raise unemployment and lower prices. Unemployment rises above the natural rate (and output falls below potential) and inflation falls below some reference level associated with expected or average inflation.[7] Similarly a surprise depreciation would have moved inflation above the reference level and unemployment below the natural rate, along a Phillips curve.

In what sense would there be a vertical long run Phillips curve? If every week there was a depreciation of the currency in the park, then this would just translate into higher inflation. Everyone would anticipate and understand the policies and there would be no real effects. How fast would the short run effects disappear and only the long run neutrality appear? It would probably not take long, probably not longer than a year, for both operators and patrons to realize that prices and exchange rates were moving over time in neutral ways.

We now go back to the Phillips curve data. Suppose, then, that the downward sloping Phillips curves are due to short run non-neutrality of

6. This is reproduced in Box 2.

7. If inflation is expected to be around 2%, then inflation would move below or above 2%. The reference level can be the target for inflation, but it doesn't necessarily have to be. It may be the case that expectations deviate from target, temporarily or possibly even permanently, if policy is unable to achieve the target.

money, of the type in Kennywood Park. Should policy exploit the non neutrality?[8] Lucas partially answers this question, but we can add to that answer with insights from the more recent literature on stabilization policy.

The idea of the Phillips curve is that there is some level of the natural rate of unemployment corresponding to potential output, but that the economy may be above or below potential, with more or less inflation. Potential output is the level of economic activity that would arise if the economy was not subject to nominal rigidities, such as sticky prices or wages. Shocks to technology or preferences, or in financial markets, can move potential output but they can also create gaps which are the deviations of equilibrium output from potential output. Those gaps manifest themselves not only as deviations of output from potential but also as deviations of inflation from target. When output is below potential, inflation is below target, as suggested by the downward sloping short run Phillips curve.

Monetary policy can act on those deviations of output from potential, and inflation from target. Monetary policy induces movements along the Phillips curve, stimulating the economy and thus inducing inflation. This can be achieved through policy on the money supply or on nominal interest rates. The economy can be stimulated by raising the money supply or by cutting interest rates. Why the movements in these two instruments are opposites is a much harder question to answer. We would need a more complex model than Kennywood Park in order to give a convincing answer. Since this is something no central banker has doubts about, we will just assume it here.

Other shocks, other than monetary, may also cause movements along the Phillips curve, in particular when potential output also changes, inducing also a shift of the curve to the right or left. The role of monetary policy in this context ought to be to bring the economy back to potential whenever because of other shocks, the economy is either above or below potential. In so doing, inflation is also brought back to target.

The nonneutrality of money in the short run is responsible for the gaps, but it is also the reason why monetary policy is effective in dealing with them. The more severe is the nonneutrality, the wider are the gaps created, but also the more effective policy is. As it turns out, the same policy can be used in more or less rigid environments, to deal with wider or narrower gaps, because the effectiveness of policy is exactly right to deal with those different gaps (see Adão. *et al.* (2004)). The policy that can fully deal with the gaps is a policy of full inflation targeting.

Inflation targeting can keep output at potential, or unemployment at its natural rate. Given that if inflation is stable and at target, the agents would be in a stable nominal environment, there would be no reason for nominal

8. One straightforward way to exploit the short run Phillips curve for policy is to use measures of slack to forecast inflation. This turns out not to be very useful as discussed in Box 3.

rigidities to be relevant. In that environment there would be still movements in the natural rate of unemployment, but there would be no deviations from it. The Phillips curve would be horizontal with inflation at target. Unemployment would be moving with shocks, but it would correspond to movements in the natural rate, not to deviations from it.

The efficient way to induce the movements along the curve in reaction to shocks is to use monetary policy. Fiscal policy can also be used, but conventional fiscal policy adds costs because it also changes the potential output in ways that are not desirable. If by using the money supply or the interest rate it is possible to bring the economy back to potential why building airports or roads for that purpose? Roads should be repaired when needed, not when the economy is below potential. Distributive policies should be used for distribution, not as standard macro stabilization policy.

One exception to the rule that monetary policy should be used first is when monetary policy is deprived of instruments.[9] This happens when interest rates are so low that they cannot be lowered further. As it turns out when that is the case, money supply policy also looses its effectiveness. When the nominal interest rate is very low, close to zero, the opportunity cost of money is also very low. People may just as well hold money, so that increasing the supply of money has no effects. In particular, banks may hold very high reserves at zero cost, or close to zero. Figure 8 is evidence of this.

Monetary policy can play a role in stabilizing the economy in response to shocks. This does not mean that economic fluctuations should be eliminated. It just means that the fluctuations would be the desirable ones (not the patologies that Lucas talks about in his lecture). It means that, when productivity is high, production is able to rise fully, and when productivity is low, production is able to go down fully. It may very well be the case, with this way of looking at stabilization policy, that instead of reducing economic fluctuations, policy would be increasing them.

Now, should monetary policy try to induce systematic movements along the Phillips curve in order to reduce unemployment? The model of Kennywood Park, again, helps to understand that the answer is no. Monetary policy is not very effective when used systematically. Systematic policy feeds into expectations and instead of lowering unemployment (and raising inflation) along the Phillips curve, only inflation rises. The Phillips curve shifts up and the movement is along the long run vertical Phillips curve. But there is another, more important reason not use policy to systematically increase output above potential. It is that potential output may very well be the optimal level of output, even if associated with unemployment.

9. There is fiscal policy that can mimic monetary policy and that can be used even at the zero bound (Correia *et al.* (2013))). It is not simple policy because in principle many taxes would have to be used. In a monetary union that is not fiscally integrated, a lot of explaining and coordinating, and experimenting would have to take place.

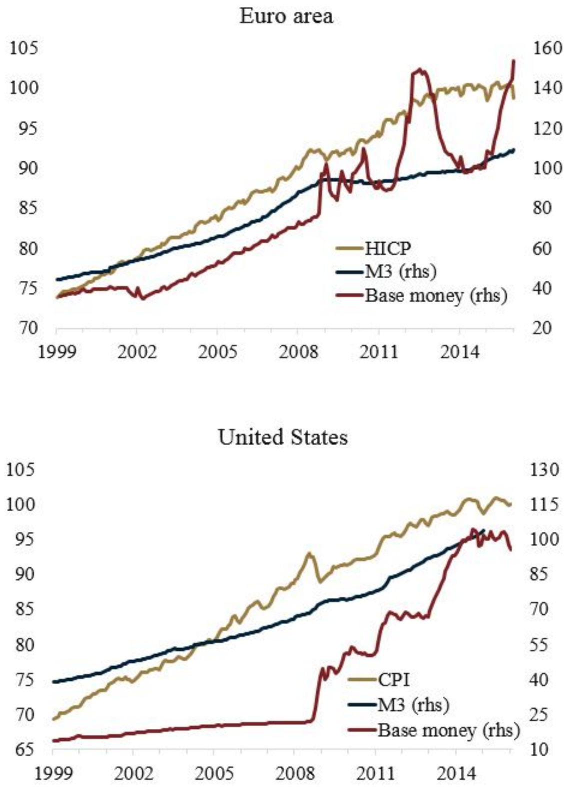

FIGURE 8: Money and inflation

Source: Bureau of Labour Statistics, ECB, Eurostat, Federal Reserve Economic Data and own calculations.

Monetary policy can also act directly on inflation by shifting upwards or downwards the Phillips curve. A higher Phillips curve corresponds to one with higher reference (average, expected, or target) inflation. That can only be supported by higher average nominal interest rates and growth rates of the money supply.[10]

Inflation is currently very low in the Euro area. The natural question to ask after this discussion is whether the low inflation is because of a movement along a Phillips curve associated with output below potential, or whether it is because of a shift downwards of the curve associated with lower inflation

10. Expectations may adapt in a way such that a shift along the curve may shift the curve. Agents that are unsure about the way policy is conducted, or are uncertain about the true model, may perceive temporary high inflation for higher average inflation, so that a movement along the curve may induce a shift of the curve.

expectations. If it is a movement along the curve there is not much monetary policy can do. If along the curve, the way to stimulate is to reduce rates, rates are already at zero and cannot be lowered further. If the answer is that the curve has shifted down, then there is a lot more that policy can do. A shift upwards of the Phillips curve with higher inflation can be supported by higher rates, and interest rates are not bounded above.

Concluding with one pressing policy question

Currently in the Euro area there is one pressing policy question that can be broken in two. The first question is whether the current low inflation is the result of a movement along a Phillips curve associated with slack in the use of economic resources. There is certainly considerable slack in the Euro area in the countries exposed to the sovereign debt crisis. If there was room to cut rates, should policy rates be cut down further in order to address that slack? Yes, most central bankers would agree. But the answer using a more complete model could very well be no. One problem with the countries exposed to the sovereign debt crisis is that savings, both public and private, were not high enough, and lower rates would reduce savings.

The slack in countries like Portugal is indeed very high. Now, is monetary policy in the context of the euro area the right way to address that slack? Countries with sovereign currencies that go through the type of external account adjustment that Portugal went through have their currency devalue up to the point where real wages in units of tradeables go down on impact by 50%. In that context what difference does European inflation of 2% make? If labor market restrictions, such as minimum wages, are adjusted to inflation to keep those restrictions active, whatever inflation could be produced, unemployment would not be reduced. In the end, the solution to the considerable slack in countries like Portugal is not a technical one, but a political one.

The second question is whether the low inflation is due to a shift down of the Phillips curve, because of persistently low nominal interest rates. The answer to this is likely to be yes. The reason is very simple. Nominal interest rates have been very low for the last eight years and they are expected to remain low for a long time. That looks a lot like the long run, when inflation and interest rates move in the same direction.

If indeed the answer to the second question is yes, how can inflation be brought back to target? One thing we have no doubts about is that eventually interest rates will have to be higher, if inflation is to return to target. What is not so clear is how fast policy rates should go up. That is why monetary policy making is such a great challenge today.

Box 1. Evidence for countries with moderate to low inflation

The relationship between average inflation and growth rate of money is not so overwhelming when attention is focused on countries with relatively low inflations. There, the picture looks more like a cloud than a straight line. Teles *et al.* (2016) show that the reason for it is that when inflation is relatively low other monetary factors play a role. They make the case that if the interest rate is larger at the beginning of the sample than at the end, one would expect that the real quantity of money would be larger at the end than at the beginning so that inflation would be lower than the growth rate of money supply in that sample period. Breaking the sample period into two they correct for this effect and see the points lining up beautifully on a 45^o line, in the first sample. The 45^o line seems to fade away in the second part of the sample, after the mid-eighties. They make the case that inflation targeting, by reducing the variability of inflation in the second part of the sample, explains why the points lie on an horizontal line rather than on a diagonal.

Box 2. What Economists Do

Robert E. Lucas, Jr. December 9, 1988

Economists have an image of practicality and worldliness not shared by physicists and poets. Some economists have earned this image. Others – myself and many of my colleagues here at Chicago– have not. I'm not sure whether you will take this as a confession or a boast, but we are basically story-tellers, creators of make-believe economic systems. Rather than try to explain what this story-telling activity is about and why I think it is a useful –even an essential– activity, I thought I would just tell you a story and let you make of it what you like.

My story has a point: I want to understand the connection between changes in the money supply and economic depressions. One way to demonstrate that I understand this connection –I think the only really convincing way– would be for me to engineer a depression in the United States by manipulating the U.S. money supply. I think I know how to do this, though I'm not absolutely sure, but a real virtue of the democratic system is that we do not look kindly on people who want to use our lives as a laboratory. So I will try to make my depression somewhere else.

The location I have in mind is an old-fashioned amusement park–roller coasters, fun house, hot dogs, the works. I am thinking of Kennywood Park in Pittsburgh, where I lived when my children were at the optimal age as amusement park companions - a beautiful, turn-of-the-century place on a bluff overlooking the Monongahela River. If you have not seen this particular park, substitute one with which you are familiar, as I want you to try to visualize how the experiment I am going to describe would actually work in practice.

Kennywood Park is a useful location for my purposes because it is an entirely independent monetary system. One cannot spend U.S. dollars inside the park. At the gate, visitors use U.S. dollars to purchase tickets and then enter the park and spend the tickets. Rides inside are priced at so many tickets per ride. Ride operators collect these tickets, and at the end of each day they are cashed in for dollars, like chips in a casino.

For obvious reasons, business in the park fluctuates: Sundays are big days, July 4 is even bigger. On most concessions –I imagine each ride in the park to be independently operated– there is some flexibility: an extra person can be called in to help take tickets or to speed people getting on and off the ride, on short-notice if the day is unexpectedly big or with advanced notice if it is predictable. If business is disappointingly slow, an operator will let some of his help leave early. So "GNP" in the park (total tickets spent) and employment (the number of man hours worked) will fluctuate from one day to the next due to fluctuations in demand. Do we want to call a slow day –a Monday or a Tuesday, say– a depression? Surely not. By an economic depression we mean something that ought not to happen, something pathological, not normal seasonal or daily ups and downs.

This, I imagine, is how the park works. (I say "imagine" because I am just making most of this up as I go along.) Technically, Kennywood Park is a fixed exchange rate system, since its central bank–the cashier's office at the gate– stands ready to exchange local currency –tickets– for foreign currency –US dollars– at a fixed rate.

In this economy, there is an obvious sense in which the number of tickets in circulation is economically irrelevant. Noone –customer or concessioner– really cares about the number of tickets per ride except insofar as these prices reflect U.S. dollars per ride. If the number of tickets per U.S. dollar were doubled from 10 to 20, and if the prices of all rides were doubled in terms of tickets–6 tickets per roller coaster ride instead of 3–and if everyone understood that these changes had occurred, it just would not make any important difference. Such a doubling of the money supply and of prices would amount to a 100 percent inflation in terms of local currency, but so what?

Yet I want to show you that changes in the quantity of money–in the number of tickets in circulation–have the capacity to induce depressions or booms in this economy (just as I think they do in reality). To do so, I want to imagine subjecting Kennywood Park to an entirely operational experiment. Think of renting the park from its owners for one Sunday, for suitable compensation, and taking over the functions of the cashier's office. Neither the operators of concessions nor the customers are to be informed of this. Then, with no advance warning to anyone inside the park, and no communication to them as to what is going on, the cashiers are instructed for this one day to give 8 tickets per dollar instead of 10. What will happen?

We can imagine a variety of reactions. Some customers, discouraged or angry, will turn around and go home. Others, coming to the park with a dollar budget fixed by Mom, will just buy 80 percent of the tickets they would have bought otherwise. Still others will shell out 20 percent more dollars and behave as they would have in the absence of this change in "exchange rates." I would have to know much more than I do about Kennywood Park patrons to judge how many would fall into each of these categories, but it is pretty clear that no-one will be induced to take more tickets than if the experiment had not taken place, many will buy fewer, and thus that the total number of tickets in circulation–the "money supply" of this amusement park economy–will take a drop below what it otherwise would have been on this Sunday.

Now how does all of this look from the point of view of the operator of a ride or the guy selling hot dogs? Again, there will be a variety of reactions. In general, most operators will notice that the park seems kind of empty, for a Sunday, and that customers don't seam to be spending like they usually do. More time is being spent on "freebies", the river view or a walk through the gardens. Many operators take this personally. Those who were worried that their ride was becoming passé get additional confirmation. Those who thought they were just starting to become popular, and had thoughts of adding some capacity, begin to wonder if they had perhaps become over-optimistic. On many concessions, the extra employees hired to deal with the expected Sunday crowd are sent home early. A gloomy, "depressed" mood settles in.

What I have done, in short, is to engineer a depression in the park. The reduction in the quantity of money has led to a reduction in real output and employment. And this depression is indeed a kind of pathology. Customers are arriving at the park, eager to spend and enjoy themselves; Concessioners are ready and waiting to serve them. By introducing a glitch into the park's monetary system, we have prevented (not physically, but just as effectively) buyers and sellers from getting together to consummate mutually advantageous trades.

That is the end of my story. Rather than offer you some of my opinions about the nature and causes of depressions in the United States, I simply made a depression and let you watch it unfold. I hope you found it convincing on its own terms–that what I said would happen in the park as the result of my manipulations would in fact happen. If so, then you will agree that by increasing the number of tickets per dollar we could as easily have engineered a boom in the park. But we could not, clearly, engineer a boom Sunday after Sunday by this method. Our experiment worked only because our manipulations caught everyone by surprise. We could have avoided the depression by leaving things alone, but we could not use monetary manipulation to engineer a permanently higher level of prosperity in the park. The clarity with which these affects can be seen is the key advantage of operating in simplified, fictional worlds.

The disadvantage, it must be conceded, is that we are not really interested in understanding and preventing depressions in hypothetical amusement parks. We are interested in our own, vastly more complicated society. To apply the knowledge we have gained about depressions in Kennywood Park, we must be willing to argue by analogy from what we know about one situation to what we would like to know about another, quite different situation. And, as we all know, the analogy that one person finds persuasive, his neighbor may well, find ridiculous.

Well, that is why honest people can disagree. I don't know what one can do about it, except keep trying to tell better and better stories, to provide the raw material for better and more instructive analogies. How else can we free ourselves from the limits of historical experience so as to discover ways in which our society can operate better than it has in the past? In any case, that is what economists do. We are storytellers, operating much of the time in worlds of make believe. We do not find that the realm of imagination and ideas is an alternative to, or a retreat from, practical reality. On the contrary, it is the only way we have found to think seriously about reality.

In a way, there is nothing more to this method than maintaining the conviction (which I know you have after four years at Chicago) that imagination and ideas matter. I hope you can do this in the years that follow. It is fun and interesting and, really, there is no practical alternative.

Box 3. The Phillips curve is not useful for forecasting inflation

A standard approach to monetary policy has the policy rate move with a forecast for inflation. Can the Phillips curve be used to improve upon that forecast for inflation? The answer is a surprising no. As is turns out, in forecasting inflation at shorter horizons, one or two year-ahead, the best forecast is current inflation. Measures of slack, that according to the Phillips curve are directly related to inflation, do not significantly improve the inflation forecast, and neither do other monetary or financial variables. One reference for these results is Atkeson and Ohanian (2001). This does not mean that the Phillips curve is not to be found in the data. It just means that measures of slack do not add information to current inflation in order to forecast future inflation.

References

Adão., B., I. Correia, and P. Teles (2004). "The Monetary Transmission Mechanism: Is it Relevant for Policy?" *Journal of the European Economic Association*, 2(2-3), 310–319.

Atkeson, A. and L. E. Ohanian (2001). "Are Phillips curves useful for forecasting inflation?" *Quarterly Review, Federal Reserve Bank of Minneapolis*, pp. 2–11.

Correia, I., E. Farhi, J. P. Nicolini, and P. Teles (2013). "Unconventional Fiscal Policy at the Zero Bound." *American Economic Review*, 103(4), 1172–1211.

Fitzgerald, T.J. and J.P. Nicolini (2014). "Is There a Stable Relationship between Unemployment and Future Inflation? Evidence from U.S. Cities." Working Papers 713, Federal Reserve Bank of Minneapolis.

Lucas, Jr, Robert E (1988). "What Economists Do." *Lecture at the University of Chicago December 1988 Convocation*.

Lucas, Jr, Robert E (1996). "Nobel Lecture: Monetary Neutrality." *Journal of Political Economy*, 104(4), 661–682.

McCandless, G. T. and W.E. Weber (1995). "Some monetary facts." *Quarterly Review*, issue Sum, 2–11.

Sargent, T. J. (2001). *The Conquest of American Inflation*. Princeton Universisty Press.

Stockman, A.C. (1996). *Introduction to Economics*. Harcourt College Pub.

Teles, P., H. Uhlig, and J. Valle e Azevedo (2016). "Is Quantity Theory Still Alive?" *Economic Journal*, 126, 442–464.

Teles, P. and J. Valle e Azevedo (2016). "On the Long Run Neutrality of Nominal Rates." *mimeo, Banco de Portugal*.

Firm default probabilities revisited

António Antunes
Banco de Portugal and NOVA SBE

Homero Gonçalves
Banco de Portugal

Pedro Prego
Banco de Portugal

April 2016

Abstract
This article describes a tool to assess the creditworthiness of the Portuguese non-financial firms. In its design, the main goal is to find factors explaining the probability that any given firm will have a significant default episode vis-à-vis the banking system during the following year. Using information from the central credit register for period 2002–2015 and a comprehensive balance sheet data set for period 2005–2014, we develop a method to select explanatory variables and then estimate binary response models for ten strata of firms, defined in terms of size and sector of activity. We use this methodology for the classification of firms in terms of one-year probability of default consistent with typical values of existing credit rating systems, in particular the one used within the Eurosystem. We provide a brief characterisation of the Portuguese non-financial sector in terms of probabilities of default and transition between credit rating classes. (JEL: C25, G24, G32)

Introduction

This article describes a tool to assess the creditworthiness of the Portuguese non-financial firms. The main goal is to find factors explaining the probability that any given firm will have a significant default episode vis-à-vis the banking system during the following year. The output of this tool is a probability of default in banking debt with a one-year horizon. This value is then mapped into a masterscale where companies are grouped into homogeneous risk classes. The fact that credit quality is assessed only in terms of banking debt is essentially not limiting our analysis for two reasons. First, most credit in Portugal is granted by banks. Only a few large firms typically issue market debt. Second, defaults in issued debt should be highly correlated with defaults in bank loans.

Acknowledgements: We thank Lucena Vieira for skilfully supplying the data, Manuel Lingo and Florian Resch (Oesterreichische Nationalbank) for sharing with us their expertise in the design of credit rating systems, and our colleagues at the Statistics Department and Economics and Research Department who helped us in this project.

E-mail: aantunes@bportugal.pt; hgoncalves@bportugal.pt; pmprego@bportugal.pt

Credit Quality Step	Upper default probability limit
1 & 2	0.1
3	0.4
4	1.0
5	1.5
6	3.0
7	5.0
8	100

TABLE 1. Credit Quality Steps within the Eurosystem. All values in percentage.
Source: ECB.

Each risk class will be labeled by a "credit rating" and in the rest of this article we will refer to a risk class using its label. A credit rating is then a synthetic indicator reflecting several features (e.g. solvency, liquidity, profitability) that measure the firm's ability to fulfill its financial commitments.

In the current exercise the Eurosystem's taxonomy will be used, where a credit rating is designated by "Credit Quality Step". Table 1 presents the different risk classes and the associated upper limits of the probability of default. See ECB (2015) for additional details.

This article is partly based on previous efforts made in Martinho and Antunes (2012), but there is a vast policy and scholarly literature on the topic (see, for example, Coppens *et al.* 2007; Lingo and Winkler 2008; Figlewski *et al.* 2012), as well as a variety of documents produced by public and private institutions, including the European Central Bank (ECB), the European Banking Authority (EBA), Fitch Ratings, Moody's and Standard & Poors.

Credit ratings are used in a variety of situations. The most obvious one relates to the banks' credit allocation process. Ratings are indeed an important tool for lenders to select the borrowers according to their predefined risk appetite and to determine the terms of a loan. A higher credit ranking usually means better financing terms, including lower costs and access to more diversified instruments such as, for instance, securities markets.

Periods of broader materialisation of credit risk, like the one recently experienced in Portugal, put even more emphasis on the relevance of the firms' credit assessment process. Data for 2015 show that the total debt of non-financial corporations in Portugal represents 115% of GDP, one of the highest values in the euro area. A considerable share of this debt is in banks' balance sheets, where non-financial corporations were responsible for close to 28% of the total bank credit (bank loans and debt securities). The quality of these credits has been deteriorating substantially over the last years, putting pressure on the banks' results and capital requirements. Between December 2008 and December 2015 the non-performing loans ratio of non-financial corporations increased from 2.2% to 15.9%. In the same period the share of companies with overdue loans rose 10 percentage points to 29% in December 2015.

Early warning systems that can help predict future defaults are therefore of utmost relevance to support, at the banks' individual level, the credit allocation process and, at the aggregated level, the analysis of the financial stability of the overall banking system. Credit ratings are useful because they allow regulators and other agents in the market to identify potential problems that may be forthcoming in particular strata of firms—for example, defined in terms of activity sector or size. This is particularly important in an environment where banks' incentives in terms of reporting accurately and consistently probabilities of defaults of firms have been challenged. For example, Plosser and Santos (2014) show that banks with less regulatory capital systematically assign lower probabilities of default to firms than banks with more regulatory capital. This underreporting then implies that, for a loan with the same firm, different banks will constitute different levels of capital.

Credit ratings can also be useful as input for stress tests in order to evaluate the impact that changes in the economic environment may have on the financial sector performance. These measures can be used to estimate expected losses within a given time frame and are therefore key instruments for the risk management of financial institutions as well as for supervisory purposes. For this last purpose, it is important as well to have a benchmark tool to validate the capital requirements of each financial institution.

The existence of independent credit assessment systems also supports investment. As investment opportunities become more global and diverse, it is increasingly difficult to decide not only on which countries but also on which companies resources should be allocated. Measuring the ability and willingness of an entity to fulfil its financial commitments is key for helping make important investment decisions. Oftentimes, investors base part of their decisions on the credit rating of the company. For lenders it is difficult to have access and to analyse detailed data about each individual company presenting an investment opportunity. These grades are used as well to design structured financial products and as requirements for inclusion of securities portfolios eligible for collateral in various operations of the financial institutions.

The existence of this kind of indicator is also important for the borrower as it can provide better access to funding. Moreover, management and company owners can also use credit ratings to get a quick idea of the overall health of a company and for a direct benchmark with competitors.

Under the Eurosystem's decentralised monetary policy framework, national central banks grant credit to resident credit institutions. In order to protect the Eurosystem from financial risk, eligible assets[1] must be posted

1. Eligible collateral for refinancing operations includes not only securities but also credit claims against non-financial corporations.

as collateral for all lending operations. The Eurosystem Credit Assessment Framework (ECAF) defines the procedures, rules and techniques which ensure that the Eurosystem requirement of high credit standards for all eligible assets is met. Credit assessment systems can be used to estimate non-financial corporations' default risk. On the one hand, this credit assessment dictates whether credit institutions can pledge a certain asset against these enterprises as collateral for monetary policy operations with the national central bank. On the other hand, in the case of eligible assets, the size of the haircut is also based on the credit rating.[2]

For economic analysis, credit ratings are particularly relevant to evaluate the monetary policy transmission mechanism and to gauge the health of quality of credit flowing to the economy through the financial system. For instance, this tool can be used to evaluate if companies with the same level of intrinsic risk are charged the same cost by the banks or if there are additional variables determining the pricing of loans. There are a number of theories explaining these differences, typically in terms of asymmetries of information or the level of bank capital (see, for example, Santos and Winton 2015, and also Plosser and Santos 2014). It is also particularly interesting to compare firms from different countries of the euro area and quantify the component of the interest rate that can be attributed to the company risk, and the part stemming from other reasons, namely problems in the monetary policy transmission mechanism or country-specific risk. The data used by credit assessment systems is also valuable to identify sustainable companies that are facing problems because of lack of finance. This information can be used to help design policy measures to support companies that have viable businesses but whose activity is constrained by a weak financial system.

For statistical purposes the use of credit ratings is straightforward. Indeed, any statistic based on individual company data can be broken down into risk classes. For example, it can be valuable to compile interest rate statistics by risk class of the companies or to simply split the total bank credit by risk classes.

In order to describe a rating system suitable for the uses described above, this article is structured as follows. First, the data are presented and the default event is defined based on the available data and appropriate conventions. Second, the methodology underpinning the rating system is described. Then a calibration exercise is performed to fine-tune the model to the credit assessment system used within the Eurosystem. Fourth, some results are presented in terms of model-estimated and observed default rates and transitions among credit risk classes. Finally, a conclusion is provided.

2. To assess the credit quality of collateral, the Eurosystem takes into account information from credit assessment systems belonging to one of four sources: (i) external credit assessment institutions (ECAI); (ii) national central banks' in-house credit assessment systems (ICAS); (iii) counterparties' internal ratings-based systems (IRB); and (iv) third-party providers' rating tools (RT).

Data

The analysis in this article uses Banco de Portugal's annual *Central de Balanços* (CB) database—which is based on *Informação Empresarial Simplificada* (IES), an almost universal database with detailed balance sheet information of Portuguese firms—and the *Central de Responsabilidades de Crédito* (CRC), the Portuguese central credit register. CB contains yearly balance sheet and financial statements from virtually all Portuguese corporate firms, both private and state owned, since 2005 until 2014, which is the most recent year available. One of the main benefits of using CB is the ability to perform the analysis at the micro level. CRC records all credit institutions' exposures to Portuguese firms and households at monthly frequency, providing firm- and individual-level information on all types of credit and credit lines. For the purpose of this analysis, the time span ranges from 2002 until 2015.

In this article only private non-financial firms with at least one relationship vis-à-vis the financial sector were considered, which for the sake of simplicity will only be referred to as firms. The main reason for the exclusion of firms with no bank borrowing is that the aim is to estimate default probabilities. In addition, on the CB side observations regarding self-employed individuals and firms that reported incomplete or incoherent data, such as observations with negative total assets or negative business turnover, were excluded. As for the CRC, only information regarding performing and non-performing loans was considered, and credit lines, write-offs and renegotiated credit were disregarded. Moreover, all firm-bank relationships below €50 and firms that had an exposure to the financial system as a whole (aggregated over all the firm-bank relationships) below €10,000 were excluded.

Default definition

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. It is possible for a single firm to record more than one default event during the period of analysis but, in order to make sure we are not biasing the sample towards firms with recurrent defaults, we exclude all observations of the firm after the first default event.

We only include firms that either are new to the financial system during the sample period (that is, firms which did not have banking relationships before 2005, possibly because they did not even exist) or have a history of three years with a clean credit record. We exclude firms that enter the CRC database immediately in default.

#	Size
1	Micro
2	Small, medium and large

#	Industry
1	Manufacturing, mining and quarrying
2	Construction and real estate activities
3	Wholesale and retail trade and the primary sector
4	Utilities, transports and storage
5	Services

TABLE 2. Size and industry groups of firms.

Source: Banco de Portugal.

Data treatment and definitions of variables

In order to increase group homogeneity, we split the sample into micro firms and all other firms (i.e., small, medium and large firms). These two groups were further divided based on the firms' classification into thirteen industry NACE groups. Some industries were bundled according to their affinity, as was for instance the case of the real estate sector and the construction sector. We ended up with five groups of industries (manufacturing, mining and quarrying; construction and real estate activities; wholesale and retail trade and the primary sector; utilities, transports and storage; services) and two groups for size (micro firms; all other firms), in a total of ten groups of firms to be used in the econometric estimations. See Table 2.

The CB database contains detailed balance sheet data of Portuguese non-financial firms. For the purpose of this analysis, only a subset of CB's variables were used. The large pool of variables can be categorised into specific groups such as leverage, profitability, liquidity, capital structure, dimension, and a residual group which corresponds to variables related with the balance sheet ratios that do not fit in any of the groups previously defined. All the level variables are scaled by dividing them by either the firm's total assets, current liabilities or total liabilities, depending on the case. We never use denominators that can have negative values as that would create significant discontinuities when the denominator is close to zero. To account for the possible influence of the economy as a whole on a specific firm, we consider a small set of macro factors: nominal and real GDP growth, total credit growth and the aggregate corporate default rate. This choice was motivated by previous literature on the topic; for example, Figlewski *et al.* (2012) have found that real GDP growth and the corporate default rate help explain transitions across rating classes. Table 3 summarises the subset of CB variables and the macro factors used in this analysis.

Measures of:	Variables
Leverage	Financial debt; Bank debt; Interest paid
Profitability	Value-added per worker; Profit / Loss; EBIT; Cash flow; EBITDA
Liquidity	Cash; Current liabilities
Capital structure	Equity; Current assets; Tangible assets
Dimension	Total assets; Age; Turnover; Employees
Other idiosyncratic	Wages; Trade debt
Macroeconomy	Aggregate default rate; Credit growth; Nominal GDP growth; Real GDP growth

TABLE 3. Summary of variables used in the regressions.

Source: Banco de Portugal. Precise definition of variables available upon request.

As previously mentioned, firms that had negative total assets, liabilities or turnover were removed from the analysis. Additionally, firms with total assets, turnover or the number of employees equal to zero were excluded. In order to cope with values for skewness and kurtosis far from what would be expected under the Normal distribution, strictly positive variables were transformed into their logarithms in order to reduce skewness. Because this transformation is not applicable to variables that can be negative, the set of variables was expanded with the ranks of all variables normalised between 0 and 1. The rank transformation was applied within each year-size-industry group to increase homogeneity. A final group of well-behaved variables was kept unchanged. This included variables expressed in shares and macro variables.

Methodology

In this study, we develop an approach based on a multi-criteria system of variable selection out of a large pool of potential variables. We build upon the methodology used by Imbens and Rubin (2015) of explanatory variables selection through maximum likelihood estimation. This methodology selects variables in an iterative process based on the explanatory prediction power that each variable is able to provide. A variable under scrutiny will be included if the increase in explanatory power is above a certain threshold. We adapt this approach for our own purposes.

Selection of explanatory variables

More specifically, we start by estimating a base model with fixed effects for size (only for non micro-sized firms) and for activity sector (at a disaggregation level of a few sectors per industry). For each variable of the initial pool of N variables, we estimate a model with the fixed effects plus that variable. These regressions will then be compared to the base model by using a likelihood ratio (LR) test. The algorithm then picks the variable associated to the model with the highest likelihood statistic under the condition that it is above the initial likelihood at a 5% significance level; this corresponds to an LR ratio of at least 3.84.

The process is then repeated but the base model is now the model with the fixed effects plus the variable picked in the previous step. The next variable is to be chosen among the remaining pool of $N - 1$ variables, but from this second step on we add criteria other than the requirement in terms of the LR. These criteria address potential problems stemming from a completely agnostic inclusion of variables. More specifically, the following conditions are added in order for the candidate variable to be included in the model:

1. It must have linear and non-linear correlation coefficients with any of the variables already present in the model lower than 0.5. This condition aims at avoiding potential problems of multicollinearity.
2. It has to be statistically significant at the 5% level in the new regression, while all of the previously included variables must remain statistically significant. This is to avoid that non significant variables survive in the final model specification.
3. It has to be such that the new model estimate improves the AUROC criterion[3] relative to its previous value. In addition, the new model estimate also has to improve the AIC information criterion. This condition addresses the potential problem of over-fitting the model, as this criterion penalises the inclusion of parameters.

The process ends when none of the remaining variables in the set of potential variables fulfills all the conditions 1–3 or, to avoid the proliferation of parameters, a maximum of ten variables has been reached. In order to maintain the approach as replicable and as simple as possible, a Logit specification was chosen.

All ten models (one for each combination between two size categories and five industries) were estimated by pooling the existing observations together, spanning the period from 2005 to 2014 in terms of the balance sheet information. All explanatory variables pertain to the end of the current year t. The dependent variable is defined as an indicator of the default event during year $t + 1$. Note that when the restriction on the maximum number of variables is removed none of the ten models includes more than 13 variables. Moreover, when analysing the evolution of the AUROC with each variable added it

3. AUROC stands for "area under the Receiver Operator Characteristic". See Lingo and Winkler (2008) and Wu (2008) for the definition and the stochastic properties of this synthetic measure.

is possible to see that this benchmark tends to flatten out before the tenth variable; see Figure 1.

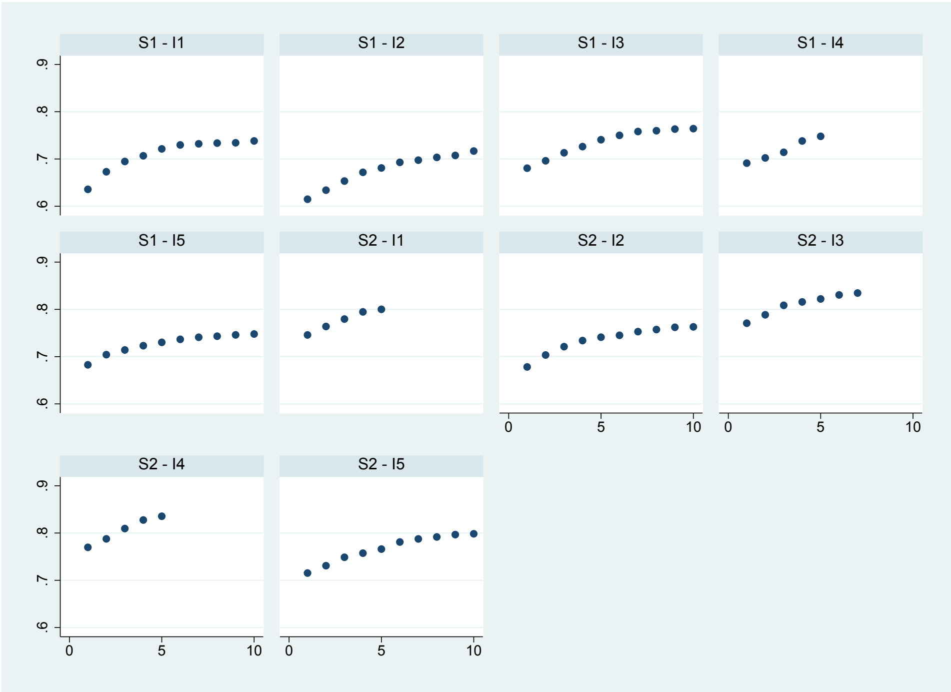

FIGURE 1: The AUROC as a function of the number of variables selected according to the methodology defined in the text. S# means size group # and I# means industry #; see Table 2 for details.

Source: Banco de Portugal and authors' calculations.

A summary of the results

After applying the proposed methodology to our data set, we obtained ten estimated Logit models; Table 4 displays some information characterising them.[4] A first observation is the overall consistent goodness-of-fit, which can be gauged by the AUROC.[5] These values lie in the range 0.72–0.84 and reject comfortably the hypothesis that the models are not distinguishable from

4. In practice we did not use the original variables, except in cases where they represented shares or growth rates, because the algorithm always chose the transformed variables (logarithm or rank).

5. For a critique of the AUROC as a measure of discriminatory power in the context of model validation, see Lingo and Winkler (2008).

Group	Obs.	Defaults	Def. ratio	# variables	AUROC	Brier Score
S1 - I1	58063	3000	5.17%	**10**	0.738	0.047
S1 - I2	53543	2965	5.54%	10	0.717	0.050
S1 - I3	178178	7696	4.32%	**10**	0.764	0.039
S1 - I4	2681	121	4.51%	5	0.748	0.041
S1 - I5	123048	5336	4.34%	**10**	0.748	0.040
S2 - I1	98065	3887	3.96%	5	0.800	0.035
S2 - I2	58325	3861	6.62%	10	0.763	0.057
S2 - I3	96738	3062	3.17%	7	0.835	0.028
S2 - I4	3903	128	3.28%	5	0.836	0.030
S2 - I5	73782	2476	3.36%	10	0.798	0.031
Overall	746326	32532	4.36%	n.a.	0.777	0.0393

TABLE 4. A summary of the Logit estimations for ten strata of firms. Values in bold mean that the procedure was stopped due to the limit on explanatory variables. S# means size group # and I# means industry #; see Table 2 for details.

Source: Banco de Portugal and authors' calculations.

a random classifier. Also, in each model the Brier score, a measure of the goodness of fit, is considerably small. The Spiegelhalter (1986) test applied to each model (not reported) also indicates that the level predicted for the probability of default is consistent with the observed defaults.

Although the methodology includes ten separate models there are several similarities among them. Table 5 presents a summary of the variables more often chosen using the procedure described above. Most importantly, the different models seem to have a core group of variables, even if they enter different models in slightly different variants: for instance, cash to total assets or cash to current assets as a measure of liquidity are always chosen, although they are never chosen together for the same model.

All ten models include a measure for profitability, alternating between cash-flow to total assets or earnings to total assets, and a measure for liquidity. Nine out of the ten models include the cost of credit as well as short-term liabilities, measured by current liabilities to total assets. Eight models include a measure for leverage and seven models include the weight of the employees' wage bill to total assets. Seven models select one macro factor among nominal GDP growth, total credit growth and the aggregate default rate. Finally, six models include the age of the firm and five models include a proxy for the firm's productivity as measured by value-added per worker.

Curiously, the weight of trade debt to total liabilities is also selected for five different models, all of them pertaining to micro-sized firms. This indicates that for this group of firms the behaviour of suppliers is particularly important.

Variable	# times selected	Avg. rank	Sign in regs.
r(Cash flow / Total assets)	6	1.0	-
r(Earnings / Total assets)	4	1.0	-
r(Interest paid / Financial debt)	9	3.1	+
r(Current liabilities / Total assets)	5	3.4	+
r(Age)	5	4.4	-
r(Wages / Total assets)	5	4.6	-
r(Cash / Current assets)	6	6.0	-
r(Financial debt / Total assets)	5	5.6	-
log(Current liabilities / Total assets)	4	4.5	+
r(Cash / Total assets)	4	5.8	-
r(Trade debt / Current liabilities)	5	7.2	+
log(Financial debt / Total assets)	2	3.0	+
r(Value-added per worker)	5	7.8	-
Nominal GDP growth	3	6.3	-
Credit growth	2	5.0	+
Agg. Default Rate	2	5.0	+
log(Equity / Total assets)	1	3.0	-
r(Bank debt / Total assets)	1	4.0	+
r(Employees)	2	8.5	+
log(Wages / Total assets)	2	9.0	-
log(Turnover)	2	9.5	-
log(Age)	1	7.0	-
r(Current assets / Total assets)	1	10.0	-

TABLE 5. Qualitative results of the variable selection procedure. r(·) denotes the rank of the variable within the size-industry group in the current year; log(·) denotes the natural logarithm of the variable. The second column contains the number models for which the variable is chosen by the variable selection algorithm (out of a maximum of ten models). The third column contains the average rank (1 is first, 10 is tenth) with which the variable was chosen. The fourth column contains the sign of variable's coefficient in the Logit estimation of the default event. Variables are ordered by the product of the inverse of number of times chosen by the average rank, in ascending order. S# means size group # and I# means industry #; see Table 2 for details.

Source: Banco de Portugal and authors' calculations.

Another significant result is that the variables that are more often chosen by the algorithm are also among the first variables to be selected, which indicates that these variables have the largest contribution to the explanatory power of the model. In particular, the variables measuring profitability are the first to be picked by the algorithm in the ten different models.

Another important observation is that the coefficient of each variable always enters the model with the sign that would be expected, even though the algorithm does not impose any restriction to this effect. Moreover, when a variable is selected for more than one model the variable's coefficient sign is the same across those models.

Rating class calibration

The next step in the setup of a rating tool system is to calibrate the model so that observed default rates of firms at any given credit category are consistent with the typical default rates used to define them (see Table 1). This step is usually needed because, while the average of the conditional model-estimated default probability should match the observed average default rate, this need not be so across different groups of firms, and in particular across rating classes. One basic requirement for the calibration that we want to perform is that overall the observed default rate is consistent with the conditional default rate stemming from the estimated models. While this requirement is generally fulfilled in-sample, one question remains: is the model conditional default probability consistent also across different categories of risk?

To answer this question, let us first define the concept of *z-score* in the context of our analysis. The Logit model used in the methodology described above is framed in terms of an unobserved latent variable which is then transformed into a number between 0 and 1, the probability of default. To keep the analysis simple, it suffices to say that the coefficients β of each one of the Logit models are estimated so that the probability of default is, to the extent possible, accurately given by

$$\Pr\{\text{default}_{t+1} = 1|x_t\} = \frac{1}{1 + e^{-x_t\beta}}$$

where default_{t+1} is an indicator of a default event occurring in year $t+1$, x_t is a (row) vector of regressors in year t—including a constant and variables characterising the firm and possibly the economy—and β is a (column) vector of coefficients. It is a property of these coefficients that the in-sample average of the predicted default rates (as computed by the equation above) is equal to the observed average default rate. The z-score of each observation is simply defined as the estimated value of the latent variable, that is, $z_t = x_t\beta$.

The answer to the question above is broadly positive. Figure 2 depicts the model-predicted default probabilities (the dash-dotted curve) along with average observed default rates (the dots in the graph). Each point represents the fraction of defaults for groups of firms with relatively similar z-scores. The lower (more negative) the z-score, the lower the estimated probability of default of the firm. We can see that using a Logit specification does a good job explaining the relationship between z-scores and observed default probabilities for groups of firms across the whole z-score distribution.

One way to try to improve the fit is to have a more flexible approach. While this procedure is not consistent with the estimation process, we view that as a fine-tuning exercise rather than something that invalidates the results obtained using regression analysis. The solid line is one such attempt: it is a semiparametric curve interpolating the dots. It is readily seen that the two curves (the Logit and the semiparametric) are really telling the same story, but the semiparametric one lies above the Logit for very negative z-scores. This means that, for that range of z-scores, the semiparametric curve is going to be more conservative in assigning probabilities to firms.

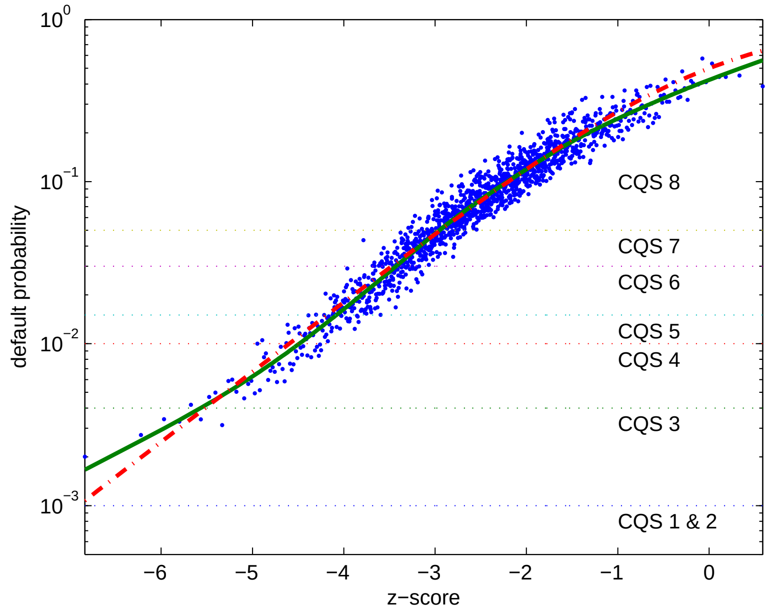

FIGURE 2: Probabilities of default of firms. Each dot represents the observed default rate for groups of firms with similar z-scores. Upper limits for default probabilities of each Credit Quality Step as defined by the Eurosystem also depicted.

Source: Banco de Portugal and authors' calculations.

We now provide additional details on the procedure of fitting the semiparametric curve to the dots, but the reader uninterested in mathematical details can safely skip the following section.

Fitting the dots

The dots in Figure 2 are empirical probabilities of default for groups of observations in the sample. Each dot in the graph represents a pair from the set of points $S^n = \{(\hat{d}^n_q, \hat{z}^n_q)\}_{q=1,\ldots,Q^n}$. These points were obtained as follows. First we sorted in ascending order all the z-scores (which are normalised and can be compared across the different groups of firms) of the sample. We then identified the first n defaults and set r^n_1 as the order number of the observation with the n^{th} default. We grouped these observations in set $A^n_1 = \{z_1, \ldots, z_{r^n_1}\}$. We then computed the ratio $\hat{d}^n_1 = \frac{n}{\#A^n_1}$ and defined $\hat{z}^n_1$ as the median of set A^n_1. We repeated the procedure for the next group of n defaults by finding set $A_2^n = \{z_{r_1^n+1}, \ldots, z_{r_2^n}\}$, default rate $\hat{d}_2^n = \frac{n}{\#A_2^n}$ and median z-score $\hat{z}_2^n$. This process was carried out in a similar fashion until we exhausted all the observations, ending up with a total of Q^n pairs of empirical default rates and z-scores. Notice that, for all q, $\hat{z}_{q-1}^n \leq \hat{z}_q^n \leq \hat{z}_{q+1}^n$, that is, these points are also sorted in ascending order in terms of the z-scores, although not necessarily in terms of default probabilities. Not all points were plotted in Figure 2; only a representative sample was.

One word about the choice of n. If this number is too small then the standard deviation of the estimated empirical probability will be relatively high. To see this, assume that the default event has a Binomial distribution within A_q^n, and take $\hat{d}_q^n$ as an estimator for the default probability. Then, an estimate of the standard deviation of $\hat{d}_q$ would be

$$\sqrt{\frac{\hat{d}_q^n(1-\hat{d}_q^n)}{\#A_q^n - 1}}$$

which decreases with $\#A_q^n$. We picked $n = 23$ in our simulations because, due to the relative scarcity of very negative z-scores (associated to relatively low probabilities of default), we wanted to have meaningful estimates for default rates even in high rating classes. With this choice we ended up with Q^{23} close to 1400. We later address the significance of the estimates obtained with this choice. The robustness of the general results of this analysis with respect to this choice is performed elsewhere. For commodity we will drop n from the notation described above.

In order to keep the analysis as standard and simple as possible, we fitted a smoothing spline to the points in the figure. The smoothing spline is a semiparametric curve that approximates a set of points in a graph while penalising the occurrence of inflexion points along the whole curve. More specifically, we chose the following specification:

$$s(\cdot) = \arg\min p \sum_{q=1}^{Q} (\log(\hat{d}_q) - s(\hat{z}_q))^2 + (1-p) \int_{\hat{z}_1}^{\hat{z}_Q} (s''(z))^2 dz\,.$$

In this formulation, function $s : [\hat{z}_1, \hat{z}_Q] \to]-\infty, 0]$ is a cubic spline defined over the set of points in S. A cubic spline is a set of cubic polynomials defined in intervals and "glued" together at the unique z-scores contained in S. By construction, $s(\cdot)$ has continuous second derivative $s''(\cdot)$ in all points. Parameter p governs the smoothness of the interpolating curve. If p is close to 1, one gets the so-called natural cubic interpolant, which passes through all the points in S.[6] If p is close to 0, the penalisation of the second derivative

6. Technically, if there are points in S with the same z-score, the natural interpolant passes through the average of the log default rates among all the points with the same z-score.

ensures that the solution will be the linear interpolant, which has zero second derivative.

The curve of the smoothing spline with $p = 0.3$ is depicted in Figure 2 as the solid line.

One thing that is clear from Figure 2 is that the empirical default probability will still be a noisy measure: while each point represents the median z-score for the set of observations leading to a given number of observed defaults (23 defaults), it is possible to have groups of very similar firms—in the sense they have very similar z-scores—and still observe relatively different observed default rates among those groups of firms. That concern is addressed by the models' performance in terms of the AUROC, which has already been presented. In any case, the general shape of the cloud of points tells us that the analytical framework captures well the probability of default across firms: a random model would yield a cloud coalescing along an horizontal line in the graph at the unconditional observed default rate. The figure then underlines that even when large AUROC measures can be obtained, the default event is still a very uncertain event.

Defining credit quality classes

The general approach chosen for the purpose of categorising firms in terms of credit default classes is (i) to obtain reference values for default probabilities from external sources, then (ii) to choose thresholds in terms of z-scores for the different credit classes, and finally (iii) to check *ex post* the observed in-sample default probabilities' consistency with the previously defined credit classes. We also provide a more detailed analysis of the transitions of firms across credit categories and to default.

We now turn to the question of defining credit quality classes. The horizontal dashed lines of Figure 2 represent upper limits of credit classes according to the Eurosystem credit quality system (see Table 1). For example, class 3 corresponds, in the standard framework of monetary policy, to the lowest-rated firms whose loans can still be posted as collateral by financial institutions for monetary refinancing operations with the Eurosystem. Instead of using the Logit curve to compute conditional probabilities—which is depicted as the dash-dot curve in the graph—we adopt a semiparametric approach and fit a smoothing spline to this set of points. Additional robustness exercises were performed but are not reported here in terms of the parameters of smoothing spline.

Comparing the semiparametric curve with the Logit curve in Figure 2, we see that for the lowest estimated default probabilities for which we have data in the sample the smoothing spline is more conservative in terms of credit class classification, while over the mid-range of z-scores the Logit is slightly more conservative. For higher estimated default rates, the two curves are equivalent, and for the highest estimated default probabilities the Logit is again more conservative than the smoothing spline.

The strategy followed here will be to use the intersections of the smoothing spline with the upper limits of the credit classes as classification thresholds in terms of z-scores.[7] These values can be observed in Figure 3, where we also depict the upper value of the probability within the class.

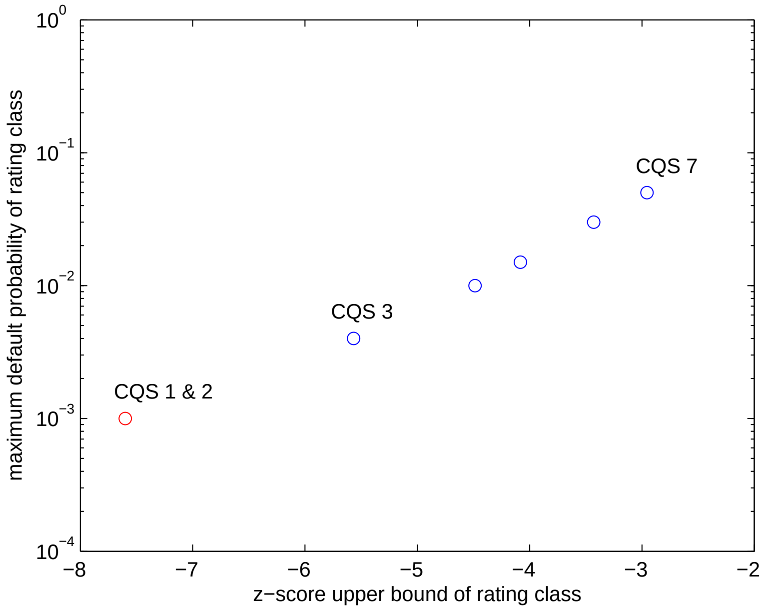

FIGURE 3: Thresholds in terms of z-scores defined according to the text.

Source: ECB, Banco de Portugal and authors' calculations.

Two observations are important at this point. First, it is clear that even with this strategy a post-classification evaluation of the method is warranted. This is because the thresholds define classes in terms of z-scores but if the observed default rates are too noisy they will have no discrimination power relative to adjacent classes. The fact that the dots represent a relatively smooth function of the probability of default with respect to the z-score gives us confidence about the capacity of the classification method to produce reasonable results.

Second, it is not possible to classify firms with credit rating classes with default probabilities below a certain value, that is, above a certain credit rating. The reason for this is the scarcity of observations classified in lower risk classes. For example, the upper limit of the default probability admissible for a

7. For class 1 & 2, the intersection was extrapolated. More on this below.

firm with a Credit Quality Step 1 would be[8] about 0.03% during one year. This means that we need approximately 67 thousand observations classified with that rating to expect observing 20 defaults.[9] If we cannot classify this number of firms with such rating in our sample, we also cannot be sure that those firms really have a probability of default compatible with the step 1 rating. Even if we are willing to lower the number of expected default events to, say, 5, we still need 17 thousand observations. In practice, for our data set we found that thresholds up to class 2 are possible: this is one class above the highest credit class for which it is possible to consistently estimate default rates. This point can made by noting that, using the notation previously introduced, $\hat{d}_1^{23} = \frac{23}{11486} = 0.002$, that is, the first 23 defaults occur for the best 11,486 z-scores. This default rate is significantly lower than the upper limit of credit class 3, and above the upper limit of credit class 2.[10] Using the fitted curve of Figure 2 to extrapolate one class above (in terms of rating) class 3 seems reasonable. For this reason we lumped Credit Quality Steps 1 and 2 into the class labeled "1 & 2". In Figure 4 we have depicted observed default rates for each class using the thresholds shown in Figure 3. Also represented are the upper default probability limits of each credit class. Since we are using a conservative approach in defining the thresholds, we see that, for all classes except class 1 & 2, the observed default rates are lower than the upper limit of each class. Moreover, assuming within-class binomial distribution[11] the lower bound of the 90% confidence interval of the default rate lies above the upper limit of the class immediately to its left (that is, with better credit quality) and the upper bound lies below the upper limit of the class.

Classes with few observations

Class 1 & 2 merits a special reference. Out of a sample of more than 740 thousand firm-year observations spanning the period 2005–2014, the above methodology allows us to classify 1177 observations in class 1 & 2. Out of these observations only two were defaults. This means that the statistical significance of the empirical default rate is low: one more or one less default would change considerably the observed default rate of the class. In Figure 4, this can be seen by the wide 90% confidence interval, whose lower limit is 0 and higher limit is 0.35%, assuming a binomial distribution of defaults within

8. This would be roughly equivalent to ratings of AA- and above (Fitch and Standard & Poors) or Aa3 and above (Moody's).

9. That is, $20 \times \frac{1}{0.0003} \approx 67,000$ observations.

10. Assuming a binomial distribution, the lower and upper limits of the 90% confidence interval of $\hat{d}_1^{23}$ are 0.13% and 0.27%, respectively.

11. Under the binomial distribution, the observed default rate of a given class is the maximum likelihood estimator of the default rate.

the class. This also means that we do not reject the null hypothesis that, under a binomial distribution, the actual probability of default is lower than 0.1%.

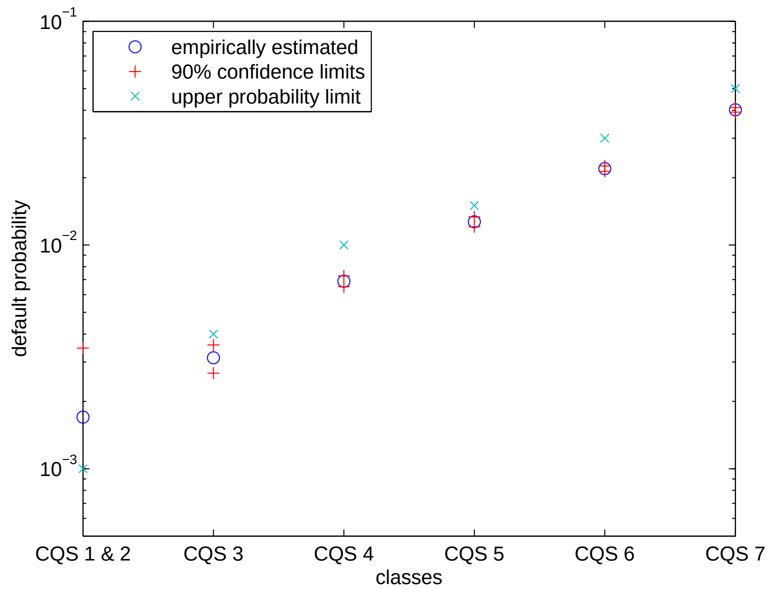

FIGURE 4: Observed default probabilities across classes using the thresholds in terms of z-scores defined according to the text. Confidence intervals are estimated assuming that within each class the default event follows a binomial distribution. Upper limits for default probabilities of each Credit Quality Step as defined by the Eurosystem also depicted as dashed horizontal lines.

Source: ECB, Banco de Portugal and authors' calculations.

All in all, one would assume that the model should be able to reliably distinguish firms in terms of all credit categories, with the best class being a residual class that lumps all high credit quality observations. The discriminating power of the model is limited by the number of observations in each class; we deem it reasonable to classify firms up to class 2. In the next section we perform an analysis of transitions of firms across classes and to default.

Some results

We now present some of the results of the rating system applied to our data. The results are consistent with the observation from Figure 2 that the z-scores seem to be effective in distinguishing firms in terms of their propensity to default.

Credit risk dynamics

Transition tables are a useful way to characterise the dynamics of firms across rating classes and to default. These tables typically contain the probability of moving to a specific credit rating class or to default, conditional on the current rating class. Table 6 contains some general statistics of our sample, including the observed default rates conditional on rating class and also exits from the sample.

Overall, we see that the default rates across classes vary considerably but are close to both their model-predicted values and the upper limit of the respective class, as seen in Figure 4. Class 8 is the most prevalent, while unsurprisingly the least numerous one is class 1 & 2, which accounts for about 0.16% of the sample. Applying the Spiegelhalter (1986) test within each class allows us not to reject (with the exception of class 8) the null that all model-estimated default forecasts match the true but unknown probability of default of the firm.[12]

As for exits without default from the sample, values vary between 11% and 18%, with an overall mean of 13.8%. These transitions are defined as permanent exits from the sample due to any of the following situations, all of them without any registered default: (i) exit from activity by merger, acquisition or formal extinction; (ii) the firm's loans are fully amortised; (iii) at least one of the regressors selected in the Logit model is not reported by the firm. Defaults can always be detected even if the firm ceases to report to CB because banks still have to report any non-performing loans by legally existing firms. These numbers compare favourably with similar measures found in the literature. For example, Figlewski *et al.* (2012) reports that, out of a sample of about 13,000 observations, the withdrawal rate was 33%.

Over time, the model-estimated default probabilities follow reasonably well the observed default rates. A notable exception is 2009, when observed default rates were considerably higher than what the respective credit risk class would suggest. This was a widespread phenomenon. See, for example, Chart 14 in Vazza and Kraemer (2015). In Table 7 this can be assessed by the differences in observed default rates in year t and the predicted default rates in year $t-1$ for year t. We see that most of the variation is due to the highest risk class, where the construction and real estate industry and the micro firms are over-represented (see Table 9 below).

Table 8 reports the overall transition matrix, which contains the share of firms migrating from one risk class to another in the subsequent year, conditional on non default and non exit. The table shows that in 3 out of 7 classes the majority of firms remained in the same risk class. It is also seen that

12. For class 8 we indeed reject the null at 5% significance. The average model-estimated default rate is 10.0% while the observed value is 10.3%. See Table 6.

the large majority of firms either stayed in the same category or moved only one category up or down. In addition, notice that, conditional on non default and non exit, firms were more likely to be downgraded than to be upgraded, except class 8 for obvious reasons.

The Markovian structure of the matrix allows us to compute a long-run distribution across credit classes (called the "ergodic" distribution). This would be the distribution prevailing in a year in the distant future if the rate at which firms entered and left the data set were those observed in the sample. It turns out that such distribution is remarkably similar to the actual shares of firms observed in Table 4. This suggests that the sample is a reasonable representation of the long-run dynamics of firms across credit rating classes.

One thing that is important to note is the relatively low persistence of credit class categories that emerges with this tool. The average persistence of a firm in the same class is much smaller than the persistence observed by ratings from rating agencies. For example, Vazza and Kraemer (2015) document that, out of 7 credit risk categories, the average fraction of firms staying in the same credit category is 87%; the comparable number in our sample is 45%. There are at least two reasons for this.

First, rating agencies typically produce ratings for relatively large corporations that have strong incentives to be rated, while in our case all firms are *ex ante* included in the sample. Moreover, several strategic considerations could bias the persistence values. While typically credit rating agencies follow firms even when they are no longer rated to detect potential defaults, firms that are currently rated might have an incentive to withdraw the rating if they suspect they will be downgraded. The other two possibilities—rating unchanged or upgrade—do not induce such a powerful incentive. This strong selection bias of the static pools of rating agencies, while not affecting the transitions to default—as ratings are conditional on the actual balance sheet of firms—would tend to produce much more persistent ratings than a rating tool that potentially includes all firms.

Second, ratings agencies and also other rating systems (such as Banco de Portugal's ICAS, currently applied to mostly large Portuguese corporations) typically involve dedicated analysts which have some latitude in adjusting the ratings coming from the statistical models underlying the system. This could also be a origin of more persistent ratings as the analyst would be reluctant to change the rating if, for example, the newly computed probability of default were marginally outside the range of the previous rating. No such adjustments are done here and even minor changes in the model-estimated default probabilities could entail changes in credit risk category.

Table 9 presents the model-estimated probabilities of default versus the empirical probabilities of default separately for each industry group and for each size category, as well as the share in terms of observations of each risk class in the group. When compared to the other sectors, the table shows that the construction and real estate sectors (industry 2) have a particularly high average default probability. This result is observed both in the comparison of estimated and empirical default probabilities and in the shares of each class. Class 8 is more than twice as large as any other risk class in this specific industry group.

Relatively risky are also micro-sized firms (size 1), none of which is considered to be in class 1 & 2 while about 74% of them are concentrated in the three worst risk classes. In contrast, about 57% of larger firms (size 2) are in the three worst risk classes.

The table shows that the five industries are generally skewed to riskier classes, particularly classes 6 and 8.

Additional validation

It is outside the scope of this article to present a detailed characterization of the method's performance out-of-sample and validation exercises. For a simple approach to this issue, the interested reader is reported to, for example, Wu (2008). Aussenegg *et al.* (2011) and Coppens *et al.* (2016) and references therein provide more advanced material.

Conclusion

The aim of this article is to present a method to assess the creditworthiness of the Portuguese non-financial firms by estimating the probability that any given firm will have a significant default episode vis-à-vis the banking system during the following year. The outcome of the model is then mapped into a masterscale where companies are grouped into homogeneous risk classes, originating a synthetic indicator of the firm's ability to fulfill its financial commitments.

By merging balance sheet information from 2005 until 2014 with credit register information from 2002 until 2015 we were able to estimate ten different models with good explanatory power in terms of the default risk of a firm. With the exception of class 8, the model-estimated default probabilities are not statistically different from the observed default probabilities.

The results also show how firms are mostly allocated to higher risk classes, with some industries and firm size classifications not represented in the lowest risk class. As expected, micro-sized firms have, on average, estimated and observed default probability higher than larger firms. The same can be seen for the construction and real estate sectors when compared to the rest of the industry sectors.

With respect to the dynamics in the transition tables presented, we can see that, from one year to the next, most firms remain in the same risk class or move to an adjacent class. Moreover, the overall transition table also seems to indicate that our model is a fairly good representation of the long-run risk distribution of the Portuguese non-financial sector.

Finally, it should be stressed that the available data do not allow us to classify firms beyond a certain credit quality. This is due to the scarcity of observations for the lower risk classes. For a finer classification among high ratings it is necessary to include professional analysts in the process and, perhaps, resort to more structural models of default as opposed to statistical approaches like the one followed here.

References

Aussenegg, Wolfgang, Florian Resch, and Gerhard Winkler (2011). "Pitfalls and remedies in testing the calibration quality of rating systems." *Journal of Banking and Finance*, 35(3), 698–708.

Coppens, François, Fernando González, and Gerhard Winkler (2007). "The performance of credit rating systems in the assessment of collateral used in Eurosystem monetary policy operations." Working Paper 118, National Bank of Belgium.

Coppens, François, Manuel Mayer, Laurent Millischer, Florian Resch, Stephan Sauer, and Klaas Schulze (2016). "Advances in multivariate back-testing for credit risk underestimation." Working Paper 1885, European Central Bank.

ECB (2015). "Questions and answers on the compilation of the static pool for performance monitoring in the Eurosystem Credit Assessment Framework (ECAF)." Available online.

Figlewski, Stephen, Halina Frydman, and Weijian Liang (2012). "Modeling the effect of macroeconomic factors on corporate default and credit rating transitions." *International Review of Economics and Finance*, 21(1), 87–105.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Lingo, Manuel and Gerhard Winkler (2008). "Discriminatory power: an obsolete validation criterion?" *Journal of Risk Model Validation*, 2(1), 45–72.

Martinho, Ricardo and António Antunes (2012). "A Scoring Model For Portuguese Non-Financial Enterprises." In *Financial Stability Report*, pp. 111–124. Banco de Portugal.

Plosser, Matthew C. and João Santos (2014). "Banks' incentives and the quality of internal risk models." Staff Report 704, Federal Reserve Bank of New York.

Santos, João and Andrew Winton (2015). "Bank capital, borrower power, and loan rates." Mimeo.

Spiegelhalter, David J. (1986). "Probabilistic prediction in patient management and clinical trials." *Statistics in Medicine*, 5, 421–433.

Vazza, Diane and Nick W. Kraemer (2015). "2014 Annual Global Corporate Default Study And Rating Transitions." Ratings direct, Standard & Poor's

Rating Services.

Wu, Xuezhen (2008). *Credit Scoring Model Validation*. Master's thesis, University of Amsterdam, Korteweg-de Vries Institute for Mathematics.

CQS	Withdrawn	Default rate			Share of
		Observed	Estimated	*Upper limit*	total sample
1 & 2	16.4	0.17	0.10	*0.10*	0.16
3	11.1	0.31	0.28	*0.40*	5.5
4	11.6	0.69	0.69	*1.00*	16.7
5	11.8	1.27	1.24	*1.50*	11.1
6	12.4	2.20	2.17	*3.00*	21.8
7	13.1	4.02	3.91	*5.00*	16.0
8	17.6	10.3	10.00	*100*	28.8
Full sample	13.8	4.36	4.25	*n.a.*	100

TABLE 6. Observed and model-estimated default rates and rate of exits from the sample without default, by rating class. Model default rates estimated using the semiparametric methodology. All values in percentage. Model-estimated default rate for CQS 1 & 2 set to the upper limit of the class.

Source: Banco de Portugal and authors' calculations.

CQS	Default rate	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
1 & 2	Estimated	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	Observed	0.00	1.75	0.00	0.00	0.72	0.00	0.00	0.00	0.00	0.00	0.17
3	Estimated	0.29	0.28	0.29	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
	Observed	0.16	0.30	0.33	0.78	0.17	0.27	0.42	0.22	0.26	0.39	0.31
4	Estimated	0.70	0.70	0.70	0.70	0.69	0.69	0.69	0.69	0.69	0.69	0.69
	Observed	0.48	0.51	0.64	0.87	0.42	0.77	1.13	0.77	0.70	0.46	0.69
5	Estimated	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24
	Observed	0.82	1.00	1.46	1.82	1.05	1.59	1.89	1.34	1.02	0.66	1.27
6	Estimated	2.17	2.17	2.18	2.18	2.18	2.17	2.17	2.17	2.16	2.16	2.17
	Observed	1.35	1.84	2.41	3.33	1.70	2.54	3.40	2.21	1.68	1.42	2.20
7	Estimated	3.90	3.90	3.91	3.91	3.91	3.91	3.90	3.92	3.90	3.89	3.91
	Observed	2.61	3.56	4.64	6.09	2.99	4.51	5.86	3.99	3.30	2.35	4.02
8	Estimated	9.06	9.20	9.32	9.52	10.30	10.25	10.62	10.95	10.20	9.78	10.04
	Observed	6.57	7.99	10.43	14.44	8.09	11.00	15.29	11.32	8.59	6.42	10.31
Total	Estimated	3.77	3.91	4.03	4.30	4.64	4.25	4.59	4.82	4.13	3.75	4.25
	Observed	2.63	3.40	4.54	6.53	3.62	4.68	6.74	4.98	3.47	2.43	4.36

TABLE 7. Observed and model-estimated default rates over time, by rating class. Model default rates estimated using the semiparametric methodology. All values in percentage. Model-estimated default rate for CQS 1 & 2 set to the upper limit of the class.

Source: Banco de Portugal and authors' calculations.

CQS in year t	CQS in year t+1						
	1 & 2	3	4	5	6	7	8
1 & 2	36.5	**55.9**	5.9	0.7	0.8		0.1
3	1.5	**56.5**	32.0	4.5	3.6	1.1	0.8
4	0.0	10.7	**51.3**	17.3	13.7	4.1	2.8
5	0.0	2.0	25.8	26.1	**30.6**	9.3	6.2
6	0.0	0.8	9.4	14.4	**40.2**	20.5	14.7
7		0.3	3.5	5.3	24.6	31.8	**34.4**
8		0.1	1.4	2.2	9.1	16.0	**71.2**

TABLE 8. Transition matrix between credit rating classes, conditional on firms being in the sample in two consecutive years and not defaulting. Rows add up to 100 percent. All values in percentage.

Source: Banco de Portugal and authors' calculations.

CQS	Statistic	Industry					Size		Total
		1	2	3	4	5	1	2	
1 & 2	Estimated def. rate	0.10		0.10	0.10	0.10		0.10	0.10
	Observed def. rate	0.00		0.18	0.00	0.00		0.17	0.17
	Share of obs.	*0.02*		*0.40*	*0.70*	*0.00*		*0.36*	*0.16*
3	Estimated def. rate	0.29	0.34	0.27	0.26	0.31	0.33	0.27	0.28
	Observed def. rate	0.40	1.38	0.30	0.00	0.19	0.29	0.32	0.31
	Share of obs.	*5.89*	*0.45*	*8.61*	*12.56*	*3.56*	*1.33*	*10.79*	*5.52*
4	Estimated def. rate	0.69	0.74	0.68	0.68	0.70	0.72	0.67	0.69
	Observed def. rate	0.68	0.94	0.75	0.79	0.56	0.70	0.68	0.69
	Share of obs.	*17.45*	*6.48*	*19.12*	*19.33*	*18.41*	*13.10*	*21.20*	*16.69*
5	Estimated def. rate	1.24	1.25	1.24	1.24	1.24	1.24	1.24	1.24
	Observed def. rate	1.44	1.45	1.25	0.56	1.14	1.24	1.31	1.27
	Share of obs.	*10.81*	*7.72*	*11.44*	*10.75*	*12.72*	*11.24*	*10.88*	*11.08*
6	Estimated def. rate	2.17	2.22	2.16	2.16	2.16	2.18	2.16	2.17
	Observed def. rate	2.24	2.25	2.10	2.22	2.26	2.21	2.17	2.20
	Share of obs.	*21.02*	*21.73*	*21.19*	*18.45*	*23.39*	*24.15*	*18.83*	*21.79*
7	Estimated def. rate	3.91	3.94	3.89	3.91	3.89	3.91	3.90	3.91
	Observed def. rate	3.89	3.76	3.98	5.28	4.32	4.11	3.86	4.02
	Share of obs.	*15.52*	*20.40*	*14.67*	*12.65*	*15.86*	*18.54*	*12.82*	*16.00*
8	Estimated def. rate	10.15	10.47	10.12	9.83	9.45	9.54	10.83	10.00
	Observed def. rate	10.37	10.80	10.47	9.75	9.59	9.71	11.26	10.31
	Share of obs.	*29.29*	*43.22*	*24.56*	*25.55*	*26.06*	*31.64*	*25.12*	*28.75*
Total	Estimated def. rate	4.30	5.96	3.81	3.70	3.88	4.51	3.93	4.24
	Observed def. rate	4.41	6.10	3.91	3.78	3.97	4.60	4.05	4.36
	Share of obs.	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*

TABLE 9. Model-estimated and observed default rate for selected groups of firms. Model default rates estimated using the semiparametric methodology. All values in percentage. Model-estimated default rate for CQS 1 & 2 set to the upper limit of the class.

Source: Banco de Portugal and authors' calculations.

The sources of the gender wage gap

Ana Rute Cardoso
Institute for Economic Analysis, Barcelona GSE

Paulo Guimarães
Banco de Portugal

Pedro Portugal
Banco de Portugal, Universidade NOVA de Lisboa and IZA Bonn

Pedro S. Raposo
UCP - Catolica Lisbon School of Business and Economics

April 2016

Abstract
In Portugal, over the last two decades, the proportion of women among employed workers increased from 35 to 45 percent. This evolution was accompanied by a sharp fall in the gender wage gap from 32 to 20 percent. The improvement in the wage outcome of the women, however, is fully accounted by the catching up of their skills in comparison to males, after two decades of human capital investments. By 2013 women already possess observable characteristics that enhance productivity identical to their male counterparts. This means that gender discrimination remained roughly constant over the 1991-2013 period. In this study, we investigate the sources of the wage gender gap and conclude that sorting among firms and job-titles can explain about two fifths of the wage gender gap. (JEL: J16, J24, J31, J71)

"Um dos aspectos da desigualdade é a singularidade - isto é, não o ser este homem mais, neste ou naquele característico, que outros homens, mas o ser tão-somente diferente dele."

"Os espíritos altamente analíticos vêem quase só defeitos: quanto mais forte a lente mais imperfeita se mostra a cousa observada."

Fernando Pessoa

Introduction

In 1991, the wages of Portuguese women used to be around two thirds of the wages of men. Since 1991, women dominated the labor market inflows, particularly the better skilled women. This evolution translated into a 10

Acknowledgements: The authors are grateful for all the help provided by António Antunes, Hugo Reis, Lucena Vieira and Hugo Vilares and benefit from insightful discussions with the students from the Labor Economics class at NOVA SBE.

E-mail: anarute.cardoso@iae.csic.es; pfguimaraes@bportugal.pt; pportugal@bportugal.pt; pedro.raposo@ucp.pt

percentage point increase in the feminization rate of the stock of employed workers (see figure 1) and a 12 percentage point decrease in the raw wage gender gap (blue line, in figure 2). By 2013, the average wages of women represented about four fifths of the wages of men.

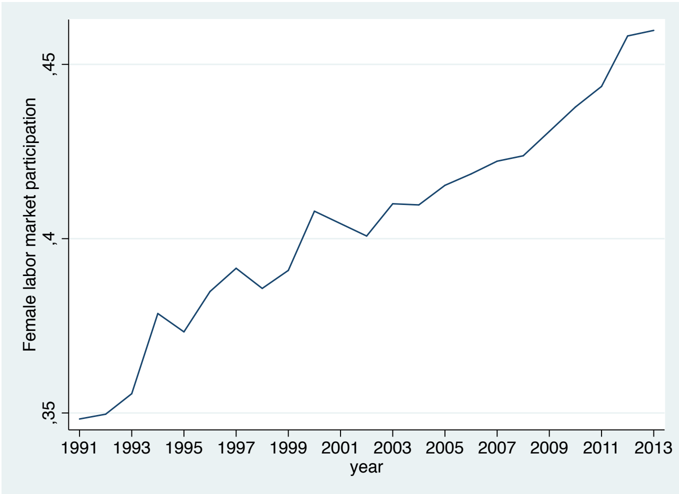

FIGURE 1: Female labor market participation

If, however, we take into account the characteristics of the workers to compute an "adjusted" gender wage gap, there is no longer an indication of improvement (red line, figure 2). In other words, the wage gain achieved by women over this period is due to the catching up of their skills (labor market experience, seniority, etc.) in comparison to males not by a reduction in unexplained component of the wage difference, which is conventionally equated to gender discrimination. Gender discrimination, in this sense, did not ameliorate, it slightly deteriorate over the 22 years period.

In this study we aim to study what hides behind the gender wage gap by executing a number of wage decomposition exercises. Firstly, we shall exploit the Machado and Mata (2005) quantile decomposition methodology to disentangle the role of the structural from the composition effects along the quantiles of the wage distribution. Secondly, we will combine the estimation of high-dimensional fixed effects regression models with the omitted variable bias decomposition suggested by Gelbach (2016) to access the importance of sorting into firms with heterogeneous wage policies and into job titles which are associated with different wage differentials. In this sense, we are updating and extending the work of Cardoso, Guimarães, and Portugal (2016).Thirdly, and finally, we will adapt the methodology of Guimarães and Portugal (2010) to incorporate the notion of high-dimensional slope effects, measuring gender wage gaps at the firm and the job-title levels.

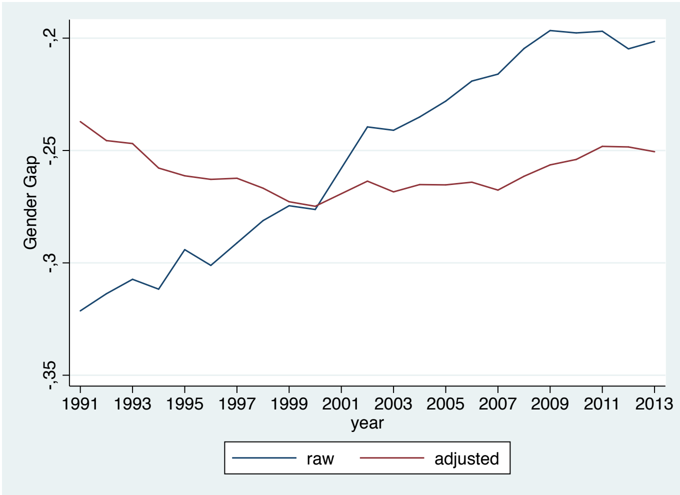

FIGURE 2: Gender wage discrimination

For this purpose, we rely upon an unusually rich data set, the "Quadros de Pessoal" survey, a longitudinal matched employer-employee- job title dataset, which covers all the establishment with at least one wage earner. The information about wages is provided by the employer annals. It is a complete and reliable source, because the main reason for its existence is to allow the officials from the Ministry of Employment to verify whether the employers are complying with the wage floors established by the collective agreement for the job-title of the worker.

The next section makes a brief literature review. Section 3 describes the data, while methods are discussed in Section 4. Section 5 provides the key results on the determinants of the gender pay gap. Section 6 concludes.

Literature review

It seems fair to claim that we are witnessing a revival of interest in the search for the determinants of the gender pay gap, under new empirical approaches, richer data, and renewed theoretical perspectives. Indeed, traditional economic analysis had focused primarily on the importance of female labor force participation and differences in observable attributes between men and women on the gender pay gap. Either of these two mechanisms can be understood intuitively. If the female participation rate is low, there is scope for the attributes of employed women to be unrepresentative of those of the female population in general. This selection could operate to raise or lower females' wages relative to males, depending on whether social norms, preferences, economic conditions and public policies disproportionately attract into the labor market more or less qualified women (along dimensions that can be observable or unobservable). In any case, as the female participation rate increases, the importance of selection influencing the gender pay gap is expected to decline (see the cross-country evidence in Olivetti and Petrongolo (2008) or the evidence over time for the US in Stanley and Jarrell (1998) and Jarrell and Stanley (2004)). Concomitantly, the qualifications of females and males in the labor market will influence their relative pay (see the ample evidence that education and experience contribute to shape the gender pay gap, in the review by Altonji and Blank (1999). Under this strand of literature, the convergence in schooling achievement across males and females (if not the reversal of the gap, in favor of women) and the increased female labor force attachment would lead us to expect the closing of the gender pay gap. Strikingly, a question lingers on: Why is the gender pay gap so persistent, despite a marked convergence in the participation rates and observable labor market attributes of men and women, in particular in developed economies?

A recent surge of literature addresses that question. Blau and Kahn (2016) report on the partial closing of the gender pay gap in the US in recent decades, remarkably so during the 1980s. Their empirical analysis, together with a review of the recent literature for other countries, points to a set of stylized facts and remaining challenges.

First of all, the convergence in attributes such as education and experience played a key role reducing the gender pay gap. These factors have currently a muted impact on pay differences between men and women. On the contrary, the industry and the occupation strive as factors generating pay differences across gender. Further research is thus needed to fully understand the allocation of gender across industries and occupations and their associated pay. Most likely, a better understanding of firm recruitment and pay policies will be helpful. A third noteworthy fact is that the gender pay gap is persistently larger at the top of the skill and wage distribution. The sources of this "glass ceiling effect" are also not yet fully understood. Plausible explanations highlighted by Blau and Khan include differences in psychological attributes (for example, competitiveness and bargaining power) that would penalize women at the top of the skill and job ladder, compensating differentials for characteristics of the top jobs (for example, longer and less flexible working hours), and pure discrimination.

Progress on some of the pending issues has recently been facilitated by availability of large longitudinal linked employer-employee datasets. Cardoso *et al.* (2016) (CGP) quantify the impact of access to firms and detailed jobs on the gender pay gap. They depart from the idea that different firms adopt different pay standards and assume that this generosity of the firm pay policy is common across gender and can be captured by a firm-specific fixed effect in a wage regression. Their subsequent step is to compare the average firm wage effect for males and females. They conclude that gender allocation to firms of different pay standards accounts for 20% of the overall gender pay gap. Similarly, the sorting or segregation of males and females across job titles accounts for approximately another 20% of the gender pay gap. Their quantification of the impact of worker allocation to firms and to jobs takes into due account the heterogeneity in worker quality. Their exercise is accomplished by adapting the methodology in Gelbach (2016), which allows for an unambiguous decomposition of the gender pay gap.

Card *et al.* (2016) (CCK) progressed along a different dimension. They aimed at formally testing a hypothesis long discussed in other fields of science, which made its entry into economic analysis more recently, namely, that females would have non-observable skills (such as competitiveness and bargaining attitudes) that would penalize them in the labor market vis a vis men. If so, women would extract lower rents from their employer than men working for the same firm. Accordingly, CCK allow for gender-specific firm wage premia and link these premia to measures of firm performance. Their analysis thus uncovers two channels contributing to the gender pay gap: the allocation of workers to firms (sorting or segregation channel) and the bargaining channel. Their decomposition of the pay gap is performed by relying on the following counterfactual exercises: by imposing the male firm wage premium on females in the same firm, they "shut down" the bargaining channel; similarly, by imposing an even distribution of males and females across firms, they "shut down" the allocation channel. The exercise requires firms that employ both males and females and it thus excludes single-gender firms.[1] They conclude, on one hand, that the bargaining effect accounts for 5% of the overall gender pay gap in Portugal. On the other hand, they confirms the relevance of the firm sorting channel, as it accounts for 15% of the overall pay gap.

Another recent strand of literature explores the role of compensating differentials for characteristics of the top jobs, in particular longer and less flexible working hours. Goldin (2014) and Bertrand and Katz (2010) are among the studies that present compelling evidence on the importance of this channel.

The aim of the current paper is to progress along the new strand of literature that relies on large longitudinal linked employer-employee data to evaluate the role of the firm shaping the gender pay gap.

1. A further requirement is that these firms are "connected" by workers of either gender moving across firms.

Data

The *Quadros de Pessoal* (QP) is, by construction, a longitudinal matched employer-employee-job title data set. QP is an annual mandatory employment survey collected by the Portuguese Ministry of Employment, and covers virtually all firms employing paid labor in Portugal. Due to the mandatory nature of the survey, problems commonly associated with panel data sets, such as panel attrition, are considerably attenuated.

The data set includes both firm-specific information (location, industry (SIC codes), legal setting, foreign ownership, employment, sales, ownership type) and and each and every one of its workers (labor earnings, worker qualifications, gender, age, tenure, hours of work, etc.). The information on earnings is very detailed, precise, and complete. It includes the base wage (gross pay for normal hours of work), regular benefits, and overtime pay. Information on standard and overtime hours of work is also available. Because the information on earnings is reported by the employer, it is likely to be subject to less measurement error than worker-provided earnings data. The fact that the information contained in the QP survey needs, by law, to be available in a public space at the establishment further reinforces our trust in the information.

A notable feature of the QP is that it collects information regarding the collective agreement that rules the wage dimension of the match between the employer and the employee. Furthermore, within each collective agreement, it identifies the particular job-title that the worker holds. The relevance of progressing from the broad classification of occupations traditionally available in datasets into a richer description of the actual tasks performed by workers has been highlighted in the literature [see for example Autor (2013), or Goos and Manning (2007), Autor *et al.* (2006) and Dustmann *et al.* (2009) on job polarization]. This recent literature illustrates that, in addition to firm and worker heterogeneity, wage outcomes are shaped by task heterogeneity, which should be explicitly accounted for in the analysis (Torres *et al.* 2013).

A number of restrictions were imposed on the raw data set. First, we limited our analysis to full-time workers in mainland Portugal, between 1986 and 2013.[2] Second, we excluded workers from the Agriculture and Fishery sectors. Third, individuals younger than 18 years old and older than 65 years were also excised. Fourth, we dropped from the analysis workers whose monthly wages were below 80 percent of the mandatory minimum wage, which corresponds to the lowest admissible wage for apprentices. Fifth, we excluded observations whose firm-job-title match included only one worker. Finally, we dropped (around 1 percent of the total number of) observations

2. The years between 1986 and 1989 were only used in order to obtain with more precision the estimates of the three high dimensional fixed effects in equation 3.

that did not belong to the largest connected group. Our final sample included 27,921,002 observations (338,580 firms; 5,126,998 workers; 95,196 job titles).

The dependent variable used in our estimating equation is a measure of real hourly labour earnings and is constructed as the ratio of the sum of deflated base wages, regular benefits (including seniority payments), and overtime pay over the sum of normal hours of work and overtime hours.

High-Dimensional fixed effects and Gelbach's decomposition

In this section we follow closely the empirical approach of Cardoso *et al.* (2016). The idea consists of employing Gelbach's (2016) decomposition to help sort out the root causes of the observed gender wage gap. The novelty here is the application of Gelbach's decomposition to a linear regression model that accounts for the main sources of variation including unobserved components that are captured by the inclusion of multiple high-dimensional fixed effects. Our departure point is the traditional workhorse Mincerian wage equation:

$$\ln w_{ifjt} = \mathbf{x}_{ift}\boldsymbol{\beta} + \gamma g_i + \varepsilon_{ifjt} \quad . \tag{1}$$

In the above equation, $\ln w_{ifjt}$ stands for the natural logarithm of the real hourly wage. The various indices attached to w serve to emphasize all potential sources of wage variation. The index i $(i = 1, ..., N)$ stands for the worker, f $(f = 1, ..., F)$ accounts for firms while j reflects the variation accrued by differences in job titles. The index t stands for time $(t = 1, ..., T)$. The vector of explanatory variables, $\mathbf{x}$, comprises both observed characteristics of the worker and of the firm. These include variables such as worker education and tenure as well as firm size. Intentionally, we leave out of the vector $\mathbf{x}$ the variable g, a dummy that accounts for gender differences. The coefficient associated with this variable is the focus of our analysis as it provides the standard estimate for the gender wage gap. Finally, it is assumed that the error term, ε_{ifjt} follows the conventional assumptions.

It is more convenient to express the above equation in matrix terms. In doing so we obtain

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{G} + \boldsymbol{\varepsilon} \tag{2}$$

where the symbology used is quite obvious. The above specification is what we call the base model and is the regression typically used the ascertain the size of the gender wage gap. Basically, it estimates the percentage difference between the wages of men and women once we take into account the observed characteristics of the workers such as their education level and tenure and important firm characteristics such as size. However, in line with the work of Abowd *et al.* (1999), we recognize the need to explicitly account for all wage variation emanating from factors that are specific to the worker and the firm. This can only be accomplished with employer-employee data. As shown by Abowd *et al.* (1999) with the introduction of fixed-effects for firm and worker we are able to control for time-invariant characteristics of workers and firms whether or not we are able to observe them. In this framework, things such as worker ability, family background, risk aversion, etc. are all accounted for. The same applies to firm unobserved characteristics, such as managerial ability and organization, location, etc. The richness of our data allows us to go a step further. As explained earlier, since we have detailed job title information we are also able to introduce a fixed effect that absorbs all time-invariant characteristics of a specific job-title.

Adding firm or job-title fixed effects to the base equation in 2 should not affect the estimate of γ unless there is an uneven distribution of gender across firms and job-titles. Put differently, if γ changes when we fully control for firm and job-title effects than this means that the sorting of females/males across firms or jobs is a factor that is contributing to the gender wage gap. But, to account for the main sources of variation the full regression model needs to also include a worker fixed effect. With the introduction of an individual specific fixed effect we will absorb all time-invariant individual specific characteristics, including the gender dummy variable (G). As we will see below, this does not prevent us from understanding what happens to γ when we control for all three additional sources of variation (worker, firm and job title). In order to do this we need to estimate a full model, one that includes the three fixed effects. This model is simply

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\varphi} + \mathbf{L}\boldsymbol{\lambda} + \boldsymbol{\varepsilon} \tag{3}$$

where we have added three high-dimensional fixed effects to the equation in (2). $\mathbf{D}$ is a design matrix for the worker effects, $\mathbf{F}$ is design matrix for the firm effects while $\mathbf{L}$ is a design matrix for the job title effects. As usual, we maintain the assumption of strict exogeneity of the error term.

The large size of our data, with around 28 million observations, more than 5 million workers and 400 thousand firms, and around 95,000 distinct job-titles, raises some econometric challenges. Of particular concern is the high-dimensionality of the fixed effects. Estimation of a regression with three high-dimensional fixed effects is a non-trivial issue given the size of the matrices involved. The within transformation can absorb one of the fixed effects but the large dimension of the remaining fixed effects prevents the application of the conventional OLS formula. Estimation of this model is possible if we resort to the algorithm of Guimarães and Portugal (2010). This algorithm is able to provide the exact OLS solution without requiring the inversion of large matrices. [3]

3. We used the user-written command `reghdfe` coded by Sergio Correia which implements an improved version of the Guimarães and Portugal (2010) algorithm.

Since we provide secondary analysis of the estimates of the fixed effects we have to make sure that they are identifiable. This is done by restricting our data set to a connected subset. We accomplish this by using an algorithm proposed by Weeks and Williams (1964). Application of this algorithm to our data permits the identification of a subset of the data that comprises 99% of our original data set. Within this subset of data the estimates of all fixed effects are comparable up to an additive scalar factor.

The Gelbach (2016) decomposition can help us understand what happens to the estimate of γ when we move from the basic equation in (2) to the full equation (3) where the three fixed effects are simultaneously added. The approach is based on the OLS formula for omitted variable bias and has the advantage of providing an unequivocal way to quantify the parcel of change that can be attributed to the inclusion of each individual fixed effect. To see how the decomposition can be employed in this context we recall that by the Frisch-Waugh-Lovell (FWL) theorem it is possible to obtain an estimate of the γ in the base model by running a two step regression. First, we regress $\mathbf{Y}$ on $\mathbf{X}$ and calculate the residual of that regression. If we let $\mathbf{M} \equiv [\mathbf{I} - \mathbf{X}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}]$ be the residual-maker matrix then this amounts to calculating the vector $\mathbf{MY}$. Similarly, we calculate the residual of the regression of $\mathbf{G}$ on $\mathbf{X}$, that is, $\mathbf{MG}$. With this procedure we have expurgated the effect of the $\mathbf{X}$ variables from $\mathbf{Y}$ and $\mathbf{G}$. Thus, if we now run a simple linear regression of $\mathbf{MY}$ on $\mathbf{MG}$ we know by the FWL theorem that we obtain the OLS estimate for the γ in our base model. That is,

$$\widehat{\gamma} = (\mathbf{G}^{'}\mathbf{MG})^{-1}\mathbf{G}^{'}\mathbf{MY} = \mathbf{M_G}\mathbf{Y} \tag{4}$$

where we note in passing that $\mathbf{M_G} \equiv (\mathbf{G}^{'}\mathbf{MG})^{-1}\mathbf{G}^{'}\mathbf{M}$ and $\mathbf{M}$ is an idempotent matrix. We now turn to the full version of the wage equation model in (3). The fitted version of this model can be expressed as

$$\mathbf{Y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{D}\widehat{\boldsymbol{\theta}} + \mathbf{F}\widehat{\boldsymbol{\varphi}} + \mathbf{L}\widehat{\boldsymbol{\lambda}} + \widehat{\boldsymbol{\varepsilon}} \tag{5}$$

where we have replaced the coefficients and error term by their OLS estimates. Note that $\mathbf{D}\widehat{\boldsymbol{\theta}}$, $\mathbf{F}\widehat{\boldsymbol{\varphi}}$ and $\mathbf{L}\widehat{\boldsymbol{\lambda}}$ are the column vectors containing the estimates of the fixed effects. To implement Gelbach's decomposition we simply have to pre-multiply the above expression by $\mathbf{M_G}$. When we do this we obtain on the left-hand side the formula for the OLS estimate of γ while on the right-hand side the terms associated with $\mathbf{X}$ and $\widehat{\boldsymbol{\varepsilon}}$ disappear. [4] We are left with three components, each one associated with one of the fixed effects, that add up to the observed gender wage gap, $\widehat{\gamma}$. That is,

$$\widehat{\gamma} = \widehat{\delta}_{\boldsymbol{\theta}} + \widehat{\delta}_{\boldsymbol{\varphi}} + \widehat{\delta}_{\boldsymbol{\lambda}} \quad . \tag{6}$$

4. By construction $\widehat{\boldsymbol{\varepsilon}}$ is orthogonal to $\mathbf{X}$ and to $\mathbf{D}$ meaning that it is also orthogonal to $\mathbf{G}$. It follows that $\mathbf{M_G}\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}$. Using the fact that $\mathbf{MX} = \mathbf{0}$ it is easy to show that $\mathbf{M_G}\mathbf{X} = \mathbf{0}$.

In practical terms each $\widehat{\delta}$ in the left-hand side is the coefficient of a regression between the respective fixed effect and the gender variable adjusting for the $\mathbf{X}$ covariates. If, conditional on the $\mathbf{X}$ variables, the distribution of females across firms was absolutely random then we would expect $\widehat{\delta}_{\varphi}$ to be zero. This would mean that the sorting of females/males across firms was not a contributing factor to the gender pay gap. A similar reasoning can be applied to the sorting of gender across jobs.

Discussion of the results

The Machado and Mata decomposition

We rely on quantile regression methods to analyse the changes in the wage distribution between gender over a 22 year period. To that end, we use the Machado and Mata decomposition method which enables us to identify the sources of the changes in the distribution of wages between females and males. We repeat the exercise in 1991 and in 2013 in order to compare how the sources of variation have evolved between the beginning of the period (1991) and 22 years later (2013).

Gender differences in the distribution of wages may result from changes in the distribution of the conditioning variables (changes in terms of the characteristics of the population, e.g. labor force characteristics such as education and age) or from changes in the conditional distribution of wages itself (which may be thought of as changes in the way characteristics impact wages, the "coefficients"). The first is a "composition effect" and the second may be thought of as a "structural effect" (Autor *et al.* (2008)). We build the counterfactual exercise by estimating the marginal distribution of wages that would have prevailed for male if they had the characteristics of females ("composition effect"). Subsequently, we estimate the marginal distribution of wages that would have prevailed for female if they had the same returns than males ("structural effect").

In 1991, men earned more than women, most notably at higher percentiles. Whereas males earned more 35.1 log points than females at the median, the difference was 41.7 log points at the 8th decile (see the third column of Table 1). It is clear from columns 4th and 5th that (aggregate) differences in the coefficients were more influential driving the overall shift in the wage distribution than (aggregate) differences in the covariates. At the median, the gender wage gap was 10.9 log points due to changes in covariates and it was 24.2 log points due to changes in the coefficients. Interestingly, "covariate changes" are larger at the 1st decile but "coefficient differences" become more influential as we move up the wage distribution. The "coefficient changes" generated a larger gender gap at the highest percentiles.

	(1) Female x0b0	(2) Male x1b1	(3) (2)-(1)	(4) Aggregate composition effect (x1b1-x0b1)	(5) Aggregate structural effect (x0b1-x0b0)
10 percentile	-0.433***	-0.268***	0.165***	0.090***	0.074***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
20 percentile	-0.351***	-0.116***	0.235***	0.096***	0.139***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
30 percentile	-0.274***	0.010***	0.284***	0.101***	0.183***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
40 percentile	-0.191***	0.130***	0.322***	0.105***	0.216***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
50 percentile	-0.099***	0.251***	0.351***	0.109***	0.242***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
60 percentile	0.008***	0.384***	0.375***	0.111***	0.264***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
70 percentile	0.142***	0.539***	0.397***	0.114***	0.282***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
80 percentile	0.321***	0.737***	0.417***	0.117***	0.300***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
90 percentile	0.602***	1.036***	0.433***	0.123***	0.310***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)

TABLE 1. Gender wage discrimination decomposition (1991)

	(1) Female x0b0	(2) Male x1b1	(3) (2)-(1)	(4) Aggregate composition effect (x1b1-x0b1)	(5) Aggregate structural effect (x0b1-x0b0)
10 percentile	-0.163***	-0.064***	0.099***	0.021***	0.078***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
20 percentile	-0.063***	0.067***	0.130***	0.014***	0.117***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
30 percentile	0.028***	0.186***	0.158***	0.008***	0.150***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
40 percentile	0.124***	0.306***	0.183***	0.002**	0.181***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
50 percentile	0.229***	0.434***	0.205***	-0.004***	0.209***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
60 percentile	0.349***	0.575***	0.226***	-0.009***	0.235***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
70 percentile	0.495***	0.739***	0.244***	-0.015***	0.259***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
80 percentile	0.684***	0.947***	0.262***	-0.016***	0.279***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90 percentile	0.968***	1.256***	0.288***	-0.013***	0.301***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)

TABLE 2. Gender wage discrimination decomposition (2013)

	1991		2013	
	Female	Male	Female	Male
Age	33,98	38,27	40,25	40,73
Tenure	8,87	10,17	9,47	9,55
Firm size	5,09	5,50	4,82	4,77
Education	6,36	6,27	9,86	9,29

TABLE 3. Gender wage discrimination: Summary statistics (Composition)

	1991		2013	
	Female	Male	Female	Male
Age	0,0190	0,0470	0,0267	0,0452
Age^2	-0,0001	-0,0005	-0,0002	-0,0004
Tenure	0,0095	0,0104	0,0171	0,0216
$Tenure^2$	-0,0002	-0,0001	-0,0002	-0,0002
Firm size	0,0422	0,0629	0,0258	0,0426
Education	0,0739	0,0810	0,0723	0,0783
Fraction of Females	-0,2527	-0,1031	-0,2537	-0,1023
Constant	-1,0997	-1,6832	-1,2180	-1,7495

TABLE 4. Gender wage discrimination: Quantile regressions (β)

In 2013, the gender gap is still positive and statistically significant but its magnitude was reduced. Although males earn more 20.5 log points than females at the median, the difference between the highest and the lowest percentiles was reduced. In 2013 the "coefficient differences" are everywhere larger, in absolute magnitude, than "covariate differences" (Table 2).

Females in 2013 are not only more similar to their male counterparts but also show better characteristics (Table 3). Women in 2013 are older and more experienced reflecting the increase in their labor market participation rates. The educational level of the labor force increased considerably during this period reflecting the aging of the baby-boom generation. Women in 2013 are working in larger firms and they are clearly more educated than male. There are significant differences in the returns to education, both in 1991 and in 2013. Despite having similar characteristics the return to general and specific human capital is much lower for women in comparison with their male counterparts (Table 4). The high paying policies by large firms benefit males in a much larger extent than females. Finally, firms whose workforce is more heavily populated by women (more segregated) generate a wage penalty, most notably, for females.

The Gelbach decomposition

The sizable gender wage gap for total hourly earnings that we have estimated constitutes an average differential between the wages of two otherwise observably identical workers. A key question concerns the potential sources of the unobserved heterogeneity behind these differentials (see figure 3). We next consider how sorting among firms with different compensation policies, the assignment to distinct job titles, and the allocation of workers with different unobserved ability drive the gender wage gap. Our focus in decomposing the gender wage gap is therefore upon the contributions of each of these three sources of unobserved heterogeneity.

Before proceeding, it is worth to discuss the interpretation of the three high-dimensional fixed effects added in equation (3). The firm fixed effect, in essence, captures the (constant) wage policy of the firm. Firms with generous compensation policies will exhibit positive firm fixed effects, low-wage firms will generate negative fixed effects. In Figure 4 we contrast the distribution of the firm fixed effects for workers by gender.[5] It is very clear from the picture that males disproportionally populate high paying firms.

In Figure 5 the empirical distribution of the worker fixed-effects are presented. The worker fixed effects condense the influence of constant characteristics (observed and non-observed) of the individuals on their wages. They can be a proxy for the portable human capital (or productivity) of the worker or they may simply reflect gender discrimination that is not associated with sorting of workers across firms and job titles. The picture shows the wage gap between males and females is firmly rooted in the individual component of wages, more notably in the upper tail of the distribution. This outcome can be the result of observed or unobserved characteristics (say, schooling or ability). We shall, below, identify the specific role of unobserved skills.

Finally, we show the empirical distribution of the job title fixed effects. Job title fixed effects largely reflect the remuneration status of disaggregated occupations. In a way, the inclusion of job title effects builds upon first generation Mincerian wage equation which included broad definition of occupations. In the current setup, we provide an unusually fine accounting of the tasks required to fill a job. The distributions of the job title fixed effects given in Figure 6 do exhibit a discernible difference in terms gender, suggesting that the allocation of workers across job titles significantly disfavors women.

Results for the Gelbach decomposition are given in table 5. It can be seen that the wage penalty of 25.6 log points (arriving from the estimation of equation 1) can be decomposed into the contribution of three parts: worker, firm, and job title unobserved heterogeneity. A significant fraction of the gender wage gap is explained by the heterogeneity of the firms′ compensation policies. The allocation of workers into firms is responsible for 5.8 out of 25.6 log points of the gender wage gap. This means that females disproportionally belong to firms firms with less generous wage policies. Put differently, if workers were randomly assigned to firms, the gender wage gap would be reduced by about one fifth. We also find that the attribution of job-titles, either through promotion policies or through initial assignments, is significantly influenced by gender, contributing 4.3 log points to wage gap. Together, the process of sorting into firms and job titles accounts for around 40 percent of the gender wage gap. The unobserved (permanent) characteristics of the individuals is responsible for the remaining 60 percent. These unobserved (to

5. Notice, however, that in this comparison the influence of variables such as industry or firm size are still subsumed in the firm fixed effect.

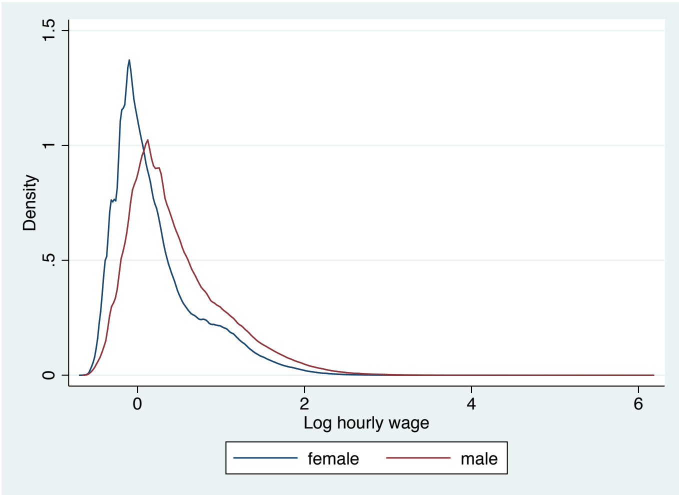

FIGURE 3: Gender wage discrimination: (log) real hourly wage

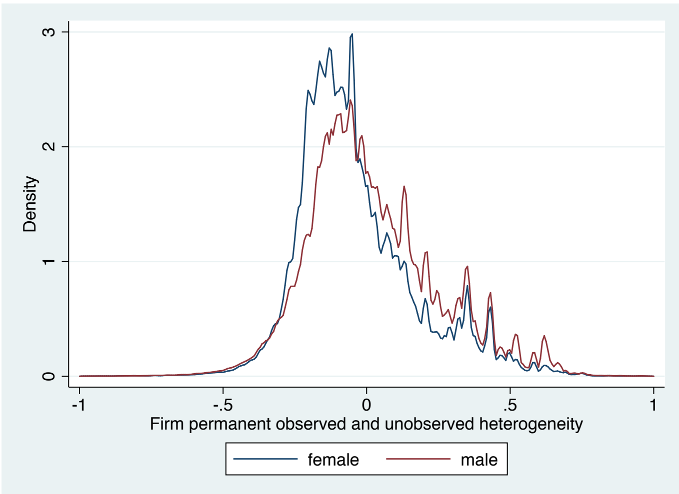

FIGURE 4: Gender wage discrimination: Firm

the researcher) worker characteristics can be equated either with unobserved skills or, simply, to some form of gender discrimination.

Figure 7 display the gender gap decomposition over time. The allocation of female workers into firms and job-titles did not improve over the last two decades. If anything, the sorting into firms and job-titles is now slightly less favorable for women (-1.7 and -1.0 log points for firms and for job-titles, respectively, over the 1991-2013 period). In compensation, the wage penalty resulting from the role of unobserved individual heterogeneity was visibly attenuated (3.2 log points), in particular since the beginning of the century.

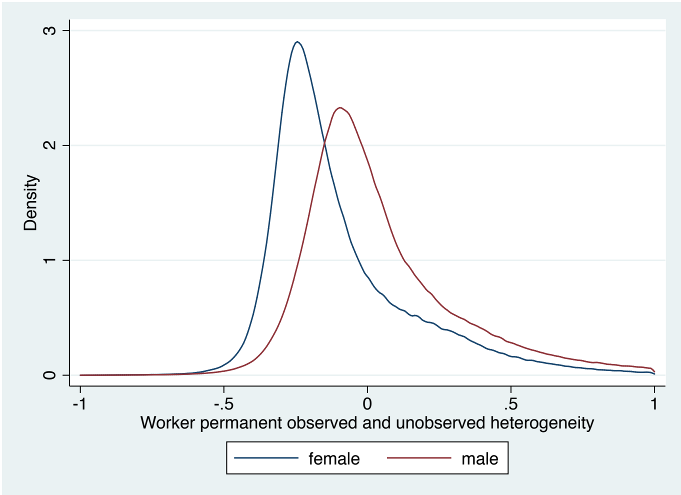

FIGURE 5: Gender wage discrimination: Worker

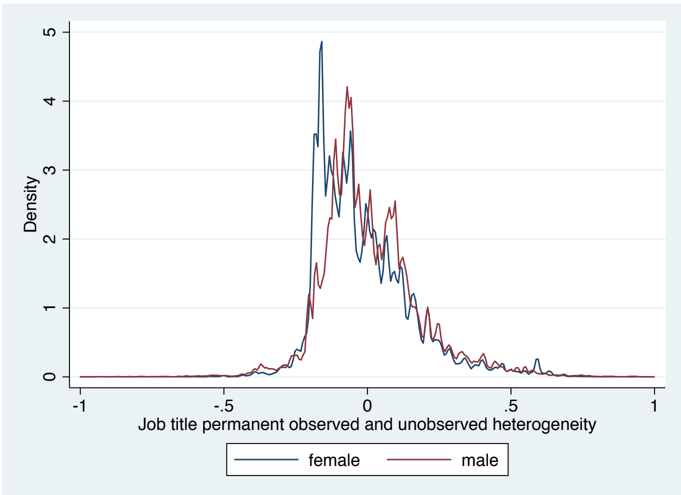

FIGURE 6: Gender wage discrimination: Job title

gap	worker fe	firm fe	job fe
-0.2560	-0.1547	-0.0580	-0.0433

Note: Decompositions based on Gelbach (2016).

TABLE 5. Conditional Decomposition of the Gender Wage Gap

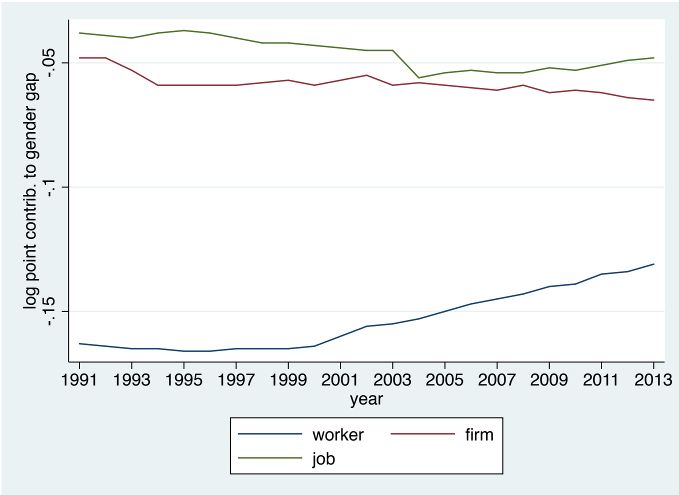

FIGURE 7: Conditional decomposition of the gender wage gap, separately by year

Overall, the combination of the evolution of the three sources of heterogeneity resulted in a tiny (0.5 log points) decrease in the gender wage gap over 22 years.

The gender gap heterogeneity

Whereas the approach based upon the high-dimensional fixed effect regression fully accounts for the distribution of workers across firms and job title cells, it is silent regarding the heterogeneity of gender gaps within the firm and job title. Firm specific gender gaps have been interpreted as evidence of gender discrimination that emerges from the shortfall in women's relative bargaining power (Card *et al.* (2016)). Here we extend our previous approach to accommodate the estimation of firm specific and job specific gender gaps. In essence, for the firm case, we estimate the following regression model:

$$\ln w_{ifjt} = \mathbf{x}_{ift}\boldsymbol{\beta} + \varphi_f + \gamma_f g_i + \varepsilon_{ifjt} \quad . \tag{7}$$

where equation (1) is generalized to include a firm fixed effect (φ_f) and a firm-specific gender effect (γ_f). It should be noted that we are not including a worker fixed effect and so the firm gender gap is not filtered from the presence of unobserved individual heterogeneity. The identification of the firm gender parameter in conjunction with the worker fixed effect would require additional normalization restrictions in order to retain a common scale.

The results from this procedure are exhibited in figure 8, where the empirical distribution of firm specific wage gender gaps for 1991 are contrasted with those of 2013. The histogram may be interpreted as the distribution of discriminating employers (or, in the sense of Card *et al.* (2016), as reflexion of the relative bargaining power of women). The graph indicates that most employers have negative gender wage gaps and that the distribution of the gender gaps only mildly improved from 1991 to 2013. It is interesting to notice that a non-negligible fraction of employers has positive gender gaps.

Whether this outcome signals the true distribution of discriminating employers or is just a product of sampling variation remains to be solved. An indication that it is not simply the consequence of sampling variation can be argued from the fact that firm specific gender gaps are highly correlated with the firm level segregation (-0.476). The notion that higher proportion of females leading to more negative firm gender gaps is consistent with the idea of a shortfall in women bargaining power.

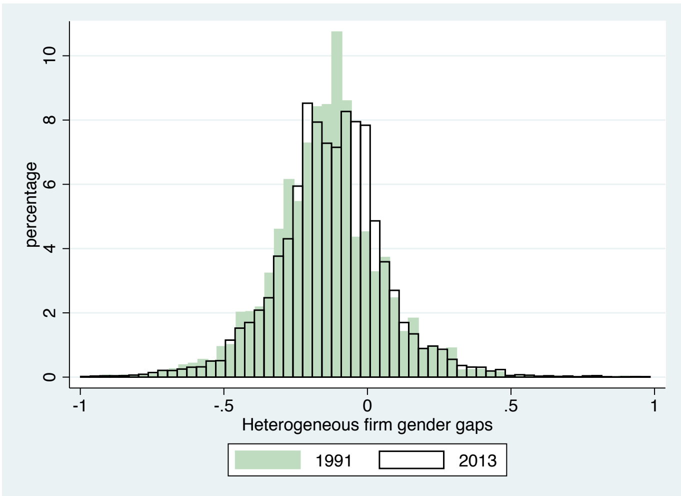

FIGURE 8: Heterogeneous firm gender gaps

The distribution of job title specific wage gender gaps is much less dispersed, in particular in 2013. In contrast with the firm gender gaps which are sensitive to the segregation at the firm level, these are poorly predicted by the measure of job title segregation (correlation equals 0.006). Whereas job title segregation leads to lower mean wages, it does not lead to larger gender gaps along the job title dimension. Put differently, whereas firm segregation leads to higher gender gaps, job title segregation leads to lower wages. This latter result is in line with Groshen (1991) and is consistent, for example, with the idea that some occupations may be overcrowded by women.

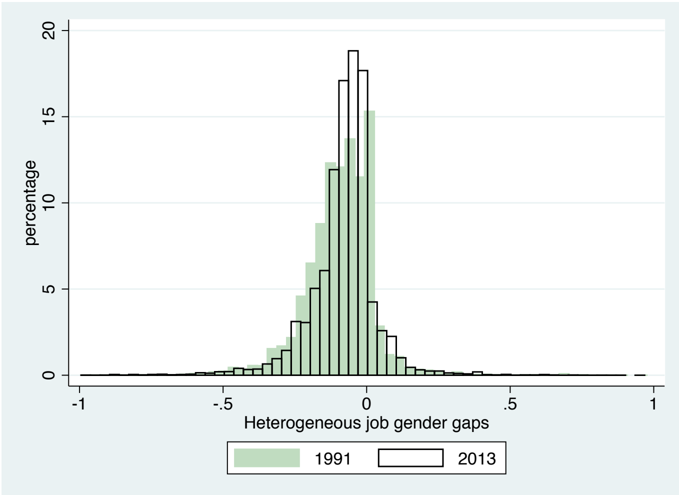

FIGURE 9: Heterogeneous job title gender gaps

Conclusions

Over the 1991-2013 period, there was a notable increase in the feminization rate of Portuguese labour market. At the same time, the average wages of women approached significantly those of the men. In this study, we argue that the fall in the gender wage gap is largely the result of a compositional change (not a structural effect), due to the fact that the women that joined the labor market detained higher level of general and specific human capital.

This means that the adjusted gender wage gap remained roughly constant at around 25 percent over the period. We show that gender plays an important role in the allocation of workers across firms with distinct wage policies. Indeed, if workers were randomly allocated to firms, the gender gap would be reduced by 5.8 percentage points. Similarly, if workers were randomly selected into job titles, the gender gap would be reduced by 4.3 percentage points. Overall, if workers were randomly sorted into firms and job titles, the gender gap would be reduced by about two fifths.

The allocation of female workers to firms and job-titles did not improve over the last two decades. In fact it deteriorate somewhat, since in 2013 females tend to be less present in firms and job titles with more generous wage policies. In compensation, the role of unobserved skills favored a small decrease of the gender gap. This may either reflect less gender discrimination or improved ability.

Firm segregation, that is, the feminization rate at the firm level, leads to higher firm specific gender gaps. In contrast, job title segregation leads to lower wages but not larger job title specific gender gaps.

References

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). "High Wage Workers and High Wage Firms." *Econometrica*, 67(2), 251–334.

Altonji, Joseph G. and Rebecca M. Blank (1999). "Race and gender in the labor market." In *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card, pp. 3143–3259. Elsevier.

Autor, David H. (2013). "The "task approach" to labor markets : an overview." *Journal for Labour Market Research*, 46(3), 185–199.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2006). "The Polarization of the U.S. Labor Market." *American Economic Review*, 96(2), 189–194.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008). "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics*, 90(2), 300–323.

Bertrand, C. Goldin, M. and L. F. Katz (2010). "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics*, 2(3), 228–255.

Blau, F. and L. Kahn (2016). "The gender wage gap: Extent, trends and explanations." *Journal of Economic Literature*, forthcoming.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016). "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *The Quarterly Journal of Economics*, forthcoming.

Cardoso, Ana Rute, Paulo Guimarães, and Pedro Portugal (2016). "What drives the gender wage gap? A look at the role of firm and job-title heterogeneity." *Oxford Economic Papers*, 68(2), 506–524.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). "Revisiting the German Wage Structure." *The Quarterly Journal of Economics*, 124(2), 843–881.

Gelbach, Jonah (2016). "When do covariates matter? and which ones, and how much?" *Journal of labor Economics*, forthcoming.

Goldin, C. (2014). "A grand convergence: Its last chapter." *American Economic Review*, 104(4), 1091–1119.

Goos, Maarten and Alan Manning (2007). "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." *The Review of Economics and Statistics*, 89(1), 118–133.

Groshen, Erica L. (1991). "The structure of the female/male wage differential: Is it who you are, what you do, or where you work?" *Journal of Human Resources*, 26(3), 457–472.

Guimarães, Paulo and Pedro Portugal (2010). "A simple feasible procedure to fit models with high-dimensional fixed effects." *Stata Journal*, 10(4), 628–649.

Jarrell, S. B. and T. D. Stanley (2004). "Declining bias and gender wage discrimination? a meta-regression analysis." *Journal of Human Resources*, 39(3), 828–838.

Machado, José A. F. and José Mata (2005). "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of Applied Econometrics*, 20(4), 445–465.

Olivetti, C. and B. Petrongolo (2008). "Unequal pay or unequal employment? A cross-country analysis of gender gaps." *Journal of Labor Economics*, 26(4), 621–654.

Stanley, T. and S. B. Jarrell (1998). "Declining bias and gender wage discrimination? a meta-regression analysis." *Journal of Human Resources*, 33(4), 947–973.

Torres, Sónia, Pedro Portugal, John T. Addison, and Paulo Guimarães (2013). "The sources of wage variation: a three-way high-dimensional fixed effects regression model." Working Papers w201309, Banco de Portugal, Economics and Research Department, URL `http://ideas.repec.org/p/ptu/wpaper/w201309.html`.

Weeks, David L. and Donald R. Williams (1964). "A Note on the Determination of Connectedness in an N-Way Cross Classification." *Technometrics*, 6(3), 319–324.