Is the Phillips curve dead? - results for Portugal

Sara Serra
Banco de Portugal

April 2018

Abstract
This article assesses the forecasting performance over recent years of Phillips curves for Portugal. Phillips curves are estimated for a large variety of slack measures and evaluated in terms of their out-of-sample performance in real time. The issues of time variation of parameters and possible non-linearities are also explored. The findings suggest that there is no unique best specification for the Phillips curve over time, and therefore a set of these models, considering different slack measures, should be considered. Furthermore, there is some evidence of flattening of the traditional linear Phillips curve in recent years, which is possibly related to non-linearities in the model. Overall, the Phillips curve maintains some forecasting power for inflation when compared to a naïve benchmark. (JEL: E31, E37)

Introduction

The Phillips curve (PC), introduced in 1958 by A. W. Phillips, postulates the existence of a negative relationship between unemployment and inflation, or of a positive relationship between output and inflation. Given the importance of the link between inflation and economic activity for monetary policy, it quickly become popular as an instrument of economic analysis. Over time, the PC has been subject to some criticism, with its standard formulation in the literature being adjusted accordingly.[1] Initially taken as a long term economic relationship, in the late 70's, with the work of Phelps and Friedman, it became seen as a short-term trade-off, dependent on inflation expectations.

More recently, both in Europe and the US, the Great Recession brought along the so called missing disinflation: inflation appears to have reacted less to the amount of slack in the economy than suggested by PC models

Acknowledgements: I thank the members of the LIFT team, particularly those from the ECB, whose codes were the starting point for the ones used in this article. I also thank José Ramos Maria, Miguel Gouveia and the participants of a seminar at Banco de Portugal for useful comments and opinions. The opinions expressed in the article are those of the author and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the author.

E-mail: srserra@bportugal.pt

1. See Macklem (1997) for a summary of the evolution of the theory surrounding the Philips curve.

(Albuquerque and Baumann (2017) and ECB (2014)). The ensuing recovery, on the other hand, has shown a weaker increase in inflation than expected given the reduction in unemployment ("missing inflation") (Ciccarelli and Osbat (2017)). PC models were one of the standard instruments used to determine this counterfactual path for inflation, which has led to questions about their reliability in explaining price developments.

The standard Phillips curve formulation is often limited, being unable to capture many aspects not related to the amount of slack in the economy (administered prices, taxes, import prices), and affected by uncertainty regarding the relevant measure of slack to be considered. It is also potentially flawed, given possible non-linearities (namely due to downward price rigidity), time varying parameters and state dependent behaviour. As stated by Dotsey *et al.* (2017), the Phillips curve is likely to be unstable, being a reduced form model which is a function of deeper structural parameters that change over time.

Notwithstanding the standard Phillips curve limitations, namely for forecasting (ECB (2014)), it has remained a central instrument of analysis for central bankers.[2] Teles and Garcia (2016) analyse the possible usefulness of Philips curves for monetary policy definition in the current context.

Several authors have shown that refinements of this instrumental are able to reduce the recent puzzles surrounding inflation. Some of these refinements deal with non-linearities in the Phillips curve.

This article aims at analysing PC models for Portugal, drawing from the work developed for the Low Inflation Team (LIFT) (Ciccarelli and Osbat (2017)). The issue of parameter instability over time and its potential relationship with nonlinearities is explored, and the forecasting ability of several PC specifications is assessed. Given that the measurement of slack has several caveats, specially when resorting to output or unemployment gap measures, the forecasting performance of the PC is assessed in real time.

The structure of the article is as follows: the following sections present the baseline specification for the PC and the details on the variables considered in estimation and data transformations. Then the selection process of PC specifications is described and the possibility of time-varying parameters and non-linear effects is explored. Finally, the forecasting performance of selected PC specifications is analysed through a real time exercise.

Phillips curve baseline specification

The baseline specification takes the form of the hybrid Phillips curve equation considered in Albuquerque and Baumann (2017), which is given by:

2. See for example Draghi (2017), Constâncio (2015) or Yellen (2013).

$$\pi_t = \theta_0 + \alpha E_t(\pi^*_{t+1}) + \sum_{i=1}^{n} \beta_i \pi_{t-i} + \sum_{j=0}^{m} \gamma_j pm_{t-j} + \delta \hat{y}_{t-1} + \varepsilon_t \tag{1}$$

where π is actual inflation, π^* is expected inflation, pm is a measure of import prices, $\hat{y}$ is a variable that measures available slack in the economy and E_t is the expectations operator. Explanatory variables are in general considered with a lag to make results more robust to potencial endogeneity (Albuquerque and Baumann (2017)).

The possibility of more restricted models (with the exclusion of one or several regressors), with the limit option of a purely autoregressive (AR) model is also explored.

A first exercise, along the lines of Albuquerque and Baumann (2017) and Ciccarelli and Osbat (2017), was to estimate PC models for a lage set of inflation, slack, import price and inflation expectations measures. This approach tries to address the fact that there is large uncertainty in the measurement of slack (Yellen (2013)) and inflation expectations and at the same time access which inflation concept is more suitable to be fitted by the PC.

Estimation details

This article will be focused on inflation measures stemming from consumer prices, namely the Harmonized Index of Consumer prices (HICP). A standard approach to Phillips curves is to consider wages as the relevant inflation measure, but administrative changes to wages during recent years hinder the quality of the data and may distort results. In addition, wage data has the additional problem of data revisions, which create an additional source of uncertainty in PC estimation.

Below there are some details on the variables considered. Data is in general seasonally adjusted, with the exception of some unemployment measures and of the inflation expectations targeted to the annual rate of change in prices.

- **Inflation measures**: the overal HICP and the HICP excluding energy and food (the most volatile components) are considered. Due to the importance of indirect tax increases in 2011 and 2012, which are administratively driven and may distort results, HICP and HICP excluding energy and food are also considered correcting for the impact of indirect taxes (see Ciccarelli and Osbat (2017) for motivation on the use of these variables, and the impact in the Portuguese case). The estimation is made on the basis of data expressed in annualized quarter-on-quarter rates of change.

- **Slack measures**: A wide range of slack measures was considered. These include several output gap estimates, both model based (Cobb Douglas, CES, UCM) and filter based (HP, BK, CF) (see Banco de Portugal (2017) for more details on these measures). The output gaps published by the European Commission (EC) and International Monetary Fund (IMF) are also considered.[3] Several measures related to unemployment were also considered. These include the unemployment rate, (both the headline and a broader measure[4]) and the unemployment gap. The short-term unemployment rate was also considered because some authors argue that it may be more representative of cyclical pressure to inflation than the headline unemployment rate (Dotsey *et al.* (2017)). The unemployment recession gap, defined as the difference between the current unemployment rate and the minimum unemployment rate over the current and previous eleven quarters, was also included (Stock and Watson (2010)). The combined unemployment and labour participation gap (UPRGAP), used by Albuquerque and Baumann (2017), aims at capturing existent slack in the labour market arising from workers that left it temporarily, like discouraged workers.[5] In addition, slack measures derived from the EC business surveys were also considered, namely capacity utilisation and demand and labour as factors limiting production in manufacturing. Finally, the real GDP and real unit labour costs were also included in this set of explanatory measures, expressed in annualized quarter-on-quarter rates of change. All other variables were considered in levels. In the case of the variables related to unemployment, the sign was flipped, to facilitate coefficient comparability.

- **Import price measures**: the options considered include overall import prices and import prices of goods. In addition, these two aggregates excluding energy are also considered. Data is expressed in annualized quarter-on-quarter rates of change.

- **Inflation expectation measures**: the information set includes past inflation (average of past four year-on-year rates of change), Consensus forecasts (both for current year and next year) and EC consumer survey expectations for price developments in the following 12 months. For the

3. This data is annual, and was converted to quarterly frequency using a cubic spline.

4. The broad measure of unemployment includes, along with unemployed, discouraged workers and a measure of involuntary part-time work. For more details see Statistics Portugal press release: `https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=281328836&DESTAQUEStema=5414314&DESTAQUESmodo=2`.

5. The trend labour participation rate implicit in this indicator was calculated as the HP filtered raw data for this variable.

latter variable the level difference vis-à-vis the same period of the previous year is considered, while for the remaining no transformation is applied.

Overall, this information set and subsets were one or several regressors are excluded amount to about 500 different specifications for each inflation aggregate.

Three estimation samples were considered. The start of all samples is 1996Q1 in the case of headline HICP measures or 1997Q1 in the case of "core" inflation measures, but in some cases could be more limited due to regressor availability. Sample 1 ends in 2007Q4, which allows an analysis of the PC behaviour over the Great Recession, a strongly disruptive period for the global economy. Sample 2 ends in 2011Q4, given that after that period Portugal and the euro area began a disinflation path not captured by traditional forecasting models (Ciccarelli and Osbat (2017)). Finally, the full sample, ending in 2017Q4, was also considered. For the smaller samples, a set of out-of-sample conditional (on the actual path of slack, import and expectation measures) forecasts were estimated. This allows an evaluation of models without the noise brought about by the need of forecasting regressors. Forecasts are dynamic, in the sense that the projected HICP for one period serves as autoregressive term in the following periods. For the HICP, the autoregressive lag order included in the model was set to three as a result of trial and error tests on the significant lag order while maintaining the expected sign of coefficients. For import prices the optimal lag to be included in the model was optimized on the basis of the Schwarz information criteria. For slack variables the first lag was considered given that is the most standard approach in the literature, as the use of contemporaneous values may lead to potential endogeneity problems.

Model selection

Table A.1 shows that the performance of PC forecasts is better for "core" aggregates, that do not include the more volatile components (food and energy), given that the average and median Root Mean Squared Errors (RMSE) are lower and in some cases the dispersion (min-max range) is also lower. For the overall inflation measures, the overwhelming majority of PC fail to forecast the decline in inflation that took place from 2008 and 2012 onwards.

Results for the core aggregates with constant taxes increase in accuracy from 2012 onwards given that the direct impact of indirect taxes increases that took place in 2011 and 2012 is excluded. This improvement is not however enough to generate a better performance than for the overall HICP excluding energy and food, which was chosen as the relevant aggregate of analysis.

For each subsample, only specifications for which the slack variable was significant with the expected sign were selected. For the other variables, this selection process allowed for the possibility that they were not significant, but if so, only included when they had the expected sign. The group of model specifications that satisfied these restrictions simultaneously over the three samples was selected, thus focusing the analysis on PC specifications that are relatively stable over time. This requirement implied the exclusion of some slack variables, despite the fact that most of them would be included if only the full sample was analysed: real unit labour costs, capacity utilization and demand as a restriction to production, the IMF output gap, the unemployment recession gap and real GDP.

Import prices are rarely significant in sample 1, appearing more frequently as a significant regressor with expected sign as the sample size increased. For the top 20 out-sample performing models, these variables are never present in the sample up to 2008 and rarely present in the sample up to 2012. This suggests that this variable only gains importance in the most recent period, which is a sign of parameter instability of the PC and may be a result of globalization (Constâncio (2015)). The same results apply to inflation expectations measures. Given their apparent growing importance over time, one import price variable (goods excluding energy) and one expectation variable (consumers' survey price expectations) were selected, given that they are present in the top out-of-sample performing specifications for the sample ending in 2011Q4.

There is not a close relationship between the best performing models in terms of the RMSE of out-of-sample errors and in terms of in sample fit. The R^2 (coefficient of determination) is nevertheless relatively low for all specifications in all samples, never reaching a value much above 50%.

After this process of selection, a group of about 50 models is left. Figures (1) and (2) show the conditional forecasts generated by this set of specifications for the top 20 group with lower RMSE.

Conditional forecasts for inflation - year-on-year rate of change

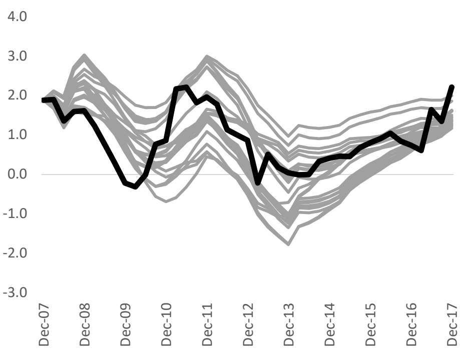

FIGURE 1: Sample 1 (2008Q1-2017Q4)

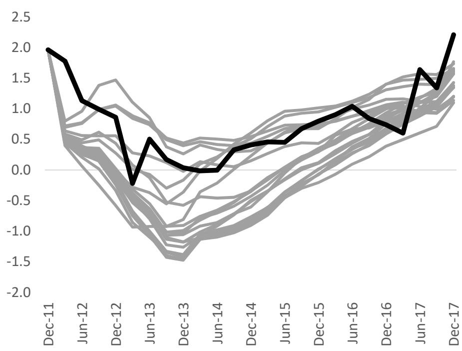

FIGURE 2: Sample 2 (2012Q1-2017Q4)

Note: Inflation is measured by the year-on-year rate of change of the HICP excluding energy and food.

The specifications chosen on the basis of sample 1 appear to capture inflation developments quite well over the out-of-sample period, despite some lag in reflecting the 2008 desinflation and not being able to fully follow the most recent increase. On the other hand, the majority of models estimated over sample 2 tend to overstate the 2012 desinflation. This may partly reflect the increases in indirect taxes that affected this period, that the PC is not able to capture.[6]

Given that the selected models are relatively similar in terms of out-of-sample performance, the remaining part of this article focuses on an even more restricted sample of models. These were selected on the basis of the criteria that they are the top best performing models in terms of out-of-sample forecasts for sample 1, while simultaneously being in the top RMSE also for sample 2 and in the top R^2 for the full sample. Tables A.2 to A.4 report the main estimation results for the top 20 performing models in terms of RMSE in case of samples 1 and 2 and in terms of R^2 for sample 3. The selected measures of slack are the short-term unemployment rate, the survey question related to labour as a limiting factor to production in the manufacturing industry and the Cobb Douglas and CES production function output gaps.[7] Import prices and inflation expectations are not included in any of these "best" models, or when included they are not significant, possibly because the series considered capture the impact of supply shocks and inflation expectations in an imperfect way.

Table A.5 shows the main estimation results for the models that include the selected slack variables and exclude both import prices and expectation measures.

The coefficients on slack variables are all strongly significant. For the output gap variables, which are expected to be nil in the long run, we can compute the long-term expected inflation as the value of the constant divided by 1 minus the sum of the autoregressive coefficients. This yields values close to 2 per cent for the three sample periods. The long run coefficients on slack, computed in the same vein, yield about 0.6, a value broadly in line with those found for other euro area countries (Ciccarelli and Osbat (2017)). The output gap measures clearly outperform the other in terms of RMSE in sample 2, while results are more similar across specifications for sample 1.

The results also show some time variation in the coefficients pertaining to slack variables, namely for all measures considered except the survey question there is a decline in the coefficient when moving from sample 1 to sample

6. The PC for the HICP excluding energy and food with constant taxes, which excludes the impact of these factors, is however even worse for the same sample period. This is because inflation is also underestimated from 2013 onwards, but it is grossly overestimated in 2012.

7. The European Commission output gap would also be a selected indicator according to these criteria, but was not included because import prices are not significant with expected sign, but when excluded from the equation the slack variable becomes non-significant.

3. This is related to the possibility of PC flattening that arose with the 2012 missing inflation puzzle (Constâncio (2015)). To test this possibility, along with a more general one of parameter instability in the PC due to non linearities, a rolling window exercise was performed. This is presented in the following section.

Parameter instability and non-linear Phillips curves

The initial window considered for the rolling sample was sample 1, and from then onwards the model was reestimated moving the window forward by one period.

Rolling Regression estimates of coefficients on slack measures

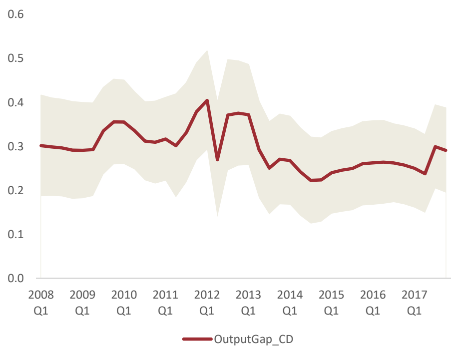

Figure 3: Slack measure: Cobb-Douglas output gap

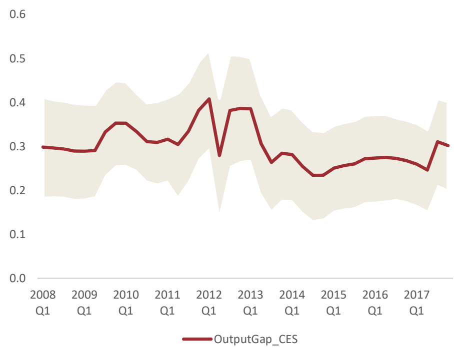

Figure 4: Slack measure: CES output gap

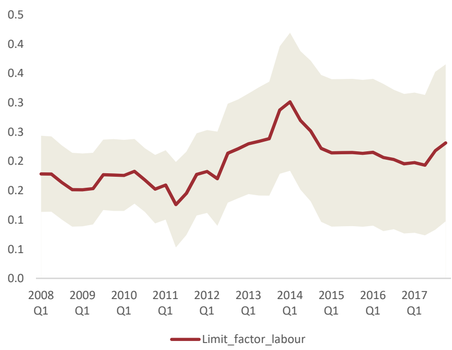

Figure 5: Slack measure: labour as a factor limiting production

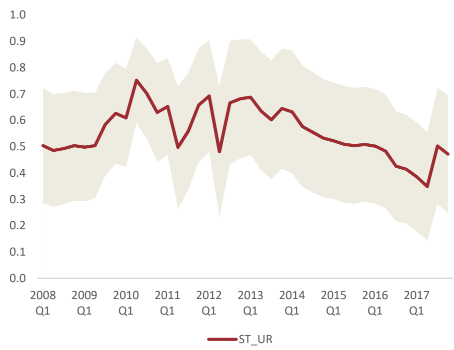

Figure 6: Slack measure: short-term unemployment rate

Note: The shaded area is defined by parameter estimate +/- one standard deviation. The dates on the x-axis define the last quarter included in the rolling regression.

Results, shown in Figures 3 to 6, display in most cases an increase in the coefficient of the PC on slack over the periods of the last two recessions (considering the 2009 decline in GDP as a separate recession) and a posterior decline to levels below those observed prior to the 2008 financial crisis, supporting the thesis of flattening of the Phillips curve. However, this conclusion is contingent on the slack measured considered (the survey-based indicator yields the opposite conclusion) and on the relevant concept of inflation considered (Ciccarelli and Osbat (2017)). This evidence however supports the idea vastly found in the literature (Dotsey *et al.* (2017)) that PC coefficients are time varying, what may be the result of non-linearities in the model. Several hypothesis for the flattening of the PC have been put forward in the literature (see Constâncio (2015) for a summary). One possibility is that inflation only reacts to slack when changes are large enough, given that there are menu costs to changing prices. Another possibility is that during a recovery firms have unused spare capacity, and do not feel the pressure to raise prices until the installed capacity is reached (Macklem (1997)). This inertia effect should be stronger in a low inflation environment.

To test for this hypothesis a threshold model was estimated, defined in equation 2:

$$\pi_t = \theta_0 + \theta_1 I_{outT} + \sum_{i=1}^{n} \beta_i \pi_{t-i} + \delta_1 I_{outT} \hat{y}_{t-1} + \delta_2 (1 - I_{outT}) \hat{y}_{t-1} + \varepsilon_t \qquad (2)$$

where I_{outT} is a dummy variable that takes the variable of 1 when $\hat{y}$ falls outside the thresholds and zero otherwise.

The thresholds considered are given by the 30th and 70th percentile of the distribution of $\hat{y}$.[8]

Table A.6 shows the result of this estimation. For the specifications in which the slack variable is the output gap, only the values outside the threshold are significant to explain changes in inflation, for all samples considered. There is no evidence of a change in the constant of the equation when slack variables lie outside the threshold. For the other two slack variables considered results suggest that this type of disaggregation does not help explaining inflation. The out-of-sample forecasting performance of the models where thresholds are significant worsens vis-à-vis the previous exercise, but this possibly reflects the fact that the non significant regressors are being used to produce these forecasts.

Real time forecasting exercise

As a final exercise, the forecasting performance of selected models is evaluated in real time. The real time analysis is particularly important for output gap measures, where data revisions can be substantial (Banco de Portugal (2017)).

8. Results are qualitatively similar with thresholds of 20th-80th and 25th-75th percentiles.

Therefore, adding to the potential forecasting errors of the PC arising from uncertainty regarding the model used and the projected path of the regressors, there is additional uncertainty due to possible revisions in potential output estimates. The other two slack variables, the short-term unemployment rate and the survey question related to labour as a limiting factor to production, are not subject to revisions, and have no projections available. In this case the actual values were used to produce the conditional forecasts[9], which favours PC results in this case, given that the uncertainty associated to regressors' projections is ruled out. In this exercise, the relative performance of the PC is confronted with a naïve random walk model and with the half-yearly Eurosystem projections. The choice of the random walk benchmark allows an evaluation of the PC performance against a very simple and standard reference in the literature, but which provided very good results in terms of inflation forecasting on a short to medium run horizon (Teles and Garcia (2016)). On the other extreme, the Eurosystem projections provide a very demanding benchmark, given that they are computed at a very detailed level and benefit from experts' judgement. In this case the relevant question is whether a simple PC model can provide inflation forecasts as accurate as this benchmark. Threshold models that include only the values of the slack variables outside the treshold are also considered for the cases where they were found significant, namely for output gap specifications. In this case the thresholds were defined also in real time, i.e., taking into account the distribution of data available at each vintage.

Forecasts are produced from 1 to 8 steps ahead and evaluated in terms of their relative RMSE vis-à-vis the benchmarks for each horizon. Moreover, the significance of these relative differences is tested with the Diebold-Mariano (Diebold and Mariano (1995)) test with Harvey *et al.* (1997) correction, considering a 10% significance level. The vintages available for this evaluation range from the June 2007 projection exercise (with observed data for the HICP up to 2007Q1) to the December 2017 projection exercise (with observed data for the HICP up to 2017Q3), thus 22 vintages in total.

The relative RMSE of the forecasts vis-à-vis the benchmark of the random walk and the Eurosystem projections are presented respectively in Tables A.7 and A.8. The relative performance of PC models is in general better than the random walk, with the exception of the short-term unemployment rate model. This outperformance is statistically significant for some medium to long term horizons. The threshold models do not perform better than their standard counterpart. On the other hand, the Eurosystem projections are only better than the random walk in a statically significant way for one step ahead forecasts. Considering the Eurosystem projections as the benchmark,

9. When necessary, the dataset for the short-term unemployment variable was extended beyond 2017Q4 with the quarterly changes in the latest unemployment projections of the Eurosystem.

the PC relative RMSE are also in general lower than 1 for medium to long-run horizons, but this difference is not statistically significant. The AR and random walk models have a relative RMSE higher than 1, which is only significant for short-term horizons.

Conclusion

Despite some difficulties in coping with inflation fluctuations since the Great Recession, Phillips curves remain a staple of economic analysis for central bankers. This article resorted to a large diversity of slack measures to estimate Phillips curves for Portugal. These models have some forecasting power for inflation, but results have shown that the best slack measure is not constant over time, and therefore it is preferable to rely on a diversified set of Phillips curves. There is some evidence that some nonlinearities are present in Phillips curves estimation, but further work is needed on this issue and how best to tackle it for forecasting purposes.

Appendix: Tables

	HICP	HICPxENFOOD	HICP_CT	HICP_CT_xENFOOD	HICP	HICPxENFOOD	HICP_CT	HICP_CT_xENFOOD
Sample up to:	2007Q4				2011Q4			
Max	5.4	2.7	4.8	6.7	3.1	3.3	3.4	2.6
Min	1.1	0.6	1.1	0.6	0.5	0.4	0.5	0.5
Average	1.7	1.3	1.8	1.4	1.6	1.0	1.5	1.0
Median	1.6	1.3	1.6	1.2	1.5	0.9	1.5	0.9

TABLE A.1. RMSE of conditional forecasts

Slack variable	slack coefficient	constant	Import prices Included?	Import price coefficient	Expectations Included?	Expectations coefficient	RMSE	R^2
EC_OG	0.22	0.99	yes	not significant	no		0.62	0.36
Limit_factor_labour	0.18	0.17	no		no		0.63	0.39
Limit_factor_labour	0.18	0.30	no		yes	not significant	0.64	0.45
OutputGap_CES	0.29	1.10	yes	not significant	no		0.68	0.41
Limit_factor_labour	0.19	0.37	yes	not significant	no		0.70	0.45
OutputGap_CD	0.29	1.08	yes	not significant	no		0.72	0.41
ST_UR	0.50	2.95	no		no		0.72	0.36
OutputGap_CES	0.28	0.94	no		yes	not significant	0.74	0.42
OutputGap_BK	0.37	0.88	no		no		0.75	0.37
OutputGap_CES	0.30	0.90	no		no		0.77	0.38
OutputGap_CD	0.28	0.93	no		yes	not significant	0.79	0.42
OutputGap_BK	0.32	1.05	yes	not significant	no		0.82	0.37
OutputGap_CD	0.30	0.88	no		no		0.83	0.38
OutputGap_HP	0.34	1.05	no		no		0.84	0.36
ST_UR	0.68	3.99	yes	not significant	no		0.85	0.48
OutputGap_CF	0.20	1.19	no		no		0.97	0.37
ST_UR	0.70	3.98	no		yes	not significant	0.99	0.48
OutputGap_UCM	0.44	1.90	yes	not significant	no		1.08	0.44
OutputGap_UCM	0.43	1.78	no		yes	not significant	1.10	0.44
OutputGap_CF	0.20	1.38	yes	not significant	no		1.14	0.39

TABLE A.2. Main estimation results for sample 1997Q1-2007Q4

Notes: the acronyms for slack variables stand for (in the order they are presented): the EC output gap, the survey question related to labour as a restrictive factor to production, the CES output gap, the Cobb-Douglas output gap, the short-term unemployment rate, the Baxter-King output gap, the Hodrick-Prescott output gap, the Christiano-Fitzgerald output gap and the unobserved components model output gap. Banco de Portugal (2017) provides details on the output gap measures.
The shaded variables denote the selected slack variables.

Slack variable	slack coefficient	constant	Import prices Included?	Import price coefficient	Expectations Included?	Expectations coefficient	RMSE	R^2
Limit_factor_labour	0.16	0.09	no		yes	not significant	0.45	0.48
ST_UR	0.46	3.14	no		no		0.46	0.39
Limit_factor_labour	0.16	0.16	no		no		0.46	0.39
OutputGap_CF	0.16	0.51	no		no		0.47	0.36
Limit_factor_labour	0.17	0.17	yes	not significant	no		0.48	0.43
ST_UR	0.58	3.87	yes	not significant	no		0.56	0.47
EC_OG	0.31	0.73	yes	not significant	no		0.56	0.43
Unemployment_gap	0.28	1.60	no		no		0.68	0.39
Unemployment_gap	0.29	1.68	yes	not significant	no		0.71	0.43
OutputGap_UCM	0.33	1.67	no		no		0.73	0.41
ST_UR	0.65	4.17	no		yes	not significant	0.74	0.54
OutputGap_UCM	0.34	1.75	yes	0.04	no		0.76	0.47
UPRGAP	0.31	1.77	no		no		0.81	0.40
UPRGAP	0.32	1.83	yes	not significant	no		0.84	0.44
OutputGap_CES	0.35	1.12	yes	not significant	no		0.95	0.49
OutputGap_CES	0.36	1.11	no		no		0.95	0.45
Labour_slack	0.21	4.21	no		no		0.97	0.41
OutputGap_CES	0.32	1.00	no		yes	not significant	1.00	0.51
OutputGap_CD	0.35	1.11	yes	not significant	no		1.01	0.48
OutputGap_CD	0.36	1.10	no		no		1.01	0.45

TABLE A.3. Main estimation results for sample 1997Q1-2011Q4

Notes: the acronyms for slack variables stand for (in the order they are presented): the survey question related to labour as a restrictive factor to production, the short-term unemployment rate, the Christiano-Fitzgerald output gap, the EC output gap, the unemployment gap, the unobserved components model output gap, the combined unemployment and labour participation gap, the CES output gap, the measure of unemployment in broad sense and the Cobb-Douglas output gap. Banco de Portugal (2017) provides details on the output gap measures.
The shaded variables denote the selected slack variables.

Slack variable	slack coefficient	constant	Import prices Included?	Import price coefficient	Expectations Included?	Expectations coefficient	R^2
OutputGap_CF	0.21	0.67	no		no		0.34
OutputGap_HP	0.25	0.83	no		no		0.34
OutputGap_BK	0.30	0.89	no		no		0.36
UR	0.22	3.39	no		no		0.38
Labour_slack	0.17	4.16	no		no		0.38
Limit_factor_labour	0.22	0.23	no		no		0.38
UR	0.21	3.35	yes	0.05	no		0.45
Unemployment_gap	0.26	2.05	no		no		0.38
UPRGAP	0.26	2.14	no		no		0.39
OutputGap_UCM	0.29	2.07	no		no		0.39
OutputGap_CF	0.18	0.70	yes	0.05	no		0.40
OutputGap_CD	0.27	1.64	no		no		0.41
OutputGap_CES	0.28	1.64	no		no		0.41
Labour_slack	0.16	4.02	no		yes	0.02	0.42
OutputGap_BK	0.27	0.89	yes	0.05	no		0.42
OutputGap_CD	0.24	1.56	no		yes	not significant	0.44
EC_OG	0.28	1.18	yes	0.05	no		0.44
OutputGap_CES	0.25	1.57	no		yes	not significant	0.44
Limit_factor_labour	0.20	0.36	no		yes	0.02	0.44
UPRGAP	0.24	2.15	no		yes	0.02	0.44

TABLE A.4. Main estimation results for sample 1997Q1-2017Q4

Notes: the acronyms for slack variables stand for (in the order they are presented): the Christiano-Fitzgerald output gap, the Hodrick-Prescott output gap, the Baxter-King output gap, the unemployment rate, the measure of unemployment in broad sense, the combined unemployment and labour participation gap, the unobserved components model output gap, the Cobb-Douglas output gap, the CES output gap and the EC output gap. Banco de Portugal (2017) provides details on the output gap measures.
The shaded variables denote the selected slack variables.

Slack variables	sample 1: 1997Q1-2007Q4 sum of AR coefficients	slack	constant	RMSE	R^2
ST_UR	0.60	0.50 (0.22)	2.95 (1.08)	0.72	0.36
Limit_factor_labour	0.44	0.18 (0.07)	0.17 (0.56)	0.63	0.39
OutputGap_CD	0.53	0.30 (0.11)	0.88 (0.52)	0.83	0.38
OutputGap_CES	0.52	0.30 (0.11)	0.90 (0.52)	0.77	0.38

Slack variables	sample 3: 1997Q1-2011Q4 sum of AR coefficients	slack	constant	RMSE	R^2
ST_UR	0.47	0.46 (0.18)	3.14 (1.06)	1.12	0.39
Limit_factor_labour	0.46	0.16 (0.06)	0.16 (0.4)	1.13	0.39
OutputGap_CD	0.41	0.36 (0.1)	1.10 (0.38)	0.50	0.45
OutputGap_CES	0.41	0.36 (0.1)	1.11 (0.38)	0.51	0.45

Slack variables	sample 2: 1997Q1-2017Q4 sum of AR coefficients	slack	constant	R^2
ST_UR	0.35	0.49 (0.14)	3.53 (0.9)	0.38
Limit_factor_labour	0.27	0.22 (0.06)	0.23 (0.28)	0.38
OutputGap_CD	0.23	0.27 (0.07)	1.64 (0.38)	0.41
OutputGap_CES	0.23	0.28 (0.07)	1.64 (0.37)	0.41

TABLE A.5. Main estimation results for selected models

Notes: figures between brackets refer to the standard deviation of the corresponding coefficients. The acronyms for slack variables stand for (in the order they are presented) the short-term unemployment rate, the survey question related to labour as a restrictive factor to production, the Cobb-Douglas output gap and the CES output gap.

slack variables	sample 1: 1997Q1-2007Q4 Sum of AR coefficients	OutT	InT	OuT Dummy	Constant	RMSE	R^2
ST_UR	0.66	-0.09 (0.63)	-0.25 (0.47)	1.89 (0.73)	-0.69 (2.12)	1.53	0.47
LF_labour	0.64	-0.01 (0.15)	0.11 (0.23)	1.62 (0.79)	-0.04 (0.99)	1.35	0.46
OutputGap_CD	0.48	0.61 (0.23)	-0.39 (0.57)	-0.51 (0.38)	0.74 (0.5)	2.15	0.45
OutputGap_CES	0.48	0.60 (0.22)	-0.42 (0.57)	-0.49 (0.68)	0.72 (0.75)	2.02	0.45

slack variables	sample 2: 1997Q1-2011Q4 Sum of AR coefficients	OutT	InT	OuT Dummy	Constant	RMSE	R^2
ST_UR	0.53	0.33 (0.2)	0.37 (0.19)	0.45 (0.39)	2.32 (1.21)	0.59	0.42
LF_labour	0.52	0.13 (0.08)	0.18 (0.14)	0.50 (0.46)	-0.14 (0.55)	0.58	0.40
OutputGap_CD	0.29	0.62 (0.12)	0.02 (0.31)	-0.89 (0.38)	1.56 (0.5)	2.21	0.55
OutputGap_CES	0.28	0.62 (0.12)	0.02 (0.32)	-0.91 (0.37)	1.58 (0.46)	2.12	0.56

slack variables	sample 3: 1997Q1-2017Q4 Sum of AR coefficients	OutT	InT	OuT Dummy	Constant	R^2
ST_UR	0.38	0.48 (0.14)	0.49 (0.16)	0.30 (0.37)	3.28 (0.37)	0.39
LF_labour	0.22	0.23 (0.07)	0.31 (0.12)	0.08 (0.42)	0.06 (0.42)	0.39
OutputGap_CD	0.22	0.27 (0.07)	0.41 (0.35)	-0.31 (0.38)	1.89 (0.5)	0.41
OutputGap_CES	0.21	0.28 (0.07)	0.42 (0.36)	-0.33 (0.37)	1.91 (0.37)	0.42

TABLE A.6. Main estimation results for threshold models

Notes: figures between brackets refer to the standard deviation of the corresponding coefficients. The acronyms for slack variables stand for (in the order they are presented) the short-term unemployment rate, the survey question related to labour as a restrictive factor to production, the Cobb-Douglas output gap and the CES output gap. OutT and IntT stand for the slack variable values outside and inside the thresholds, respectively. OutT Dummy stands for the dummy variable that has an unit value outside the thresholds.

	Slack variable	OutputGap CD	OutputGap CES	OutputGap CD (outT)	OutputGap CES (outT)	LF_labour	ST_UR	AR Model	Eurosystem projections	Random Walk
Steps ahead	1	0.8	0.8	0.8	0.8	0.9	0.9	1.0	0.5	1.0
	2	0.8	0.8	0.8	0.8	0.9	1.0	1.0	0.7	1.0
	3	0.7	0.7	0.7	0.7	0.9	1.0	1.2	1.0	1.0
	4	0.7	0.7	0.7	0.8	0.9	1.1	1.1	0.9	1.0
	5	0.7	0.7	0.8	0.8	0.9	1.0	1.1	1.0	1.0
	6	0.8	0.8	0.9	0.9	0.8	1.1	1.1	0.7	1.0
	7	0.8	0.8	0.9	0.9	0.8	1.1	1.1	0.6	1.0
	8	0.9	1.0	1.1	1.1	0.8	1.1	1.1	0.7	1.0

TABLE A.7. Relative RMSE - Benchmark Random Walk

Notes: shaded values stand for statistically significant differences between the forecasts according to the Diebold Mariano test. The acronyms ST_UR and LF_labour stand for, respectively, the short-term unemployment rate the survey question related to labour as a restrictive factor to production. The "outT" models refer to threshold models.

	Slack variable	OutputGap CD	OutputGap CES	OutputGap CD (outT)	OutputGap CES (outT)	LF_labour	ST_UR	AR Model	Eurosystem projections	Random Walk
Steps ahead	1	1.6	1.6	1.6	1.6	1.8	1.7	2.0	1.0	2.0
	2	1.1	1.1	1.1	1.1	1.3	1.4	1.5	1.0	1.4
	3	0.7	0.7	0.7	0.7	0.9	1.0	1.2	1.0	1.0
	4	0.8	0.8	0.8	0.8	1.0	1.2	1.2	1.0	1.1
	5	0.7	0.7	0.8	0.8	0.8	1.0	1.1	1.0	1.0
	6	0.7	0.7	0.8	0.8	0.8	1.3	1.5	1.0	1.2
	7	0.8	0.7	0.8	0.8	0.7	1.2	1.5	1.0	1.2
	8	0.8	0.8	0.9	0.9	0.7	1.3	1.5	1.0	1.2

TABLE A.8. Relative RMSE - Benchmark Eurosystem projections

Notes: shaded values stand for statistically significant differences between the forecasts according to the Diebold Mariano test. The acronyms ST_UR and LF_labour stand for, respectively, the short-term unemployment rate the survey question related to labour as a restrictive factor to production. The "outT" models refer to threshold models.

References

Albuquerque, Bruno and Ursel Baumann (2017). "Will US inflation awake from the dead? The role of slack and non-linearities in the Phillips curve." *Journal of Policy Modeling*, 39(2), 247–271.

Banco de Portugal (2017). "Potential output: challenges and uncertainties." *Economic Bulletin*.

Ciccarelli, Matteo and Chiara Osbat (2017). "Low inflation in the euro area: Causes and consequences." Occasional Paper Series 181, European Central Bank.

Constâncio, Vítor (2015). "Understanding Inflation Dynamics and Monetary Policy: Panel remarks at the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming." https://www.ecb.europa.eu/press/key/date/2015/html/sp150829.en.html.

Diebold, Francis X and Roberto S Mariano (1995). "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics*, 13(3), 253–263.

Dotsey, Michael, Shigeru Fujita, and Tom Stark (2017). "Do Phillips Curves Conditionally Help to Forecast Inflation?" Working Papers 17-26, Federal Reserve Bank of Philadelphia.

Draghi, Mario (2017). "Accompanying the economic recovery: speech at the ECB Forum on Central Banking." https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp170627.en.html.

ECB (2014). "The Phillips Curve Relationship in the Euro Area." *ECB Monthly Bulletin*, 45(July), 99–114.

Harvey, David, Stephen Leybourne, and Paul Newbold (1997). "Testing the equality of prediction mean squared errors." *International Journal of Forecasting*, 13(2), 281–291.

Macklem, Tiff (1997). "Capacity constraints, price adjustment, and monetary policy." *Bank of Canada Review*, 1997(Spring), 39–56.

Stock, James H. and Mark W. Watson (2010). "Modeling Inflation After the Crisis." NBER Working Papers 16488, National Bureau of Economic Research, Inc.

Teles, Pedro and Joana Garcia (2016). "How can the Phillips curve be used for today's policy?" *Economic Studies*, 2(2).

Yellen, Janet (2013). "Labor Market Dynamics and Monetary Policy: a speech at the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming ."