ufficio parlamentare di bilancio

upB. ufficio parlamentare di bilancio

WORKING PAPER

Which gap? Alternative estimations of potential output and the output gap in the Italian economy

by Cecilia Frale and Sergio De Nardis

Working Paper no. 2 / *July 2017*

The Working Papers series comprises technical and institutional studies conducted by the PBO, with the collaboration of independent experts where appropriate. The Working Papers seek to explore in greater detail specific issues relevant to the institutional work of the PBO.

The contents of the Working Papers represent the views of their authors and do not necessarily reflect those of the PBO or the institutions with which the independent experts are affiliated.

The electronic version can be downloaded from: www.upbilancio.it

Parliamentary Budget Office
Via del Seminario, 76
00186 Rome, Italy
segreteria@upbilancio.it

Which gap?
Alternative estimations of potential output and the output gap in the Italian economy

by Cecilia Frale and Sergio De Nardis[1]

Abstract

This paper presents an estimation of the output gap based on a multivariate filter that extracts the signal from output indicators (GDP, unemployment rate and capacity utilisation) and from a Phillips curve. More specifically, we present a number of unobserved components models that differ in their information sets and the specification of the trend, which differs in its fit with the observed data. The empirical findings demonstrate that as the specification of the model changes the estimation of the output gap and potential output vary over a broad interval. The various models, which are assessed using multiple measures (statistical fit, extent of revision of the estimates and inflation-predicting capacity), are optimal only in relation to a specific measure or a time period, and none of them is better in absolute terms. In addition, models that appeared optimal prior to the recent economic crisis appear to have lost their explanatory power in recent years. The policy implications of these findings point to the need to base economic policy decisions on multiple models or in any case on a model that is sufficiently flexible to incorporate different economic assumptions (such as anomalous cyclical conditions or hysteresis) and to accompany the estimates obtained with appropriate measures of error.

JEL classification code: C32, C53, E32.

Keywords: Unobserved components models; output gap; Phillips curve; inflation forecasts.

[1] Cecilia Frale, UPB, cecilia.frale@upbilancio.it and Sergio De Nardis, UPB, sergio.denardis@upbilancio.it.
The authors would like to thank Davide Delle Monache and Roberto Golinelli for their suggestions, which much improved the paper. We also thank Michele Lenza, Tommaso Proietti, Carlo Favero, Marco Cacciotti and Federico Giammusso for helpful discussions of the issue of estimating the output gap.

Introduction

The concepts of potential output and output gap (OG, i.e. the percentage deviation of actual output from potential output) have in recent years taken on considerable importance within the framework of EU rules. The definition of these variables determines the assessment of the cyclical position of the economy and the implications for the fiscal surveillance process within the scope of the Stability and Growth Pact. Nevertheless, these variables are not observable, and their estimation is characterised by considerable uncertainty, with their values differing significantly depending on the approach adopted. For example, that used by the European Commission, which was agreed with the Member States, provides measures of potential output and the OG that have diverged substantially from those produced by other international organisations (OECD and IMF), especially in recent years.

The approach adopted by the European Commission is a production function method, for which potential GDP depends on a combination of the factors of production (labour and capital) and total factor productivity (TFP), expressed at their trend level. The trend components of the factors of production are obtained with the use of statistical filters (univariate filters such as Hodrick and Prescott or bivariate filters such as Kalman). More specifically, in estimating the NAWRU (Non-Accelerating Wage Rate of Unemployment) included in the equation of potential labour supply, a Kalman filter is applied to a bivariate model that includes the unemployment rate time series and the Phillips curve. The trend for TFP is extracted with a Bayesian Kalman filter applied to a bivariate model that includes the so-called Solow residual and an indicator of capacity utilization.2

The OECD also uses a production function method, although it differs somewhat from that adopted by the European Commission. For example, the two methods differ considerably in the concept of structural unemployment, which for the OECD corresponds to the equilibrium unemployment rate consistent with stable inflation (NAIRU) and is equal to the official target of the monetary authorities, therefore incorporating inflation expectations.[3]

Finally, the IMF uses different methods to estimate the output gap depending on the country involved, using both multivariate filters (e.g. GDP, the unemployment rate and consumer prices) and the production function.[4]

Underscoring the uncertainty associated with these measures, there is considerable variability not only among the various approaches but also within the same method depending on the vintage of the estimates, as the latter have been revised over time,

[2] For more details, see Karel Havik et al. (2014), "The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps", European Economy, Economic Papers n. 535.
[3] For more details, see Ollivaud P. and Turner D., (2014) "The Effect of the Global Financial Crisis on OECD Potential Output" OECD Economics Department Working Papers No. 1166.
[4] For more details, see Alichi A., Bizimana O., Domit S., Fernandez Corugedo E., Laxton D., Tanyeri K.,Wang H., and Fan Zhang F. "Multivariate Filter Estimation of Potential Output for the Euro Area and the United States", IMF working paper, WP/15/253.

sometimes substantially (this is the case with the OECD and EC estimations in particular; Table1).

The recent economic crisis appears to have made the determination of these variables even more complex. The long recession may have permanently damaged the potential productive capacity of economies, thereby increasing the difficulty of distinguishing between the temporary (cyclical) component and the trend (structural) component of GDP growth.[5] On the other hand, it is also possible that methods for extracting the output gap based on the standard economic conditions of the past, characterised by standard cyclical frequencies, might not be appropriate for the anomalous cyclical conditions (caused by repeated or more persistent cyclical shocks than historical standards) seen in recent years. If we wish to take account of the greater variation of the cycle around the trend that characterised the last recession, it would be necessary to identify a more stable and less pro-cyclical estimation of potential output.[6] ECB researchers recently demonstrated that using different functional forms for the trend from those commonly adopted in the past, compatible with a larger cyclical component than in previous examples, we obtain estimates of potential output for the euro area which diverge substantially from those proposed by international institutions and which have a better capacity for predicting inflation in the crisis years.[7]

With regard to other issues associated with the increased difficulty of identifying cyclical/trend components, we must also bear in mind the fact that the variables of the economic surveys based on the assessment of economic agents, used for example in the Commission method for correcting the cycle for developments in total factor productivity, may have become less informative in the last few years, as the recent crisis may have changed the assessments and expectations of economic agents about what is considered a normal level of economic activity.[8]

---

[5] On the effects of hysteresis (recession that reduces the level of potential GDP) and super-hysteresis (recession that reduces the growth of potential GDP), see Blanchard O., Cerutti E. and Summers L. (2015), "Inflation and Activity: Two Explanations and their Monetary Policy Implications", IMF Working Paper, no. 15, November, and Fatàs A. and Summers L. (2015), "The Permanent Effect of Fiscal Consolidations", CEPR Discussion Paper Series no. 10902, October. For Italy, instances of hysteresis in the GDP series are identified by Proietti, T. (2002), "Some Reflections on Trend-Cycle Decompositions with Correlated Components", EUI working paper ECO no. 2002/23. Later estimations for Italy that adopt the same approach have underscored the intensification of the phenomenon in recent years. See for example, Parliamentary Budget Office (2016) "2016 Budgetary Planning Report", April, pages 28-33.

[6] For more on this issue, see Ledvai J., Salto M. and Thum-Thysen A. (2015) "Structural unemployment vs. NAWRU: Implications for the assessment of the cyclical position and the fiscal stance", European Commission, Economic Papers 552/June. Considerable doubts about the reliability of the pro-cyclical evolution of estimates of potential output, posited by various institutions for the advanced countries, are raised in Coibion O., Gorodnichenko Y. and Ulate M. (2017) "The cyclical sensitivity in estimates of potential output", NBER Working Paper Series, no. 23580, July.

[7] Jarocinski M. and Lenza M. (2016), "An inflation-predicting measure of the output gap in the euro area", *ECB working paper series* no. 1966, September.

[8] See European Commission (2016) "New normal? The impact of the financial crisis on business and consumer survey data" in European Business Cycle Indicators-3rd quarter 2016, Technical paper 011, October; Bruno G., Crosilla L., and Margani P. "Inspecting the relationship between business confidence and industrial production: evidence based on Italian survey data", paper presented at the 2016 CIRET

Table 1 – Estimations of the output gap by selected international institutions in successive vintages

	OECD			IMF			EC		
	Jun 17	Nov 16	Jun 16	Apr 17	Oct 16	Apr 16	May 17	Nov 16	May 16
2002	1.2	1.1	1.0	-0.4	-0.4	-0.4	1.1	1.1	0.5
2003	0.3	0.2	0.1	-1.1	-1.1	-1.1	0.2	0.1	-0.4
2004	0.5	0.5	0.4	-0.3	-0.3	-0.3	0.5	0.5	0.4
2005	0.8	0.7	0.7	0.0	0.0	0.0	0.6	0.6	0.7
2006	2.0	2.0	2.0	1.5	1.5	1.5	1.8	1.8	2.4
2007	2.7	2.7	2.8	2.7	2.7	2.7	2.4	2.4	3.9
2008	1.2	1.2	1.3	1.8	1.8	1.8	1.2	1.1	3.0
2009	-4.4	-4.5	-4.2	-3.2	-3.2	-3.2	-4.0	-4.0	-2.3
2010	-2.7	-2.9	-2.4	-1.3	-1.3	-1.3	-2.0	-2.0	-0.2
2011	-1.9	-2.3	-1.5	-0.5	-0.5	-0.5	-1.3	-1.4	0.9
2012	-4.4	-4.9	-3.9	-2.8	-2.8	-2.8	-2.9	-2.9	-1.4
2013	-5.8	-6.3	-5.0	-4.1	-4.1	-4.1	-4.2	-4.1	-2.8
2014	-5.3	-5.9	-4.8	-4.1	-4.1	-4.1	-3.8	-3.7	-2.8
2015	-4.5	-5.1	-3.7	-3.3	-3.3	-3.3	-2.8	-2.6	-1.9
2016	-3.4	-4.2	-2.4	-2.4	-2.5	-2.5	-1.7	-1.6	-0.8

In order to take account of these elements of uncertainty in the Italian case within a single framework, in this study we estimate an unobserved components model of the type proposed by the ECB economists,[9] in which the output gap is represented by a cyclical component common to a set of economic activity variables (GDP, unemployment rate and capacity utilisation). Meanwhile, specific trends identify the structural component of each of those variables. The model is completed by a Phillips curve in which inflation evolves as a function of the output gap, a trend component of consumer price developments and the oil price.

There is an extensive literature on estimating the output gap using unobserved components models, beginning with bivariate GDP and unemployment or GDP and inflation models, supplemented with assessments of the effects of monetary policy by including the real interest rate in the equation for the output gap. Further developments saw the construction of trivariate models (GDP, unemployment and inflation or, alternatively, capacity utilisation). More recently, a Bayesian version of the bivariate

conference, https://www.ciret.org/media/ciret_papers/copen25; Malgarini M., (2011) "Industrial production and Confidence after the crisis: what's going on?", CIRET/KOF/HSE Workshop on National Business Cycles in the Global World, Moscow, September 16-17; Fantacone S., Garalova-Stieg P. and Malgarini M., "Is business confidence still a good indicator for industrial production? Evidence from the EC survey", paper presented at the Confindustria CSC seminar on 12 May, available at: http://www.confindustria.it/wps/wcm/connect/www.confindustria.it5266/d6d2c5f8-dff3-48d6-9d0a-7af4326088d4/Fantacone-Garalova-Malgarini.pdf?MOD=AJPERES&CONVERT_TO=url&CACHEID=d6d2c5f8-dff3-48d6-9d0a-7af4326088d4.

[9] Jarocinski M. and Lenza M. (2016), op. cit.

GDP and unemployment model has been proposed.[10] The unobserved components approach has a number of advantages over the more complex production function estimation method adopted by the European Commission and the OECD.[11] Firstly, the resulting estimate is compatible with Okun's conception of potential output,[12] i.e. the maximum level of output that an economy can sustain without generating inflationary pressure. In addition, it is relatively simple to interpret because it is based on a limited number of variables. Finally, given that the estimate cannot diverge too much from the observed data, it is especially appropriate for capturing shocks with permanent impacts on the economy that reduce potential output, as in the event of severe contractions in output. The main limitation is that it is a statistical filter, and as such is affected by end-point-bias[13] and does not offer an immediate economic interpretation.

To take account of the factors of uncertainty noted earlier (breakdown into cycle/trend and informative capacity of the cyclical variables in surveys), various output gap estimation models are considered in this paper. These vary in the different types of trend involved, whose developments reflect the strength of the influence of actual GDP. More specifically, we adopt two specifications in which trend growth slows in recent years as actual GDP deteriorates, indicating relatively larger hysteresis effects, and one specification in which the trend is more stable over time, displaying a smaller response to the effects imparted by the cycle. Secondly, for each approach to specifying the trend, we estimate different versions of the model that are differentiated by the inclusion or absence of the capacity utilization rate variable drawn from economic surveys.

[10] On the bivariate models, see Kuttner K.N. (1994) "Estimating potential output as latent variable", Journal of Business and Economic Statistics 12: 361-368. For more on the effects of monetary policy on those models, see Gerlash S. and Smets F. (1999), "Output gaps and monetary policy in the EMU area", European Economic Review 43: 801-812. For more on the development of trivariate models, see Apel M. and Jansonn P. (1999)."System estimates of potential output and the NAWRU", Empirical Economics 24: 373-388. The Bayesian estimation in developed by Planas, C., Rossi, A. and Fiorentini, G. (2008), "Bayesian analysis of the output gap", Journal of Business and Economic Statistics, 26(1): 18-32. This latter approach inspires Jarocinski M. and Lenza M. (2016). On the use of capacity utilisation in estimating the output gap, see Graff M. and Sturm J-E, (2010) "The Information Content of Capacity Utilisation Rates for Output Gap Estimates", CESIFO Working Paper no. 3276, December.

[11] For a discussion of the methods and estimation problems of the OECD and the European Commission, see Fioramanti M., Padrini F. and Pollastri C. "La stima del PIL potenziale e dell'output gap: analisi di alcune criticità", Nota di lavoro UPB, 1/2015. For a survey of the different methods, with applications to the Italian case, see Bassanetti A., Caivano M. and Locarno A., (2010) "Modelling Italian potential output and the output gap", Banca d'Italia Working Paper, No. 771.

[12] Okun (1962), "Potential GNP: Its Measurement and Significance," in Proceedings of the Business and Economic Statistics Section, pp. 98-104 (Washington: American Statistical Association).

[13] This is the situation in which statistical filters are especially influenced by the observations at the end of the sample period and are therefore subject to revision when new observations become available. However, even methods based on the production functions use statistical filters in dealing with the variables and are therefore not immune from this problem.

Very briefly, the results of the analysis can be summarised as follows:

- Alternative forms – in the specification of the trend and in the information set - of the estimation model produce a broad range of values for the output gap, confirming the considerable uncertainty characterising this variable.
- The uncertainty is essentially not eliminated by adopting a criterion for assessing the goodness of the various models. Purely statistical criteria (maximum likelihood or stability of parameters) lead to the selection of different models from those that would be chosen using a more economic criterion based on the goodness of their capacity to forecast inflation.
- However, any choice based on a given criterion may still not be stable over time. Models that appeared optimal prior to the recent economic crisis seem to have lost their explanatory power in recent years. In 2012-2016, the model with the best forecasting capacity is that with a less pro-cyclical trend and, therefore, a larger output gap. This superiority does not hold up, however, if forecasting ability over a longer time period is considered. This seems to represent a signal of the differences in the characteristics of the cycle in more recent years compared with the standard behaviour found in the past.
- The evidence for the predictive capacity of the various models for inflation points to a deterioration in 2012-2016 in the signal provided by capacity utilisation. This appears to be yet another indication of the anomalous behaviour of the last economic cycle.

Given the considerable uncertainty, the policy implication of this evidence would be to base economic policy decisions on multiple models rather than a single estimation of the trend and the output gap, or in any event specify a model that is sufficiently flexible to incorporate different economic assumptions (such as anomalous cycles or hysteresis) and to accompany the estimates obtained with appropriate measures of error.

The paper is organised as follows. Following a description of the model in section 1, we present in section 2 the results of the estimation for Italy. Section 3 is devoted to assessing the goodness of the alternative models, while a brief assessment concludes.

1. The econometric model

The model used to estimate potential output and the output gap belongs to the class of unobserved components models and envisages the breakdown of a number of output variables (GDP, the unemployment rate (UR) and capacity utilization (CP)) into a cyclical component and a structural component. The model is then completed by an equation linking inflation to the output gap and to exogenous cost-push factors (Phillips curve).

The model is formalised as follows:

$y_t{}^n = \mu^n{}_t + b^n(L)\varphi_t + \varepsilon_t^n$ (output) n=1,..N $\varepsilon_t{}^n \sim NID(0, \sigma^2_{\varepsilon^n})$ (1)

$\pi_t = \mu^\pi{}_t + a(L)\varphi_t + \gamma(L)X_t + \varepsilon_t^\pi$ (Phillips curve) $\varepsilon_t{}^\pi \sim NID(0, \sigma^2_{\varepsilon^\pi})$

where

$y_t{}^n$ are the N output variables (GDP, unemployment rate, productive capacity utilisation rate);

$b^n(L)$ and $a(L)$ are the polynomials of the lag operator; ε_t^n and ε_t^π are Gaussian and uncorrelated errors.

π_t is the rate of core inflation (i.e. excluding energy products and unprocessed food), whose fluctuations around the trend ($\mu^\pi{}_t$) depend on developments in the output gap and the price of oil (X_t), considered as a cost-push variable, with contemporaneous and lagged effects expressed through the $\gamma(L)$ coefficients in the lag operator.

The N output variables ($y_t{}^n$) are therefore derived from the sum of a common cycle (φ_t) defined by a second-order autoregressive process:

$$\varphi_t = \theta_1\varphi_{t-1} + \theta_2\varphi_{t-2} + k_t \qquad k_t \sim NID(0, \sigma^2_k)$$

and specific trends ($\mu_t{}^n$) such that:

$$\mu_t{}^n = \mu_{t-1}{}^n + \beta^n{}_t + \eta^n{}_t \qquad \omega_t \sim NID(0, \sigma^2_{\xi^n})$$

in which the errors k_t and $\eta^n{}_t$ are assumed to be Gaussian and uncorrelated.

For GDP, the following restrictions are introduced: $b^1(L)$ =1 and $\sigma^2_{\varepsilon^1} = 0$ and equation (1) becomes:

$$y_t^1 = \mu_t^1 + \varphi_t \qquad \text{(1a)}$$

such that φ_t corresponds to the divergence of output from its trend, thereby representing the output gap.

To account for the possibility of more or less flexible trends, we consider the following alternative specifications of μ_t that are characterised by a higher/lower closeness to the observed data:

1. Local level model (LLT): with level $\mu_t^1 = \mu_{t-1}^1 + \beta^1{}_t + \eta_t$ and drift (which gives the slope of the trend) $\beta^1{}_t = \beta^1{}_{t-1} + \xi_t$ varying over time and uncorrelated errors $\eta_t \sim NID(0, \sigma^2_\eta)$ and $\xi_t \sim NID(0, \sigma^2_\xi)$.
2. Random walk plus drift (RW): obtained by setting, in the LLT model equation, $\sigma^2_\xi = 0$; with constant drift β.

3. Integrated random walk (IRW): obtained by setting, in the LLT model equation, $\sigma_\eta^2 = 0$; with variable drift β_t.

As noted, the different specifications of the trend represent situations of higher/lower closeness to the data . More specifically, the IRW model produces a smoother trend, mainly influenced by the maximum and minimum points of the series, while the RW trend displays a closer fit with the data observed in the various sub-periods.

The trends adopted for the other output variables and inflation follow a RW process.

Finally, a cost-push variable contributing to the determination of inflation is the price of oil (X_t) with contemporaneous and lagged effects expressed through the $\gamma(L)$ coefficients in the lag operator.

2. Empirical results

Model (1) was applied to Italian data for 1985Q1 -2016Q4.[14] The GDP series was drawn from seasonally adjusted quarterly Istat accounts at constant prices. The inflation estimate regarded core inflation, i.e. consumer prices excluding the prices of unprocessed food and energy products drawn from Istat.[15] The series for capacity utilisation regards the degree of plant utilisation as measured in Istat surveys. The oil price is the World Bank spot crude prices in euros.

The model is cast in state-space form and was estimated using the maximum likelihood method with a Kalman filter and its associated smoothing algorithm.[16] The results, which are reported in Table 2, show that the parameters of the output gap, θ_1 and θ_2, are significant in all the specifications and are relatively stable, with the exception of IRW, which shows the greatest divergence from the other models. There is greater variability for the parameters associated with the output gap in the other two equations. The inflation equation also includes the contemporaneous and lagged oil price, the coefficients for which appear significant in the models that do not include capacity utilisation.

[14] The models that incorporate the series for capacity utilisation were estimated as from 1986Q1, when that data became available.

[15] The GDP series prior to 1996 was constructed using NACE Rev 1 quarterly national accounts data in 2010 prices and linked to NAVE Rev. 2 national accounts. The consumer price index net of energy products prior to 1996 was calculated using OECD Main Economic Indicator data. The series enter the model through transformations: the GDP series is included in the model in logs, while the CPI enters in log differences, the unemployment rate through a logit transformation and capacity utilisation in levels. For the latter variable, Dickey-Fuller unit root tests are rejected at a 95 per cent confidence level.

[16] For a complete discussion of the estimation method, see Harvey 1989, "Forecasting structural time series and the Kalman filter", Cambridge University press, Cambridge. The estimation used the SsfPack 2.2 library provided by Koopman SJ, Shepard N and Doornik JA (1999), "Statistical algorithms for models in state space using SsfPack22", Econometric Journal 2 113-166, of the Ox Metrics 6.0 software.

The results in terms of estimated components, which are reported in Table 3 and the subsequent charts, show that with changes in the specification of the model, the estimated OG and potential output vary over a very broad range.[17] The greatest differences are found during expansions and, above all, in the recent crisis at the end of the sample period. More specifically, in the last period the most flexible models, IRW and LLT, generate a lower estimate of potential output and, therefore, an output gap that is closed more quickly with the recovery in output than the RW model, which assumes more stable potential growth during the crisis (Figure 1). All of the models identify the double decline in the output gap in the recent recession, which is larger in 2012 than in 2008, but within a very broad estimated range, which in 2016 for example goes from -0.5 (IRW) to -5.3 (RW-CP). For the years in the forecasting period,[18] the negative output gap narrows at different paces in the different models. In the IRW model, GDP returns to its potential in 2017.[19] These results essentially reflect the different hypotheses for the slope of the trend in the various random-walk processes adopted. In the RW model, the assumption of constant drift makes the process dynamics (i.e. the slope of the trend) relatively stable over time (the shocks in that model are only on the level of the trend) and therefore less pro-cyclical and less exposed to hysteresis effects (or more precisely, super-hysteresis, in the sense of a persistent decrease in the growth rate of potential output in a recession). The opposite is found in the LLT and IRW, which the variability of drift over time tends to produce changes in the slope of the trend (i.e. the growth rate of potential output) which may persist in prolonged recessions (this is especially the case for the IRW model, in which the shocks exclusively affected the slope of the trend).

[17] For comparative purposes, an LLT model without the price of oil was also estimated, although it did not produce relevant differences with the base method and has therefore been omitted for brevity.

[18] The values for the variables are obtained on the basis of the parameters estimated with the application of the Kalman filter for future years.

[19] In order to compare these results with the available estimates of the Italian OG (shown in Table 1), note that the official estimates of the European Commission are closest to the result produced by the IRW model, indicating the elimination of the OG in 2018. The series estimated by the OECD shows the output gap closing more slowly and is closest to that generated by the RW-CP model. Finally, the estimate of the IMF nears that of the LLT-CP model from 2012 onwards.

Table 2 – Estimated parameters and their significance in the different specifications of the model (1)

		LLT	IRW	RW	LLT-CP	IRW-CP	RW-CP
OG	θ_1	1.77 ***	1.44 ***	1.81 ***	1.74 ***	1.81 ***	1.79 ***
	θ_2	-0.79 ***	-0.52 ***	-0.82 ***	-0.76 ***	-0.82 ***	-0.81 ***
	σ_η^2	2.75 ***		3.31 ***	2.36		2.58
trend GDP	σ_ξ^2	0.06 **	0.03 ***		0.02 ***	5.31	
trend UR	b_1^2	-3.60 ***	-0.93 ***	-2.98 ***	-1.93 **	-6.91 ***	-1.60 ***
	b_2^2	-2.17 *	-1.46 ***	-2.17 ***	-2.64 ***	-1.18	-2.35 ***
	$\sigma_{\varepsilon^2}^2$	57.84 ***	91.25 ***	63.72 ***	69.66 ***	58.52 ***	73.64 ***
trend infl	a_1	0.75 **	0.11 *	0.66	0.38 **	0.91 ***	0.35
	a_2	-0.68 **	-0.02	-0.61	-0.31 *	-0.86 ***	-0.29
	$\sigma_{\varepsilon^\pi}^2$	1.33	1.68 ***	1.42	1.46 *	1.40 ***	1.47
	γ_0	0.005 ***	0.006 ***	0.005 ***	0.001 *	0.001	0.001
	γ_1	0.003 *	0.003 *	0.003 *	0.001	0.001	0.001
trend CP	b_1^3				0.26 ***	0.45 **	0.25 ***
	b_2^3				-0.15 **	-0.41 **	-0.16 **
	$\sigma_{\varepsilon^3}^2$				0.11	0.13 *	0.11 ***
	logLik	1252	1239	1242	1865	1839	1859
	BIC	-2446	-2424	-2431	-3657	-3610	-3650
	Akaike	-2480	-2455	-2463	-3700	-3650	-3689

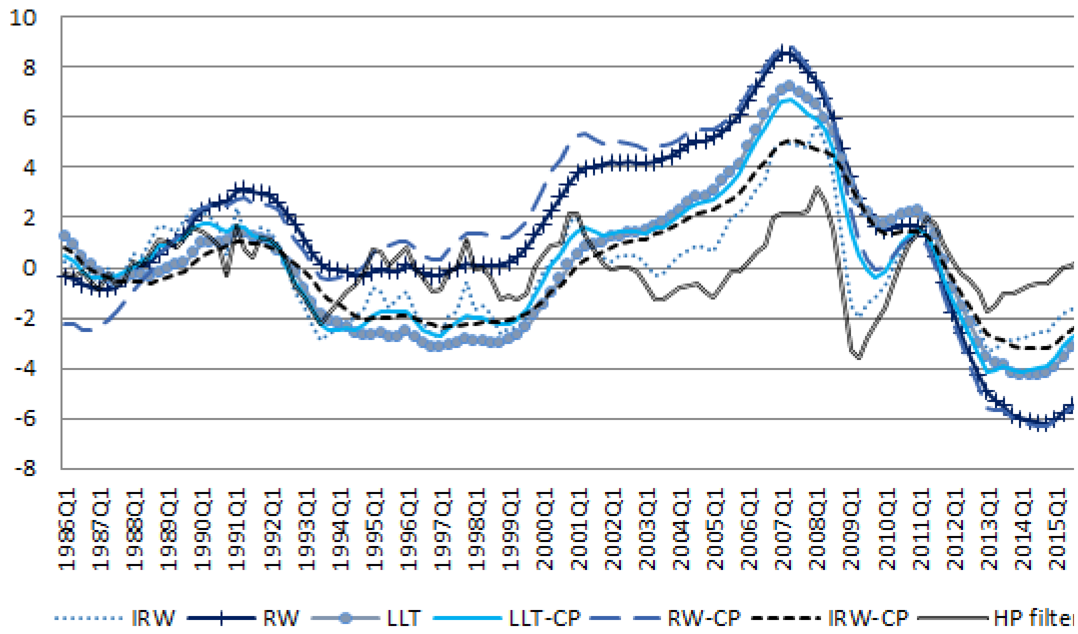
LogLik is likelihood (in log form). A comparison is only possible between nested models and therefore in the two groups with and without capacity utilisation. BIC (Schwarz) and AIC (Akaike) are criteria for the selection of a model from among a class of models with a different number of parameters. (***) indicates a significance of 99 per cent for the parameter, (**) indicates 95 per cent and (*) indicates 90 per cent. The variances are multiplied by 10^5.

Table 3 – Output gap in different specifications of the estimation model

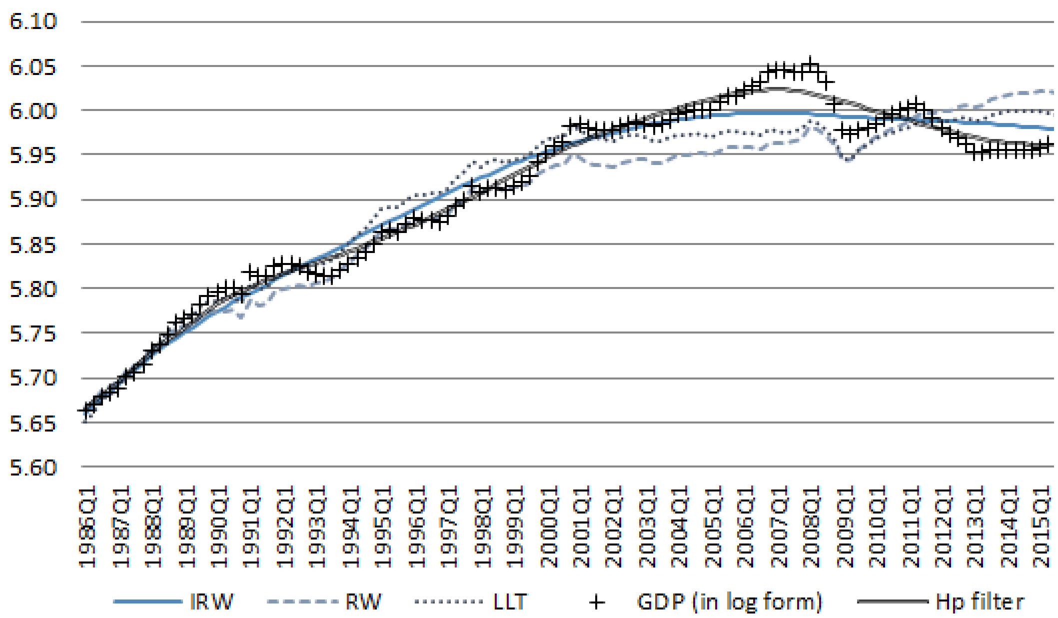
	LLT	RW	IRW	LLT-CP	RW-CP	IRW-CP
2010	2.0	1.6	0.3	0.6	0.6	1.4
2011	1.5	0.7	1.2	0.8	0.3	0.9
2012	-1.9	-3.0	-1.6	-2.4	-3.4	-1.3
2013	-3.9	-5.4	-3.1	-4.0	-5.7	-2.9
2014	-4.2	-6.1	-2.7	-4.0	-6.2	-3.2
2015	-3.4	-5.7	-1.8	-3.0	-5.7	-2.6
2016	-2.7	-5.2	-0.5	-2.2	-5.3	-2.0
2017	-2.4	-5.0	0.1	-1.7	-5.0	-1.7
2018	-1.9	-4.3	0.1	-1.1	-4.1	-1.3

Figure 1 – Output gap and potential output based on different specifications of the estimation model (1)

Output gap



Potential output



(1) The benchmark in the charts is the decomposition of cycle and trend as the result of the application of the Hodrick and Prescott filter on the GDP series.

3. Assessment of the alternative models

In view of the considerable variety in the results, it is necessary to adopt an assessment criterion to choose a model.[20] Three different criteria are considered below: 1) goodness of fit with the data; 2) the stability of the estimated parameters; 3) the ability to identify the phase of the cycle on the basis of the capacity to predict inflation. However, the various criteria do not provide unequivocal answers, leading to the selection of different functional forms.

In terms of likelihood (the first criterion), the model best supported by the data is LLT, with or without capacity utilisation (Table 1), even considering the BIC and AIC criteria.[21] If, on the other hand, we base the criterion on the standard error of the OG, as obtained with the smoothing algorithm, the best model is that with the integrated trend (IRW, Figure 2a) even though in this case the error around the specific estimate increases at the end of the sample period. The possibility of constructing confidence intervals around the specific value of the output gap is especially useful in these circumstances, considering the difficulty of measuring the phenomenon. The estimation of an interval rather than a single value could be more helpful for economic policy decision-making.

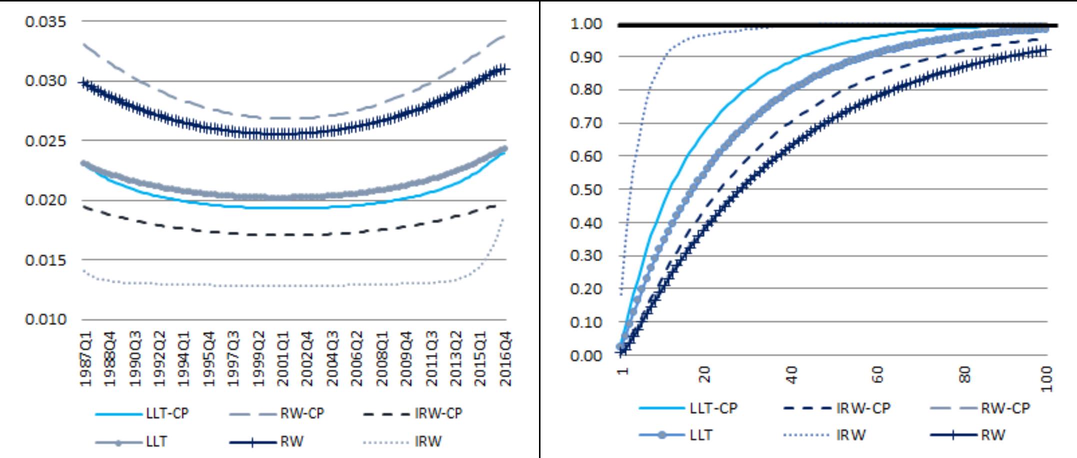
An additional criterion for assessing models is the size of revision of estimates (second criterion), measured in terms of the number of observations necessary to have stable parameters. In these terms, the best model among those analysed was IRW (Figure 2b). In this case, we have large revisions with the initial observations, but it rapidly approaches the final estimate and after only ten or so quarters error is reduced by 90 per cent (however, only in the case in which capacity utilisation is not included). Among the remaining models, LLT-CP converges most rapidly, with estimation error falling by 90 per cent after about 40 observations (10 years), while almost twice that number are needed for RW and RW-CP. The LLT-CP model appears to be the best in statistical terms (both for LogLik and in terms of the variability of the estimates).

[20] With regard to the selection criteria issue, note that method of the European Commission, while not incorporating a metric for assessing the result of the model, does give preference to the objective of minimizing revisions of estimates in the discretionary choice of the initialisation parameters (of the NAWRU and TFP). A tool was recently introduced to assess the plausibility of results (the plausibility tool). It produces a band of acceptability to identify anomalies in the estimation of the output gap. For more details, see Hristov A., Raciborski R. and Vandermeulen V., (2017)"Assessment of the Plausibility of the Output Gap Estimates", European Economy Economic Briefs, April. In this paper, we adopt a selection procedure similar to that used by Jarocinski and Lenza (2016). These authors choose between multiple models with different specifications of the trend, comparing them on the basis of likelihood and inflation predicting capacity.

[21] The BIC – Bayesian information criterion or Schwarz criterion and the AIC – Akaike Criterion assess the likelihood of a model by introducing a penalty term for the number of parameters estimated, thereby preventing overfitting of the model. They are as follows: BIC=-2ln(L) + k ln(n) and AIC=2k-2ln(L) in which ln(L) is the log-likelihood, k is the number of parameters and n is the number of observations.

Figure 2 – Standard error of the output gap and revision of estimates[22]

A – Standard error	*B – Percentage reduction of error after n periods*



We obtain substantially different results when we consider the third criterion, i.e. the capacity of the various models to predict inflation (Table 4). A validation exercise was conducted based on a recursive estimation of the models for the sample period 2002Q2-2016Q4. More specifically, each model was estimated over a moving time interval that increased by one observation at each step as from 2002Q2. At each step, annualised inflation was forecast up to four quarters ahead and the predicted value was compared with the observed data. We used the mean square error (MSE) as a metric for error.[23]

In addition, the various models were compared with a simple benchmark, represented by a random walk model for levels of the consumer price series, commonly used as a naive model in forecasting inflation.[24]

[22] The revision of the estimate is calculated using a fixed-point smoother, comparing the real-time variance with the steady-state variance obtained considering an infinite sample. See Anderson, B. D.O. and Moore J. B. (1979), "Optimal Filtering" (Prentice-Hall: Englewood Cliffs, NJ) and de Jong, P. (1989), "Smoothing and Interpolation with the State Space Model", Journal of the American Statistical Association, 84, 1085-1088.

[23] Analogous results are obtained using mean absolute error as the error measure. These have been omitted for brevity.

[24] For more on the issues associated with forecasting inflation, see Stock and Watson (2007),"Why has U.S. inflation become harder to forecast?", Journal of Money, Credit and Banking, vol. 39, February.

Table 4 – Annualized inflation forecasting error (MSE) and Diebold-Mariano test h quarters ahead for selected samples (1)

		h=1	h=2	h=3	h=4	Average
LLT	2002-2016	0.09	0.24	0.44	0.64	0.36
	DM test	*0.08*	*0.09*	*0.04*	*0.03*	
	2002-2007	**0.05**	0.18	0.37	0.62	0.31
	2008-2011	0.14	0.35	0.56	0.61	0.42
	2012-2016	0.09	0.22	0.43	0.71	0.36
RW	2002-2016	0.13	0.36	0.62	0.86	0.49
	DM test	*0.35*	*0.19*	*0.01*	*0.01*	
	2002-2007	**0.05**	0.18	0.37	0.61	0.30
	2008-2011	0.29	0.81	1.35	1.68	1.03
	2012-2016	0.08	<u>**0.17**</u>	**0.28**	**0.44**	**0.24**
IRW	2002-2016	**0.05**	**0.14**	**0.30**	**0.49**	**0.25**
	DM test	*0.07*	*0.07*	*0.05*	*0.04*	
	2002-2007	**0.05**	0.10	0.21	0.35	0.18
	2008-2011	**0.05**	**0.16**	0.31	0.40	0.23
	2012-2016	<u>**0.06**</u>	<u>**0.17**</u>	0.41	0.77	0.35
LLT-CP	2002-2016	0.12	0.23	0.37	0.52	0.31
	DM test	*0.01*	*0.03*	*0.03*	*0.04*	
	2002-2007	**0.05**	**0.09**	**0.16**	**0.27**	**0.14**
	2008-2011	0.07	**0.16**	**0.24**	**0.27**	**0.18**
	2012-2016	0.27	0.48	0.76	1.08	0.65
RW-CP	2002-2016	0.15	0.28	0.45	0.65	0.38
	DM test	*0.00*	*0.01*	*0.01*	*0.01*	
	2002-2007	**0.05**	0.10	0.20	0.33	0.17
	2008-2011	0.11	0.20	0.31	0.36	0.24
	2012-2016	0.33	0.58	0.93	1.35	0.80
IRW-CP	2002-2016	0.23	0.39	0.58	0.76	0.49
	DM test	*0.00*	*0.00*	*0.00*	*0.00*	
	2002-2007	0.11	0.17	0.28	0.43	0.25
	2008-2011	0.28	0.47	0.65	0.74	0.54
	2012-2016	0.34	0.59	0.91	1.24	0.77
Benchmark	2002-2016	0.04	0.10	0.15	0.21	0.13
	2002-2007	0.03	0.07	0.13	0.24	0.12
	2008-2011	0.03	0.08	0.15	0.21	0.12
	2012-2016	0.08	0.17	0.18	0.18	0.15

(1) The table reports the minimum error among models in boldface, excluding the benchmark. The value is underlined if the error is also smaller than the benchmark. The Diebold-Mariano test is a test of significant differences between forecasts and is conducted by comparing each model with the benchmark. Thus, a value equal to or lower than 0.05 indicates that the forecasting errors are significantly different for a confidence level equal to or greater than 95 per cent. Comparisons based on the MSE should be treated with caution when the number of parameters estimated is very different among the models considered.

The goodness of the models in terms of their capacity to forecast inflation is not stable over time, but depends on the period under consideration. In general, inflation forecasting errors are larger in the last period (2012-16), when inflation falls to historically very low levels. Over the entire 2002-2016 sample, the IRW model is the best, on average and for all horizons, and the error is statistically smaller than that found for the other models according to the Diebold-Mariano test[25] (with a maximum error of 7 per cent). However, if the sample is segmented and, more specifically, the recent economic crisis is isolated, the result changes considerably. Until 2011, the LLT-CP model is the best. From 2012 onwards, i.e. during the last recession and the subsequent weak recovery, the model that produces the smallest error (on average and for each forecasting step except the first) is RW, which is that with the most stable and least pro-cyclical trend growth. The latter could represent confirmation of the anomaly that has distinguished recent economic conditions (a much greater cyclical variation than the standard) and the consequent implications for estimating potential output (less pro-cyclical trend growth).

Another factor to consider regards the role of capacity utilisation in forecasting inflation. The LLT-CP model, which uses the Istat survey series on plant utilisation, generates a smaller average error than the other models until 2011, but becomes less accurate in more recent years. This may be related to the fact that the signal from the Istat surveys on capacity utilisation has become less correlated with actual economic developments. This could reflect a number of conditions consequent upon the recent economic crisis. For example, the severity of the contraction in output may have changed the level considered "normal" for businesses, who now have lower expectations for production than in the past. Moreover, over the course of the crisis, firms were exposed to intense selective pressure and the surviving businesses were the most efficient and more likely to make the best use of the plant that they already have.[26] This change of composition means that any assessment of capacity utilisation would not be perfectly comparable with past periods. Finally, it should be noted that during the crisis, many companies postponed the renovation of plant to better times, so that the relatively rapid increase in capacity utilisation observed during the recovery would reflect the consequences of that postponement rather than a cyclical phenomenon.

Conclusions

This paper applies an unobserved components model to estimate potential output and the output gap for Italy in the 1985-2016 sample period, considering a number of specifications of the information set and the functional form for the trend. In short, the

---

[25] Diebold FX. and Mariano RS. (1995), "Comparing predictive accuracy", Journal of Business and Economic Statistics, 13: 253-263.

[26] For more on the change in the composition of the population of firms after the recession, see Linarello A. and Petrella A. "Productivity and reallocation: evidence from the universe of Italian firms". Banca d'Italia (2016), "Questioni di Economia e Finanza (Occasional Papers)" no. 353, September.

results confirm the high level of uncertainty that characterises the estimation of those variables. More specifically:

- alternative forms – in the specification of the trend and in the information set - of the estimation model produce a broad range of values for the output gap;
- the assessment of the goodness of the various models is subject to the selection of the criterion for assessment, which does not, however, lead to the unequivocal identification of one model to prefer over the others. Overall, the LLT model appears to be the best in statistical terms. If we consider the capacity to forecast inflation for the entire 2002-2016 period, the IRW specification is preferable;
- nevertheless, this relative superiority in forecasting inflation changes over time. Considering the more recent period, marked by the double-dip recession, the decrease in inflation to very low levels and the slow economic recovery (2012-2016), the model with the best forecasting performance is the RW specification, which has a less pro-cyclical trend;
- in addition, the evidence of the predictive capacity for inflation of the various models points to a deterioration in the cyclical signal in the 2008-2011 period, in which the forecasting error in all of the models except the benchmark was larger than in all other time periods. Following 2011, the benchmark also registered an error that was substantially larger than its historic average;
- the anomalous behaviour of recent economic conditions could also explain the weakening of the explanatory capacity of changes in capacity utilisation.

These results confirm that the estimation of potential output and the output gap, which is already burdened by uncertainty, has become even more complex after the crisis. The main question mark is clearly the fact that we do not know if and, if yes, by how much the large cyclical contractions in output that have occurred since 2008 have also impacted the trend of the economy (both the rate of growth and the level). The framework used in this paper underscores the consequences of such uncertainty, indirectly providing indications on how we can seek to tackle the problem. A policy body that permits economic policy decisions to depend on unobservable variables affected by considerable uncertainty cannot base its choices on a single estimation. It must have access to a variety of estimations, accompanied by confidence intervals, obtained using different methodological approaches or, alternatively – if a single model is desired – it should be sufficiently flexible to incorporate different assumptions about the nature of the cycle and the trend within its structure.[27] The selection of one of the estimates from among the various alternatives, which as this paper demonstrates may differ considerably, could be based, for example, on the central value of the interval,[28] or may

[27] See, for example, Proietti T., Musso A. and Westermann T. (2007), "Estimating Potential Output and the Output Gap for the Euro Area: a Model-Based Production Function Approach, Empirical Economics", 33, 85-113.
[28] See, for example, Banca d'Italia, "Annual Report 2016", pages 53-54.

be determined (as in this paper) by combining statistical criteria with others linked to the specific features of the position in the cycle, especially in cases in which the length and depth of the negative phase of the cycle are highly uncertain. From this perspective, the plausibility tool[29] adopted by the European Commission to assess the reliability of its estimations of the OG is a step in the right direction towards assessing the uncertainty associated with the estimates but cannot be considered sufficient given the scope of the problem. Using such a tool only determines whether the EC estimation of the OG falls within an acceptable bound, which is not a true confidence interval as it is based on an auxiliary regression model that still takes the trend (and OG) hypothesis of the Commission model as given. And it is precisely this variable, as this paper also finds, that has been affected by the greatest uncertainty in recent years.

With regard to future developments in the framework presented here, one road forward could regard the use of flexible functional forms for output variables other than GDP and for inflation. In particular, verifying the hypothesis that the recent recessionary phase was an anomaly could be pursued by estimating a model with parameters that vary over time. In addition, in order to rigorously verify the hysteresis hypothesis, we could consider a model whose cycle and trend components are correlated, as proposed by Proietti et al.[30] Finally, the specification of the Phillips curve could benefit from a more flexible parameterisation that take account, for example, of stochastic variance, as proposed by Stock and Watson.[31]

[29] See note 20.
[30] See the work cited in note 27.
[31] Stock H.J. and Watson M. (2010), "Modeling Inflation After the Crisis", NBER Working Paper No. 16488, October.