BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

Modelling Italian potential output and the output gap

by Antonio Bassanetti, Michele Caivano and Alberto Locarno

Number 771

September 2010

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Modelling Italian potential output and the output gap

by Antonio Bassanetti, Michele Caivano and Alberto Locarno

Number 771 - September 2010

MODELLING ITALIAN POTENTIAL OUTPUT AND THE OUTPUT GAP

by Antonio Bassanetti*, Michele Caivano*, Alberto Locarno*

Abstract

The aim of the paper is to estimate a reliable quarterly time-series of potential output for the Italian economy, exploiting four alternative approaches: a Bayesian unobserved component method, a univariate time-varying autoregressive model, a production function approach and a structural VAR. Based on a wide range of evaluation criteria, all methods generate output gaps that accurately describe the Italian business cycle over the past three decades. All output gap measures are subject to non-negligible revisions when new data become available. Nonetheless they still prove to be informative about the current cyclical phase and, unlike the evidence reported in most of the literature, helpful at predicting inflation compared with simple benchmarks. We assess also the performance of output gap estimates obtained by combining the four original indicators, using either equal weights or Bayesian averaging, showing that the resulting measures (i) are less sensitive to revisions; (ii) are at least as good as the originals at tracking business cycle fluctuations; (iii) are more accurate as inflation predictors.

JEL Classification: E37, C52.
Keywords: potential output, business cycle, Phillips curve, output gap.

Contents

1. Introduction 5
2. Estimation methods 7
   2.1 Four alternatives 8
   2.2 Combining output gap indicators 10
3. Output gap indicators, business cycle dating and 'quasi real-time' estimation 11
   3.1 Business cycle dating: definitions and problems 13
   3.2 'Quasi real–time' estimates 20
4. Phillips curve and the predictive power of the output gap 29
   4.1 Model estimation 30
   4.2 Forecast comparison 32
5. Summary and conclusions 37
References 40
Appendix 1: methods for estimating potential output 42
Appendix 2: model averaging 50

* Bank of Italy, Economic Outlook and Monetary Policy Department. E-mail: antonio.bassanetti@bancaditalia.it, michele.caivano@bancaditalia.it, alberto.locarno@bancaditalia.it.

1. Introduction[1]

Measures of demand-supply imbalances in the goods and labour market - or in the economy as a whole - are frequently used as business cycle indicators and play a key role in economic analysis. Estimates of the cyclical position are important to determine prospective inflationary pressures and to choose the appropriate monetary policy stance; besides, they are a key input into calculations of structural government-sector budget balances, which are necessary to assess the sustainability of fiscal policies.

The most common proxy of demand-supply imbalances is the output gap, defined as the difference between an economy's output and its potential level. Although macroeconomic analysis often takes measurement of the output gap for granted, its construction is subject to considerable uncertainty. Supply capacity is a latent variable and must accordingly be estimated, but there is no general consensus on which is the best way to proceed; it is usually identified by trend GDP, while the gap is measured by the residual stationary component: unfortunately there is an infinite number of ways to operate such decomposition, with widely different business cycle and policy implications. Theory-based and statistical methods compete for prevalence, but each approach presents shortcomings and deficiencies: data-driven techniques in general provide a good fit and exhibit reasonable properties, but resemble a black-box whose outcome is difficult to interpret and whose working is rigid and unsuitable for more comprehensive frameworks; theory-based approaches are more easily related to economic shocks and – at least in principle - allow some role for policy variables, but tend to perform less well empirically.

Besides model uncertainty, three practical problems plague the estimation of potential output and the use of gap measures: (1) sensitivity to data revisions; (2) end-of-sample (as well as beginning-of-sample) uncertainty; (3) sensitivity to structural changes.

1. Observations used to filter potential output are subject to substantial revisions, implying that the gaps estimated from real-time preliminary data may differ from

[1] This paper summarizes the findings of the working group set up at the Bank of Italy with the mandate to estimate Italian potential output. The other members of the working group are Fabio Busetti, Silvia Fabiani, Fabrizio Venditti and Francesco Zollino. We gratefully acknowledge their contribution. We would like to thank Fabio Busetti for his

those obtained from final data: it takes usually three years before National Account statistics become final and the difference between the preliminary and the final release can be quite large, possibly affecting the location of turning points and the length and shape of the business cycle. Data uncertainty may be particularly harmful because by definition the output gap fluctuates around a zero mean and its sign may thus switch when revisions are non-negligible.

2. The trend and cyclical components of actual GDP are inherently two-sided concepts, in the sense that current potential output is a function of past as well as future growth. This characteristic greatly complicates estimates of the value of trend GDP at the beginning (where there are no past observations) and at the end (where there is no future data) of the sample. The end-of-sample uncertainty is particularly problematic because the latest observations are those that are most relevant for real-time policy analysis.[2] A possible solution to overcome this latter problem is to extend the sample with forecasts: this procedure should mitigate somehow the end-of-sample bias, but rests entirely on the accuracy of the projections and is likely to perform poorly in the proximity of turning points.

3. The sensitivity to structural changes bears on the usage of the output gap as an indicator of demand pressures on prices From a policy perspective, the output gap is a potentially useful predictor for inflation because it summarises the demand-supply imbalances arising in the goods and labour markets. It proxies for marginal costs and mark-up pricing and captures the domestic component of inflation.[3] Regardless of the attention paid to the specification of the equation, the link between the output gap and inflation is non-structural, heavily dependent on the institutional framework and, as such, subject to the Lucas' critique: it is therefore not surprising that most

many useful comments to earlier versions of the paper and two anonymous referees for their valuable suggestions. All remaining errors are ours.

[2] Watson (2007) finds that one-sided real-time estimates forecast only 50 percent of the variability in historically measured gaps and business cycle components.

[3] Accordingly, the output gap is best suited for predicting the value added deflator, rather than consumer prices. Moreover, since the government sector does not usually respond to market incentives, it is more reasonable to define

empirical analyses have found at best a weak correlation between the two variables, particularly when real-time data are used.

This paper presents some evidence on the trend-cycle decomposition of Italian GDP. Four different techniques are considered: two of them are based on statistical methods, one uses a production function approach; one relies on structural VAR modelling[4]; results are also presented for output gap measures obtained by equally weighting and Bayesian averaging those four estimates.

The article is organised as follows. The next part contains a brief description of the techniques used to estimate potential output as well as of the methods used to calculate combined measures by Bayesian model averaging; a more detailed account of the procedures is provided in Annex 1. Section 3 examines whether the alternative output gap indicators are consistent with the reference dating of the business cycle and compares their performance with other available estimates. Section 4 analyses whether quasi real-time estimates of the output gap are reliable predictors of inflation and studies how the forecasting accuracy deteriorates as the projection horizon lengthens. Section 5 concludes.

2. Estimation methods

Potential output describes the supply side potential of an economy and represents the level of output that can be produced without generating inflationary pressures; since it cannot be directly observed, it needs to be estimated. A large body of research has been devoted to this endeavour. Unfortunately, there is not a univocal way to estimate potential output and several alternative methods have been proposed, each having its own advantages and disadvantages. Currently, there are three leading approaches to measuring potential output and the output gap: (i) theory-free time series methods, based on filtering techniques

---

the gap with respect to the value added deflator of the private sector, rather than the whole economy, as in most empirical studies, and before indirect taxes are levied.

[4] The sample period for the statistical approaches starts in 1982Q1 and ends in 2008Q2; the methods based on the production function and on structural VAR modelling use also observations from the 1970s. Using the sample starting in the 1970 delivers similar results for the UNC and TVAR methods (at the cost of increased computing time), while it

such as the Hodrick-Prescott and band-pass filters; (ii) the production function approach, relying on measures of factor endowments, labour skills, capacity utilisation, equilibrium unemployment rates and technological progress; (iii) multivariate estimation procedures, which make use of theory-driven econometric techniques; the two economic relationships which are typically exploited in this approach are the Okun's law, relating output and unemployment fluctuations, and the Phillips curve, linking changes in inflation to labour market tightness.

In some cases these methods are applied to growth rates of output instead of levels, therefore delivering estimates for potential growth; an estimate of the output gap can still be recovered after linking these underlying growth rates to some agreed level of potential output at a specific point in time, considered to be a period of zero demand-supply imbalances.

2.1 Four alternatives

This paper shows the results obtained by pursuing four alternative estimation methods, covering the three leading approaches mentioned above and exploiting a quarterly sample that covers the period 1980Q1-2008Q2.[5]

The availability of alternative measures has the advantage of providing ranges - rather than point values - of potential output, which is an intuitive way to gauge the uncertainty surrounding the value of the supply capacity of the economy; it also allows to test whether aggregate measures improve the fitting and the forecasting performance of the gap indicator.

1. The first approach uses unobserved components methods (UNC) to filter potential output out of actual GDP data. Unlike the common practice, which focuses on maximization of the likelihood, here we present Bayesian estimates imposing informative priors on the parameters that define the cyclical component. The goal is

---

slightly improves those of the PF and SVAR methods. For this reason we use a longer sample for the latter methods only.

to achieve a better identification of the model coefficients by increasing the curvature of the posterior distributions and possibly to avoid multiple local maxima. The approach can be extended to a multivariate framework by including an additional equation tracking inflation dynamics, or a Phillips curve. One strength of a Bayesian setup is that, unlike more standard trend-cycle decompositions, it yields smoother estimates of potential output growth and provides more persistent output gap measures.

2. Another way to extract potential from actual output is to fit a time-varying autoregressive model (TVAR) to the log change of real GDP and then use the estimated intercept (adjusted to take into account the lag structure of the model) as the (time-varying) growth rate of potential supply. This is again an unobserved component model, but applied to growth rates rather than levels. When the variance of the trend component (i.e. potential growth) is small, the maximum likelihood estimator may be inappropriate since there is a high probability of obtaining zero as estimate of its volatility. Thus the method of median unbiased estimation is adopted, since it allows to fit to the data the variances of the trend and cyclical components, providing a series of (slowly) time-varying potential growth rate of output.

3. The production function approach (PF) assumes that the production possibilities set is well described by a Cobb-Douglas function: potential output is obtained as the level of net production compatible with the equilibrium level of capital, labour and total factor productivity. The method combines theory-driven inputs (the functional form of the production possibilities set; the assumption of the malleability of capital; the definition of the equilibrium rate of unemployment) and statistical tools (filtering devices to estimate the low-frequency value of total factor productivity, the participation rate and the unemployment rate). Its main advantage is that of providing not only a measure of supply capacity, but also an estimate of the contribution of each production input to the growth rate of potential output.

[5] Moreover, in order to reduce the distortion induced by the end-of-sample bias, the sample is augmented with forecasts up to 2010Q4, obtained by fitting a univariate AR(4) model to each of the input series.

4. The final approach is theory-based and requires estimating a structural VAR model (SVAR) of output, inflation, real wages and the unemployment rate. The model is an extension of the Blanchard and Quah (1989) framework, augmented with a price-wage block, and identifies four different shocks: a technology shock, a nominal disturbance, a wage-push shifter and a demographic shock. Potential output growth is defined as the time series obtained from actual GDP after removing the effect of the nominal shock. Compared with the other approaches, this method is more theory-driven; moreover, it can be easily adjusted to allow for the inclusion of the price of oil and of financial shocks among the determinants of supply capacity. However, it suffers from the well-known weaknesses of VARs, namely the volatility of the estimated parameters and the ad-hocness of the identification of shocks. Unlike the other approaches, no smoothing procedure is used to compute potential output, which implies that the estimated series is more volatile and less serially correlated.

2.2 Combining output gap indicators

The four estimation methods outlined in the previous paragraph may have, in general, virtues and advantages with respect to one another. This suggests that finding a way of combining them into a single indicator could be an effective way to improve their reliability and reduce their estimation error. To this end, we have used model averaging techniques in order to calculate appropriate weights for combining the output gap measures. More specifically, we have used two different aggregation procedures: a) equal weights (EQW) for all models and b) weights that are proportional to the forecasting accuracy of each indicator in predicting inflation.

The procedure can be approximately described as follows (see Appendix 2 for details):

- first, we estimate, for each of the methods outlined above, a Phillips curves of the form:

$$\pi_t = \alpha + \sum_{s=1}^{p} \delta_s \pi_{t-s} + \sum_{j=1}^{L} \gamma_j x_{t-j}^i + \sum_{h=1}^{H} \omega_h w_{ht} + \varepsilon_t$$

where π_t denotes inflation, x_t^i the output gap delivered by method i (with i = *UNC, PF, TVAR, SVAR*), w_{ht} denote additional explanatory variables (e.g. dummy variables), p and L denote the maximum lag order of inflation and output gap respectively; ε_t is an i.i.d. disturbance term[6].

- second, we compute a set of preliminary weights based on the predictive performance of the Phillips curves; to this end, we calculate the posterior predictive densities $p(Y_f|Y^*, M_i)$, representing the marginal densities of future data Y_f conditional on the Phillips curve model M_i and on past data Y^*.
- third, as the various models may bring similar pieces of information, giving rise to correlated forecasts, we compute a correction factor $l[\rho(M_i, \overline{M})]$ based on the correlation between (i) the estimated residuals implied by M_i and (ii) those implied by a benchmark model $\overline{M}$. Such a scaling factor penalizes models delivering similar forecasts, thus maximising the extraction of the original information content embedded in each output gap measure.
- Finally, we combine the posterior predictive weights with the scaling factors.

After normalizing, the resulting weights are 0.346 for the UNC method, 0.33 for the PF, 0.234 for the TVAR and 0.09 for the SVAR.[7]

3. Output gap indicators, business cycle dating and 'quasi real-time' estimation

Any attempt to measure potential output and the output gap faces the difficulty of assessing the plausibility and reliability of estimates of variables which are not observable. Among the properties that must be possessed by a reasonable output gap measure, two have been usually given priority: the ability to capture the cyclical fluctuations of the economy and the lack of excess sensitivity to data revisions and re-estimation. In the existing literature, such evaluation is implemented by resorting to three types of checks, which involve: (i) testing the precision of turning-points detection; (ii) measuring the

[6] For additional details about the actual specification of the Phillips curves, see Section 4.

[7] The role of the scaling factor in computing the weights is limited: without accounting for it, the weights would change only marginally, the SVAR method being the most affected (its weight would decrease to 0.07).

volatility of the set of vintages of estimates; (iii) assessing the accuracy of the output gap as an inflation predictor.

To appraise whether the estimated output gap indicators succeed in detecting the sequence of turning points, we first focus on what we consider the 'final' estimates of the Italian output gap, i.e., those obtained exploiting the whole dataset at our disposal (1980Q1- 2008Q2), and compare them with the reference dating of the Italian business cycle.

We then move on to consider 'quasi real-time' estimates of the output gap and compare them with the 'final' ones. Basically, the 'quasi real-time' estimate of potential output in *t* is obtained by excluding, from the most updated sample (1980q1-2008q2), all information unknown as of time *t*, but retaining revisions for observations of previous periods made after time *t*.[8] Starting from 1999Q1, we add one observation each period and, at each step, we estimate potential GDP and the gap, storing only the value referred to the last quarter of the sample; joining all these estimates, we obtain the 'quasi real-time' series.[9] For a more formal definition, let ${}^{t}\Im^{T}$ be the information set including data up to period *t*, released in *T* (with $T > t$). For estimation method $\mu_k \in \mathcal{M}$, time-*t* real-time estimates of potential output are defined as $E\left(\bar{y}_t \middle| {}^{t}\Im^{t}\right)$, while the 'quasi real-time' counterpart is denoted by $E\left(\bar{y}_t \middle| {}^{t}\Im^{T}\right)$;[10] the sequence of 'quasi real-time' estimates is therefore given by

$$E\left(\bar{y}_t \middle| {}^{t}\Im^{T}\right) \qquad for \quad 1999Q1 \le t \le 2008Q2 .$$

The advantage of 'quasi real-time' estimates is that they do not require to keep track of all the vintages of data, but only of the latest release; the main disadvantage is that the importance of data revisions is neglected and results in a somewhat biased assessment of the variability and predictive power of output gap indicators. However, since this simplification

---

8 In other words, we always use the data vintage 1980Q1-2008Q2, but we cut it at *t* when estimating potential output at *t*.

9 For the TVAR and SVAR methods, potential GDP is estimated by cumulating the estimated potential growth rates to the GDP level of 1992Q1, which is deemed to be a period of zero gap between actual and potential output according to the OECD indicator. Besides, both the UNC and the PF methods, that allow to directly estimate the level of potential GDP (and, as such, to endogenously determine the output gap) deliver a zero-gap in the same period.

affects all estimation methods, presumably it has a minor impact in ranking the performance of the competing measures of potential output.

The assessment of the reliability of gap indicators is made by comparing the 'quasi real-time' and the 'final' estimates, to get a quantification of the total revisions that each approach undergoes before settling at its 'final' value. The criteria that we adopt at this stage are based on the magnitude of total revisions, on their bias (their mean being possibly positive or negative), persistence and statistical significance; we also focus on the capability of the 'quasi real-time' estimates to detect, timely and correctly, the turning points of the business cycle.

All the empirical analyses documented in this section are based on a sample that excludes the last six observations; this choice has the advantage of (i) reducing the end-of-sample problem and (ii) discarding estimates that are less accurate, being based on still preliminary National Accounts data.[11]

3.1 Business cycle dating: definitions and problems

Since the output gap measures the deviation of actual GDP from its long-run component, its performance as a business cycle indicator should be assessed with respect to a 'growth-cycle' reference dating.[12] Unfortunately, unlike the U.S., an "official" dating of the Italian business cycle does not exist; moreover, the very few authors that have attempted to identify the peaks and troughs of the fluctuations of the Italian economy have not covered the time frame we are interested in.

Accordingly, we have dated the business cycles by combining different sources of information: for the first part of the sample (up to 1996Q4) we have used the unofficial

---

[10] The output gap estimates are $y_t^t - E\left(\bar{y}_t \middle| {}^t\Im^t\right)$ and $y_t^T - E\left(\bar{y}_t \middle| {}^t\Im^T\right)$, with $y_t^t \in {}^t\Im^t$ and $y_t^T \in {}^t\Im^T$ denoting GDP released in *t* and *T*, respectively.

[11] In the literature it is common practice to discard a larger number of observations; given the reduced length of our sample and the amount of revisions that the methods we tested usually undergo, we chose to eliminate as few as possible estimates.

[12] The alternative option is to refer to the upswings and downswings in the level of economic activity (and possibly of some other macroeconomic variables), as in the classic NBER approach.

(though widely used) growth-cycle dating provided by Altissimo *et al.* (2000); for the 1997-2005 period, we have applied the Bry-Boschan non-parametric algorithm[13] to the detrended coincident indicator of the Italian economy estimated by the Bank of Italy;[14] for the subsequent years, as the peaks and troughs cannot be determined by standard dating algorithm,[15] we have relied on purely judgemental criteria, picking up, in accordance with a widespread consensus, a trough at the beginning of 2005.[16]

The reference business cycle dates have been compared with the cyclical turning points, detected by means of the Bry and Boschan algorithm, identified by each estimated 'final' output gap indicator.[17] As an additional performance test, we have compared our measures with those derived from well-known filters (Hodrick-Prescott, HP for short, and Christiano-Fitzgerald, CF for short) and with the one published by the OECD.

A visual inspection of Fig.1 (upper panel) suggests that all methods generate 'final' gaps that are consistent both with the Italian business cycle and with the output gap measure published by the OECD, the only institution that, to the best of our knowledge, provides quarterly estimates of the degree of economic slackness.

From Fig.1 clearly stands out that in the early 1980s the gap indicators are much less alike than in the following period, most likely due to a non-negligible beginning-of-sample bias; indeed, the volatility across methods is much higher, though all estimates clearly capture the cyclical downturn. The 1984-86 downturn is instead badly tracked, with the

---

[13] See Mönch and Uhlig (2004) for a concise but clear exposition of the algorithm; Bry and Boschan (1971) provides the fully-fledged description.

[14] Over the overlapping period - the 17 years between 1980 and 1996 – the Bry and Boschan algorithm applied to the detrended coincident indicator identifies nearly exactly the same peaks and troughs as those picked out by Altissimo *et al.*

[15] This is due to the impossibility of verifying whether standard requirements about the length and shape of the business cycle are satisfied. Besides, the need to apply bilateral filters to detrend the input time-series prevents from having updated information.

[16] Since the exact position of the following peak, which should be at the beginning of 2007, is still largely uncertain even on a judgemental basis, we preferred to exclude it from the analysis.

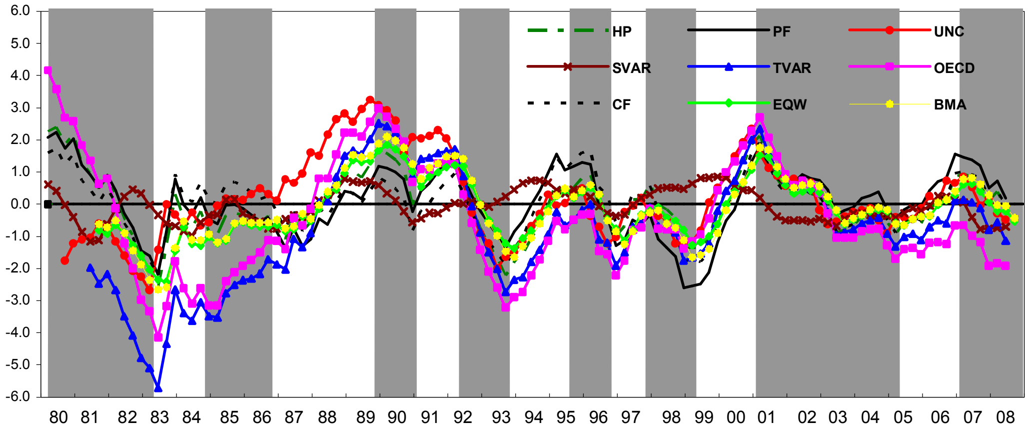
[17] The location of peaks and troughs in the reference business cycle dating is quite robust to changes in the setting of the Bry-Boschan algorithm; in particular, it hardly varies when the minimum length of either the complete cycle or a single phase (i.e. a downswing or an upswing) is modified. The output of the algorithm has not been taken at face value: in those cases in which it did not classify turning points that were evident upon visual inspection, it has been amended on the basis of judgemental considerations.

partial exception of the PF and SVAR methods; the HP and CF filters perform reasonably well, while the OECD indicator mistakes the peak for the trough and vice-versa.

Fig.1 also shows that in general the SVAR-based gap performs poorly compared with the other measures: it tends to underestimate the width of output fluctuations, exhibits a very small inertia and crosses too many times the horizontal zero line. Unlike single methods, which display "partial" failures at selected point in time, the sequence of expansions and recessions described by the combined measures tracks quite closely the movements of the OECD indicator, albeit with narrower fluctuations, and the location of the turning points coincides with the peaks and troughs of the reference business cycle dates.

Figure 1

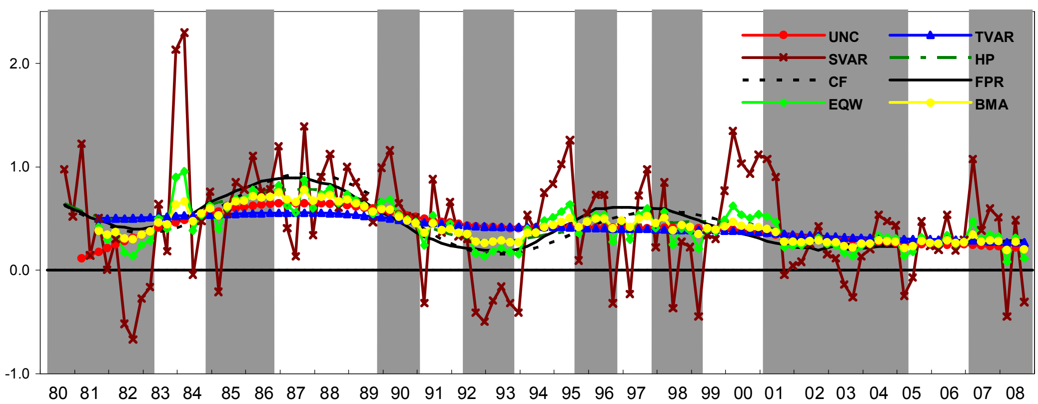
Final Estimates of the Output Gap



Note: shaded areas represent recessions (from peak to trough); the cyclical phases of the last two years cannot be determined by standard dating algorithm, therefore they have to be intended as merely tentative. Gaps are in percentage points.

Final Estimates of Potential Output

(quarter on quarter rates of growth)



Note: shaded areas represent recessions (from peak to trough); the cyclical phases of the last two years cannot be determined by standard dating algorithm, therefore they have to be intended as merely tentative.

The visual evidence of Fig.1 is confirmed by the statistics presented in Table 1. The amplitude and variability of the output-gap measures are broadly similar across methods, the notable exception being the SVAR-based estimates. All gaps have a sample mean that is very close to zero, apart from TVAR, which tends to amplify the intensity of the downturns and shows a negative bias, partly reflected also in the combined gaps.[18] Regardless of the choice of the weighting scheme, in the majority of cases the combinations turn out to be less volatile and to exhibit narrower fluctuation ranges.

Table 1

'Final' Output Gap Summary Statistics

(time range: 1981.2 - 2006.4)

Method	MEAN	SD	MIN	MAX
Hodrick-Prescott	-0.09	0.94	-2.59	2.13
Christiano-Fitzgerald	-0.05	0.84	-2.24	1.91
Production Function	-0.05	0.97	-2.59	1.79
Unobserved Components	0.22	1.29	-2.67	3.25
Structural VAR	-0.02	0.49	-1.17	0.85
Time-varying AR	-0.81	1.77	-5.72	2.51
Comb. indicator: EQW	-0.17	0.91	-2.41	1.84
Comb. indicator: BMA	-0.13	1.04	-2.65	2.10
OECD (for comparison)	-0.53	1.64	-4.14	2.98

Note: the table shows the mean, standard deviation (SD), minimum and maximum of the 'final' estimates.

As regards potential growth, Fig. 1 (lower panel) shows that methods displaying a higher volatility in the output gap obviously deliver a smoother estimate of the trend component. This reflects different weighting of the relevant frequencies defining the long run and cyclical components as well as the usual trade off between non-parametric and model-based filters. Additional results are shown in Table 2, where we report the leads or lags at turning points exhibited by 'final' output gap measures with respect to the benchmark business cycle dating. For the sake of comparison, in Table 2 we report also the location of the turning points based on the level of economic activity (classic dating).

[18] For the sake of comparison with the results obtained for other countries, tables 1, 3, 4 and 6, are identical to those

Table 2

Turning Point Analysis: Leads(-) and Lags (+) with Respect to the Growth-Cycle Reference Dating

Method	Trough	Peak	Trough	Peak	Trough	Peak	Trough	Peak
Growth-cycle Reference dating	**1983q2**	**1984q3**	**1986q3**	**1989q4**	**1991q1**	**1992q1**	**1993q3**	**1995q3**
Hodrick-Prescott	0	+4	+2	0	-1	0	0	+2*
Christiano-Fitzgerald	0	+4	+4	0	-1	0	0	+2
Production function	0	+4	+2	0	-1*	0*	0	-2*
Unobserved components	-1	--	--	-1	--	--	-1	+1*
Structural VAR	+4	+4	0	-4	-1	--	--	-5
Time-varying AR	0	--	--	0	-1*	0*	0	+2*
Comb. indicator: EQW	0	--	--	0	+1*	0	0	+1
Comb. indicator: BMA	-1	--	--	0	+1*	+1*	0	+1
for comparison: **Classic (in levels) Reference dating**	**1983q1**	**--**	**--**	**--**	**--**	**1992q1**	**1993q3**	**1995q4**

Method	Trough	Peak	Trough	Peak	Trough	# of extra (+) or missed (-) cycles
Growth-cycle Reference dating	**1996q4**	**1997q4**	**1999q2**	**2000q4**	***2005q1*****	
Hodrick-Prescott	0*	0	-2	+1	*0***	0
Christiano-Fitzgerald	0*	0*	0	+1	*0***	0
Production function	0*	0	0	+1	*0***	0
Unobserved components	-1*	-1	-3	0	*-1***	-2
Structural VAR	0	--	--	-4	*0***	-2
Time-varying AR	0*	0	-2	+1	*0***	-1
Comb. indicator: EQW	0	0	-2	0	*0***	-1
Comb. indicator: BMA	0	0	-2	0	*0***	-1
for comparison: **Classic (in levels) Reference dating**	**1996q4**	**--**	**--**	**2000q4**	***2005q1*****	

Note: * indicates the lead or lag (with respect to the reference dating) of output gap turning points that are not classified as such by the Bry and Boschan algorithm but that are evident on a visual inspection; ** indicates that the last turning point of both the reference dating and of the output gap estimates cannot be determined by standard dating algorithms but just according to a judgemental criterion.

If one considers only the peaks and troughs identified by the Bry-Boschan algorithm, all methods seem to be somewhat underperforming, missing one or more cycles, in particular those in the early 1980s and mid 1990s. However, if one includes also the peaks and troughs identified judgementally (labelled with an asterisk in Table 2), the evidence is much more supportive: (i) the univariate filters and the PF indicator correctly pick out all the turning points, though at times with some leads or lags; (ii) the TVAR gap misses only the business cycle of the mid 1980s; (iii) none of the gap measures identifies extra cycles. The results concerning the UNC and SVAR methods are less favourable: both miss two

presented in Orphanides and van Norden (2002), that have become the standard for the profession.

complete cycles and the latter fails to timely identify most of the turning points. To sum up, it seems that the best method to detect the business cycle maxima and minima is the one based on the production function.

Combined measures inherits some of the failures of single-method indicators, although the empirical evidence shows that averaging the output gap estimates is beneficial. Like the UNC and TVAR indicators, the aggregated estimates miss the downturn starting in 1984Q3 and ending in 1986Q3, but unlike them (and the OECD gap as well) they do not mix it up with an upturn; all the other turning points are correctly detected and no extra cycles are identified.

Table 3 provides additional information on how well the competing gap indicators track economic fluctuations, focusing in particular on the timeliness of the detection. The empirical evidence suggests that, on average, the TVAR series outperforms other single-method indicators: it tends to slightly anticipate the reference dates of the troughs and to defer those of the peaks, but the mean bias is basically zero. The PF estimates also show very good tracking properties, in particular at troughs, while the UNC and especially the SVAR indicator clearly underachieve, the former systematically underestimating the length of the downswing, the latter largely overestimating it. The combined measures detect the peaks and troughs with sharp precision, showing neither a systematic lead nor a systematic lag at turning points.

Table 3

Average Lags of the 'Final' Estimates

Method	Average lag at		
	Peaks	Troughs	All
Hodrick-Prescott	+1.2	-0.1	+0.6
Christiano-Fitzgerald	+1.2	+0.4	+0.8
Production function	+0.5	+0.1	+0.3
Unobserved components	-0.2	-1.4	-0.8
Structural VAR	-2.2	+0.6	-1.8
Time-varying AR	+0.6	-0.5	+0.1
Comb. indicator: EQW	+0.2	-0.2	0
Comb. indicator: BMA	+0.4	-0.3	+0.1

Note: a positive (negative) sign indicate a lag (lead).

So far the results of the analysis confirm the adequacy of (nearly all of) the output gap estimates, showing that in most cases they (i) pick out correctly the turning points, (ii) do not find nonexistent peaks and troughs, (iii) assess quite accurately the length of the cyclical upswings and downswings.

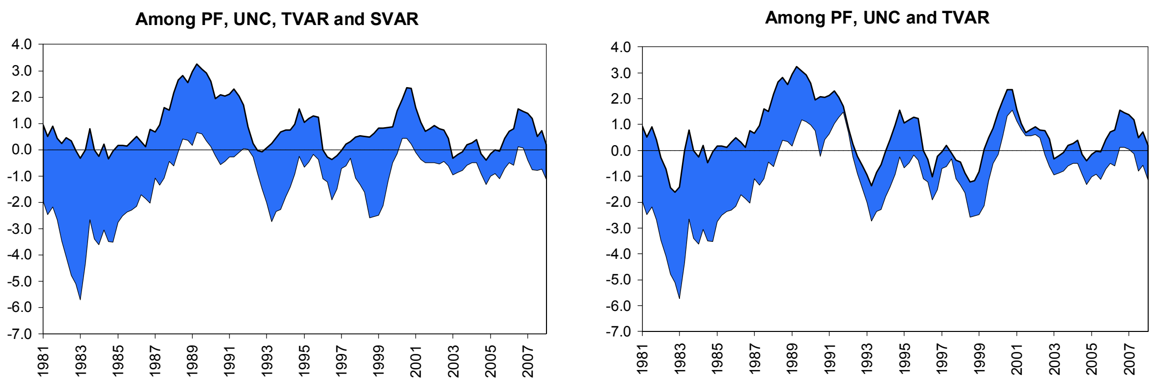
The combined measures perform at least as well as single-method indicators, although they are not uniformly superior along all dimensions. However, single-method indicators differ considerably from one another in some instances, so conveying the sense of a non-negligible degree of uncertainty about the cyclical position of the economy.

It is well known that when model uncertainty is high, a viable and well-performing solution is to rely not on a single estimator, but rather on a combination of measures, thus providing further support to our model averaging approach. A rough proxy of model uncertainty can be retrieved by Fig. 2 (left panel), in which the noisiness in the estimates is summarized by plotting the area comprised between the maximum and the minimum value of the output gap as assessed by the four indicators.

Across-method divergence seems to be higher in the (early) 1980s, but still somewhat sizable in other parts of the sample; the indeterminacy region would shrink somewhat if we dropped the SVAR estimates (Fig. 2, right panel).

Figure 2

Uncertainty About the Quarterly Output Gap



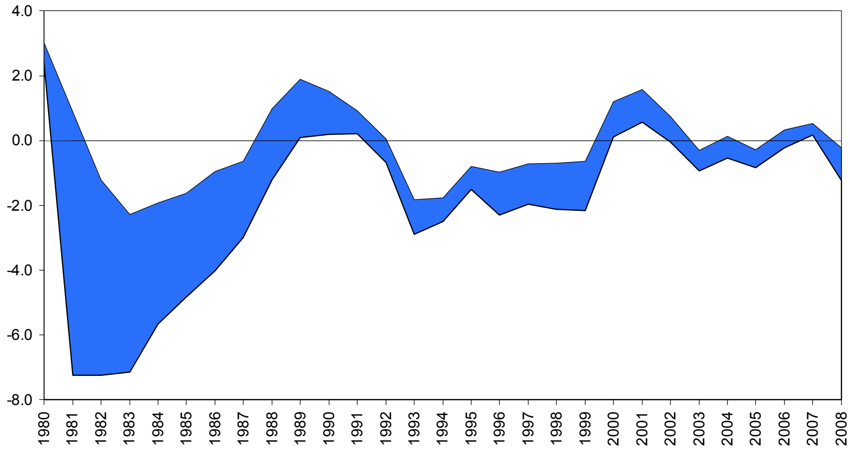
Note: the shaded areas are comprised between the maximum and minimum value among the output gap estimates.

Similar results are obtained when considering the annual estimates released by international institutions (OECD, IMF, ECB, European Commission; Fig. 3).

Again, measures of the output gap in the early 1980s turn out to be noisy and unreliable: since in this case some of the gap series starts well in advance of the 1980s, the beginning-of-sample bias does not seem a convincing explanation of this pattern. An economic, rather than a statistical reason could account for the inaccuracy of the estimates, that is the difficulty to gauge the impact on potential output of the restructuring process that took place after the second oil shock in the Italian business sector.

Figure 3

Uncertainty About the Annual Italian Output Gap According to the Estimates of International Institutions



Note: the shaded area is comprised between the maximum and minimum value among the OECD, IMF, ECB and European Commission output gap estimates.

3.2 'Quasi real–time' estimates

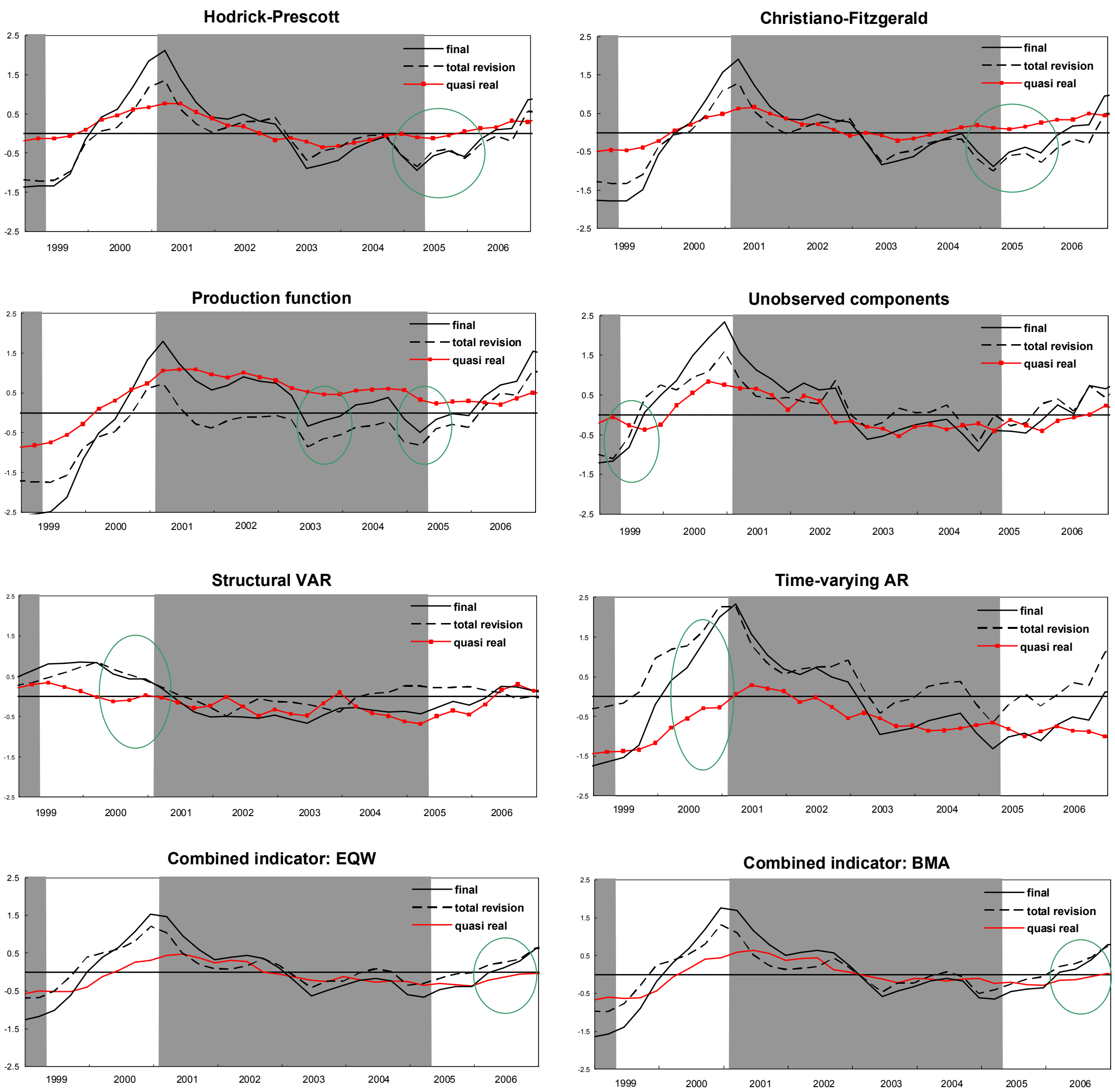
In Figure 4 we show the 'quasi real-time' sequence of quarterly output gap $\left\{y_t^{2008Q2} - E\left(\bar{y}_t \middle| {}^t\Im^{2008Q2}\right)\right\}$ for *1999Q1* $\le t \le$ *2008Q2* (i.e. the sequence formed by collecting the time-t estimates of the output gap computed on a sample containing up to time t observations belonging to the 2008Q2 vintage of National Accounts data) and the difference between the 'final' and the 'quasi-real time' estimates.

As it can be seen, the size of the revisions is often of the same order of magnitude of the gap itself; this drawback, already emphasized for the U.S. by Orphanides and van Norden (2002), implies that in a few instances, highlighted in the circled areas, the 'final' estimates have an opposite sign with respect to the 'quasi real-time' ones, raising doubts on the reliability of the corresponding gap as an indicator of business cycle conditions. The performance of the combined indicators is not dissimilar from that of their components: the size of the revisions is at times large, entailing the possibility of sign changes.

Figure 4

'Quasi Real-Time' Estimate of the Output Gap and its Total Revision

(total revision = 'final' - 'quasi real')



Note: shaded areas represent growth-cycle recessions (from peak to trough). The cyclical phases of the last two years cannot be determined by standard dating algorithm, therefore they have to be intended as merely tentative. Gaps are in percentage points.

The amplitude of total revisions (min. and max.) is non-negligible for all methods (Table 4). A few results are worth stressing: (i) the mean bias is low for the univariate filters (less than one third of the standard deviation), but larger for the more sophisticated approaches, in particular for the TVAR; (ii) in five out of eight cases the revision is predominantly upwards; in the others it is downwards; (iii) the TVAR indicator exhibits the largest positive revision, while the PF-based gap shows the biggest overestimation error; (iv) the first-order serial autocorrelation of the revisions is high in all cases (between 0.72 and 0.91) and reaches a maximum for the SVAR method; (v) combined measures display revisions of much smaller amplitude compared with those of single-method indicators, though their persistence is by and large the same.

Table 4

Summary Revision Statistics
'Final' versus 'Quasi Real-Time' Gap Estimates
(time range: 1999.1 - 2006.4)

Method	MEAN	SD	RMS	MIN	MAX	AR
Hodrick-Prescott	-0.09	0.60	0.60	-1.21	1.36	0.83
Christiano-Fitzgerald	-0.19	0.62	0.64	-1.32	1.28	0.83
Production function	-0.32	0.64	0.71	-1.75	1.05	0.87
Unobserved components	0.25	0.55	0.60	-1.11	1.59	0.72
Structural VAR	0.13	0.33	0.35	-0.49	0.85	0.91
Time-varying AR	0.51	0.73	0.88	-0.66	2.28	0.83
Comb. indicator: EQW	0.14	0.43	0.45	-0.69	1.21	0.86
Comb. indicator: BMA	0.11	0.50	0.50	-0.99	1.32	0.86

Note: the table shows the mean, standard deviation (SD), root mean square error (RMS), minimum, maximum and first order autoregressive coefficient (AR) of the output gap revisions obtained as the difference between the 'final' and the 'quasi real-time' estimates.

In order to assess the statistical significance of the revisions, we followed Koske and Pain (2008) and regressed the ‘final’ estimate on the ‘quasi real-time’ one, testing the joint null hypothesis that the intercept is zero and the slope equals one: if the null hypothesis is not rejected, the ‘final’ and ‘quasi real-time’ estimates aren’t statistically different, while if it is, they do differ. Unfortunately the empirical evidence points to a rejection of the null for all the approaches, implying that the impact of the revisions is not negligible (Table 5).

Table 5

Test for the Statistical Significance of the Revisions of the 'Quasi Real-Time' Estimates

(time range: 1999.1 - 2006.4)

Method	α (t-stat)	β (t-stat)	Wald test H0: α=0, β=1	
			Fstat, df=2,30 (prob)	Chi, df=2 (prob)
Hodrick- Prescott	-0.26	2.38	17.71	35.42
	(-3.25)	(10.04)	(0.00)	(0.00)
Christiano-Fitzgerald	-0.39	2.51	19.68	39.36
	(-4.67)	(9.56)	(0.00)	(0.00)
Production function	-0.71	1.89	20.28	40.55
	(-6.26)	(10.79)	(0.00)	(0.00)
Unobserved components	0.25	1.83	13.12	26.25
	(3.10)	(8.98)	(0.00)	(0.00)
Structural VAR	0.17	1.27	3.25	6.49
	(2.55)	(6.13)	(0.05)	(0.04)
Time-varying AR	1.03	1.84	15.67	31.34
	(5.32)	(7.24)	(0.00)	(0.00)
Comb. indicator: EQW	0.23	2.03	16.12	32.23
	(3.88)	(10.08)	(0.00)	(0.00)
Comb. indicator: BMA	0.12	2.03	17.21	34.42
	(1.98)	(11.06)	(0.00)	(0.00)

In spite of the size, persistence and biasedness of final revisions, the ‘quasi real-time’ output gap estimates maintain a potentially useful information content for business cycle analysis: as shown in Table 6, they exhibit a correlation with the ‘final’ estimate that is always very high (equal or above 80 per cent, with the combined measures reaching 90 per cent and the SVAR indicator, which is the worst performer, scoring 75 per cent).

Table 6

Summary Reliability Indicators 'Final' and 'Quasi Real-Time' Estimates

(time range: 1999.1 - 2006.4)

Method	COR	NS	NSR	OPSIGN
Hodrick-Prescott	0.88	0.70	0.70	0.13
Christiano-Fitzgerald	0.87	0.72	0.74	0.31
Production function	0.89	0.62	0.68	0.31
Unobserved components	0.85	0.65	0.70	0.16
Structural VAR	0.75	0.69	0.72	0.16
Time-varying AR	0.80	0.70	0.85	0.28
Comb. indicator: EQW	0.88	0.65	0.68	0.19
Comb. indicator: BMA	0.90	0.63	0.64	0.13

Note: the table shows the correlation between the 'final' and 'quasi real-time' estimates (COR) in the 1st column; the ratio between the standard deviations of the revisions and that of the 'final' estimates (NS) in the 2nd column; the same ratio computed using the mean square error (NSR) in the 3rd column; the fraction of sign switches due to revisions (OPSIGN) in the last column.

Moreover, the share of quarters in which the 'quasi real-time' measure has had opposite sign with respect to the 'final' one is fairly low, especially for the UNC and the SVAR methods.

Even the noise to signal ratios, that following Orphanides and van Norden (2002) is proxied by the ratio of the standard deviation (the root mean square) of the revisions to the standard deviation of the 'final' estimates, always take on values well below 1: the PF and UNC approaches seem to ensure the most stable estimates, while the TVAR method is the most sensitive to changes in the length of the sample.

Combined measures improve substantially over single-method indicators: the noise-to-signal ratio is 0.65 (0.68 when the MSEs are used) or less, well below that achieved by nearly all the other estimates; the share of sign changes is less than 0.2 and reaches 0.13 with predictive weights (BMA), the best performance of all methods.

Convergence towards the 'final' output gap estimates appears to be fairly quick. Fig. 5 shows how the output gap series change through time; one simple way to grasp how long it takes to have reliable estimates, is to see how the 1st release is modified after *1, 2, 3, …* years.[19] The panels plot four series: the revisions made in the first 3 years – namely $E\left(\bar{y}_t \middle| ^{t+k+4}\Im^T\right) - E\left(\bar{y}_t \middle| ^{t+k}\Im^T\right)$, with k = 0, 4, 8 – and the overall difference between the 'final' and the 1st gap estimates – i.e. $E\left(\bar{y}_t \middle| ^{T}\Im^T\right) - E\left(\bar{y}_t \middle| ^{t}\Im^T\right)$.[20]

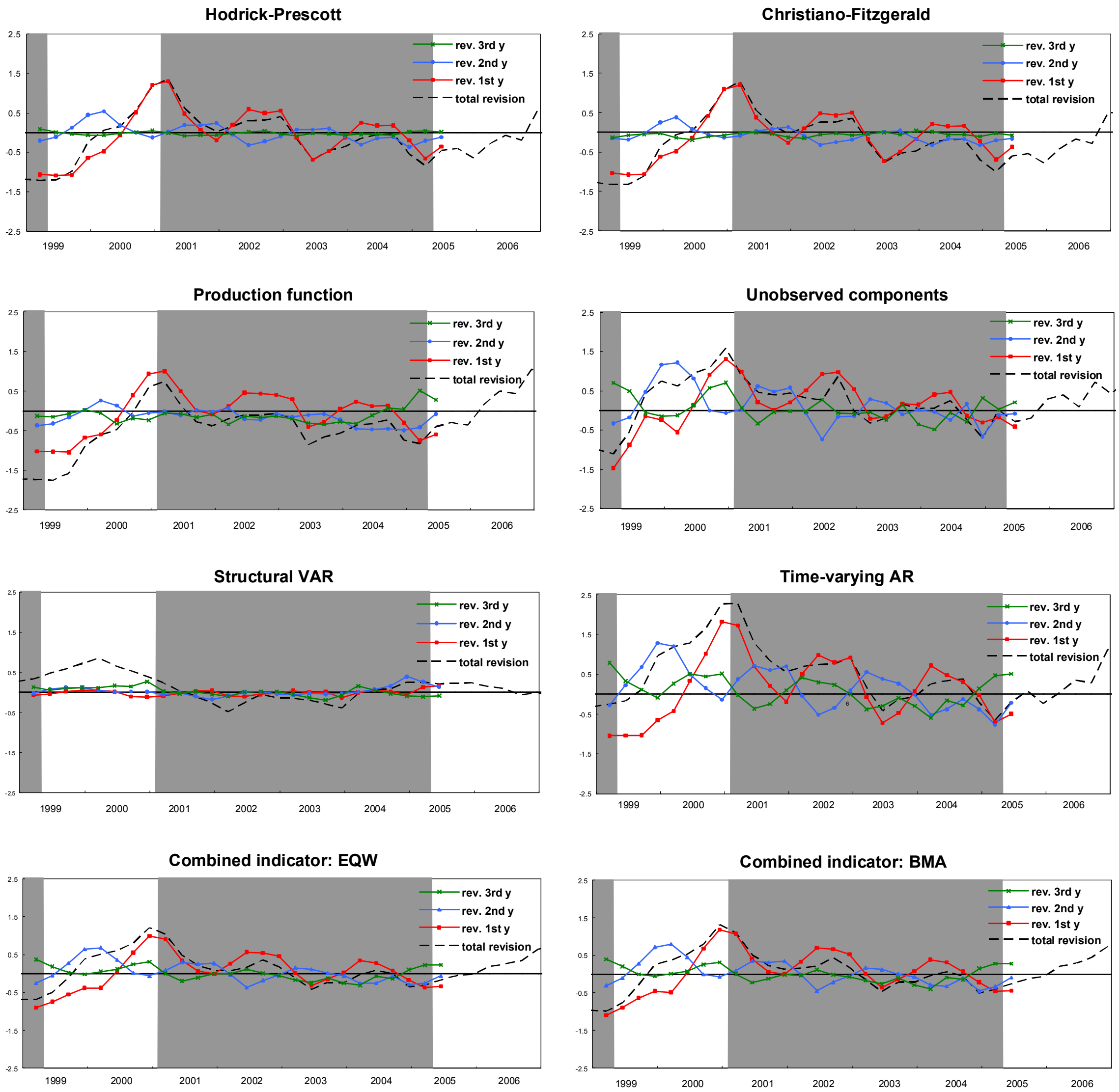
The univariate filters, the PF indicator and the combined measures are revised mostly in the 1st year (the lines 'total revision' and 'rev. 1st y' almost coincide); the UNC and TVAR gaps are subject to non negligible changes in the 2nd year as well, while the SVAR estimates are characterized by relatively small revisions.

[19] This analysis has been suggested by Tosetto (2008).

[20] Revisions depend only on potential output estimates because actual GDP enters both the minuend and the subtrahend and accordingly cancels out.

Figure 5

Revisions of the Output Gap in the First, Second and Third Year After the 'Quasi Real-Time' Estimate



Note: shaded areas represent growth-cycle recessions (from peak to trough). The cyclical phases of the last two years cannot be determined by standard dating algorithm, therefore they have to be intended as merely tentative. The four lines shows the output gap revisions made in the 1st, 2nd and 3rd year and the overall difference between the 'final' and the 'quasi real-time' estimates.

The graphical evidence is supported by the results shown in Table 7, which reports a few summary statistics (mean, min, max, standard deviation, correlation and percentage of sign changes) of the revisions.

The univariate filters and the PF indicator exhibit a correlation between the intermediate and the 'final' estimates close to 1, a low mean revision and not too extreme min and max corrections; the percentage of sign reversals is instead on the high side. The UNC series seems to be slightly more biased and volatile, though not as much as the TVAR estimates, while be SVAR gap stands out as the one with the smallest revisions. Overall, it seems fair to say that most of the methods convey a potentially useful information, even in 'quasi real-time'; their aggregation in combined measures allows to improve many of the statistics shown in Table 7, using both equal (EQW) and predictive (BMA) weights.

A crucial requisite that any business cycle indicator has to satisfy is the correct detection of business cycle reversals (i.e. turning points) in (quasi) real-time. Intuitively, a turning point identifies the time when the output gap, regardless of whether it is positive or negative, stops widening and starts narrowing. A working procedure to identify cyclical reversals is borrowed from Altissimo *et al.* (2007) and applied to our indicators: there exists a *candidate* signal of a turning point in *t-1* whenever there is a sign switch between the change of the 'quasi real-time' output gap at time *t* and that at time *t-1*, both estimated with information updated at time *t* (i.e. sign $\Delta OG_{t|t} \neq$ sign $\Delta OG_{t-1|t}$). In order to rule out signals that are spurious and due only to noisy and volatile estimates, Altissimo *et al.* suggest to accept only those candidate points that satisfy two additional requirements: (i) the signals are *consistent*, i.e. the sign of the output gap change in *t-1* does not flip when the estimate is updated because a new observation is available (sign $\Delta OG_{t-1|t}$ = sign $\Delta OG_{t-1|t-1}$); (ii) no two adjacent signals exist, i.e. the output gap varies in the same direction both in *t-1* and *t-2* (sign $\Delta OG_{t-2|t-1}$ = sign $\Delta OG_{t-1|t-1}$). When this second condition is not satisfied, regardless of whether the candidate is or isn't consistent, we consider the signal as *uncertain*. Finally a candidate signal that satisfies the two conditions and locates an upturn (downturn) at *t-1* is said to be *correct* if the reference dating actually has a turning point in the interval [*t-1-2*; *t-1+2*].

Table 7

Revisions of Output Gap Estimates at Different Time Intervals

(time range: 1999.1 - 2005.2)

Method	Mean absolute revision				Mean revision				Standard deviation				t-statistics of mean revision			
	Y1-P	Y2-Y1	Y3-Y2	F-P	Y1-P	Y2-Y1	Y3-Y2	F-P	Y1-P	Y2-Y1	Y3-Y2	F-P	Y1-P	Y2-Y1	Y3-Y2	F-P
Hodrick-Prescott	0.51	0.18	0.04	0.49	-0.04	-0.01	-0.02	-0.07	0.64	0.22	0.05	0.64	-0.07	-0.04	-0.39	-0.12
Christiano-Fitzgerald	0.49	0.16	0.07	0.53	-0.09	-0.08	-0.06	-0.17	0.60	0.17	0.06	0.66	-0.15	-0.45	-1.13	-0.26
Production function	0.46	0.20	0.18	0.57	-0.08	-0.16	-0.11	-0.46	0.57	0.20	0.19	0.60	-0.13	-0.80	-0.59	-0.77
Unobserved components	0.48	0.34	0.23	0.50	0.13	0.12	0.05	0.24	0.62	0.48	0.31	0.60	0.20	0.25	0.16	0.40
Structural VAR	0.06	0.08	0.09	0.31	-0.01	0.03	0.02	0.13	0.07	0.12	0.11	0.36	-0.15	0.27	0.17	0.36
Time-varying AR	0.67	0.44	0.31	0.72	0.14	0.16	0.09	0.56	0.81	0.53	0.35	0.78	0.18	0.29	0.25	0.72
Comb. indicator: EQW	0.38	0.21	0.14	0.36	0.05	0.04	0.01	0.12	0.47	0.27	0.18	0.46	0.10	0.14	0.06	0.26
Comb. indicator: BMA	0.45	0.25	0.16	0.41	0.05	0.03	0.00	0.08	0.57	0.33	0.20	0.53	0.09	0.09	0.01	0.14

Method	Min revision				Max revision				Correlation with final				Opposite sign w.r.t. final			
	Y1-P	Y2-Y1	Y3-Y2	F-P	Y1-P	Y2-Y1	Y3-Y2	F-P	P	Y1	Y2	Y3	P	Y1	Y2	Y3
Hodrick-Prescott	-1.09	-0.36	-0.08	-1.21	1.31	0.54	0.09	1.36	0.89	0.97	1.00	1.00	7.7	15.4	3.8	0.0
Christiano-Fitzgerald	-1.07	-0.33	-0.19	-1.32	1.21	0.38	0.04	1.28	0.91	0.97	1.00	0.99	26.9	11.5	7.7	7.7
Production function	-1.04	-0.49	-0.34	-1.75	1.00	0.26	0.51	0.73	0.96	0.98	0.98	0.99	30.8	19.2	15.4	3.8
Unobserved components	-1.46	-0.74	-0.48	-1.11	1.31	1.22	0.71	1.59	0.86	0.88	0.96	0.99	11.5	23.1	11.5	7.7
Structural VAR	-0.13	-0.17	-0.19	-0.49	0.16	0.39	0.27	0.85	0.76	0.80	0.89	0.95	19.2	15.4	15.4	7.7
Time-varying AR	-1.04	-0.77	-0.60	-0.66	1.82	1.28	0.79	2.28	0.80	0.89	0.97	0.99	30.8	11.5	7.7	0.0
Comb. indicator: EQW	-0.90	-0.37	-0.31	-0.69	0.99	0.68	0.37	1.21	0.89	0.90	0.96	0.99	7.7	11.5	3.8	0.0
Comb. indicator: BMA	-1.10	-0.45	-0.40	-0.99	1.18	0.80	0.40	1.32	0.91	0.93	0.98	0.99	3.8	11.5	3.8	0.0

Note: P = ‘quasi real-time’ estimates of the output gap; F = ‘final’ estimates; Y1-P, Y2-Y1 and Y3-Y2 are the revisions made in the 1st , 2nd , and 3rd year.

According to the evidence reported in Table 8, all methods score fairly well. The UNC indicator is the only one that shows a quite high percentage of inconsistent signals, while all the others are almost always consistent; conditional on the signals being consistent, the UNC estimates are characterized by the lowest fraction of uncertain signals, clearly outperforming the PF, SVAR and TVAR gaps and, by a larger extent, the HP and CF filters.

Table 8

'Quasi Real-Time' Detection of Turning Points

(38 signals of the sample: 1999.1 - 2008.2)

Method	Consistent signals	% inconsistent signals	% uncertainty signals	Turning Point signals	TP signals until early 2005	Correct TP	% correct TP	% missed TP
Hodrick-Prescott	36	5.3	10.5	8	6	3	50.0	0
Christiano-Fitzgerald	36	5.3	10.5	8	6	3	50.0	0
Production function	37	2.6	7.9	9	7	3	42.9	0
Unobserved components	22	42.1	2.6	4	3	1	33.3	66.7
Structural VAR	37	2.6	7.9	9	7	1	14.3	66.7
Time-varying AR	37	2.6	7.9	9	6	3	50.0	0
Comb. indicator: EQW	32	15.8	2.6	6	4	2	50.0	33.3
Comb. indicator: BMA	35	7.9	2.6	7	5	2	40.0	33.3

Over the 1999Q1-2008Q2 period, all the output gap indicators identify a rather large number of turning points: the univariate filters locate 8 cyclical reversals; the PF, SVAR and TVAR gaps identify 9 such points.[21] Apparently because of the volatility of the estimates, the UNC indicator detects only 4 turning points. Most of the signals refer to the period before 2005Q1, when the reference dating identifies only 3 turning points (see Table 2), so that the frequency of the cyclical reversal detected by the 'quasi real–time' output gap indicators appears too high. As a result, the share of correct signals (those that correspond to a true turning point) is rather low: 50 per cent for TVAR and the univariate filters; 43 per cent for PF; much less for UNC (33 per cent) and SVAR (14 per cent). Most methods pick out all the turning points; UNC and SVAR just one. As to the combined measures, although the percentage of inconsistent signals is non-negligible, in particular for the EQW output gap, it does not reach too high a level; among the consistent signals, the share of those that are considered to be uncertain is extremely low. The weighted output gaps tend to signal too many non-existent turning points: over the last nine years, despite the presence of only three reference business cycle reversals, the EQW gap picks out six turning points and the

[21] Differently from the previous part of the analysis, this exercise included the last two years of the sample (post-2006Q4) in order to ascertain the signalling performance also in the most recent period.

BMA gap seven. Once again most of misplaced signals refers to quarters prior to 2005Q1, the last trough. Both indicators miss the 2000Q4 peak. As a result, the share of correct signals is relatively low, reaching 50 per cent for the EQW output gap and 40 per cent for the BMA indicator.

4. Phillips curve and the predictive power of the output gap

An output gap indicator can be considered helpful for monetary policymaking insofar as it provides information on the degree of inflationary pressure in the economy. According to the Gordon's triangle model, actual inflation depends on three factors: (i) built-in inflation, which captures wage and price setters' expectations; (ii) demand pull, which depends on the degree of slackness in the labour and goods markets; (iii) cost push (or supply shocks), which is related mostly to commodity price developments. The output gap summarises the domestic inflationary pressures, which can arise from imbalances in the labour and goods markets and may therefore be used as a predictor for consumer price changes.

The link between measures of economic slackness and inflation has weakened in recent years, due partly to the impact of globalisation and partly to the improvements in monetary policymaking: greater competition from abroad limits firms' scope to increase prices when demand rises; enhanced capital and labour mobility reduces the response of wages and prices to domestic demand-supply imbalances; more independent and credible central banks lead to better-anchored and more stable inflation expectations. The instability of the output-inflation trade-off is confirmed by a substantial empirical evidence.[22] A number of recent papers have tested the usefulness of business cycle measures to predict inflation: in terms of out-of-sample forecasting accuracy, 'final' estimates of the output gap do not seem in general to improve upon the performance of simple benchmark models, such as random-walk or autoregressive specifications; if real time-data are used, the information

[22] The IMF (2006) finds evidence that in eight advanced economies the sensitivity of inflation to economic slackness has decreased since the 1980s. Pain, Koske and Sollie (2006) find similar results for all OECD countries, but only from 1995 onwards. Gaiotti (2008) presents evidence for the Italian economy using micro data for a sample of 2000 firms.

contained in the output (or unemployment) gap has in most cases been found to lead to a deterioration in predictive power.[23]

4.1 Model estimation

In this paper the performance of the output gap measures as inflation predictors is assessed by estimating a set of Phillips curve equations. Inflation is defined as the quarter-on-quarter rate of change of the private sector value added deflator, since this measure is likely to be more tightly related than headline consumer inflation to developments in domestic real activity, being (to a large extent) independent from foreign goods and commodity prices[24].

We focus on linear forecasting models of the form:

$$\pi_t = \alpha + \sum_{s=1}^{p} \delta_s \pi_{t-s} + \sum_{j=1}^{L} \gamma_j x_{t-j}^i + \sum_{h=1}^{H} \lambda_h d_{ht} + \varepsilon_t \,, \tag{1}$$

where π_t is inflation, x_t^i the output gap delivered by method *i,* d_{ht} denotes the h^{th} impulse dummy variable, *p* and *L* indicate the maximum lag order of inflation and, respectively, the output gap, ε_t is an i.i.d. disturbance term. In each case, the lag structure of the forecasting model and the selection of dummy variables is chosen in an "objective" way, using the Autometrics option of the PcGive software (with 'final' output gap estimates). The selected models turn out to be very similar across methods, as detailed below:

- Unobserved components (UNC)

$$\pi_t = \alpha + \beta_1 \pi_{t-2} + \beta_2 \pi_{t-3} + \beta_3 \pi_{t-4} + \gamma_1 x_{t-2} + \gamma_2 x_{t-3} + \lambda_1 d_{1993Q1} + \lambda_2 d_{1995Q3} + \varepsilon_t^{UNC} \tag{2}$$

- Production function (PF)

$$\pi_t = \alpha + \beta_1 \pi_{t-2} + \beta_2 \pi_{t-3} + \beta_3 \pi_{t-4} + \gamma_1 x_{t-1} + \gamma_2 x_{t-3} + \lambda_1 d_{1993Q1} + \lambda_2 d_{1995Q3} + \varepsilon_t^{PF} \tag{3}$$

- Time-varying AR (TVAR)

---

[23] See Orphanides and van Norden (2005) for the case of the U.S.

[24] A parallel set of regressions has also been run using the private consumption deflator as a measure of inflation. The results are very similar to those reported below, with the output gap being always a significant regressor in the Phillips curve equations.

$$\pi_t = \alpha + \beta_1\pi_{t-2} + \beta_2\pi_{t-3} + \beta_3\pi_{t-4} + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \lambda_1 d_{1993Q1} + \lambda_2 d_{1995Q3} + \varepsilon_t^{TVAR} \quad (4)$$

- *Structural VAR (VAR)*

$$\pi_t = \alpha + \beta_2\pi_{t-2} + \beta_3\pi_{t-3} + \beta_4\pi_{t-4} + \gamma_1 x_{t-2} + \gamma_2 x_{t-3} + \lambda_1 d_{1983Q4} + \lambda_2 d_{1993Q1} + \lambda_3 d_{1995Q3} + \varepsilon_t^{VAR} \quad (5)$$

- *Combined measures (BMA and EQW)*

$$\pi_t = \alpha + \beta_1\pi_{t-2} + \beta_2\pi_{t-3} + \beta_3\pi_{t-4} + \gamma_1 x_{t-2} + \gamma_2 x_{t-3} + \lambda_1 d_{1993Q1} + \lambda_2 d_{1995Q3} + \varepsilon_t^{A}, \quad (6)$$

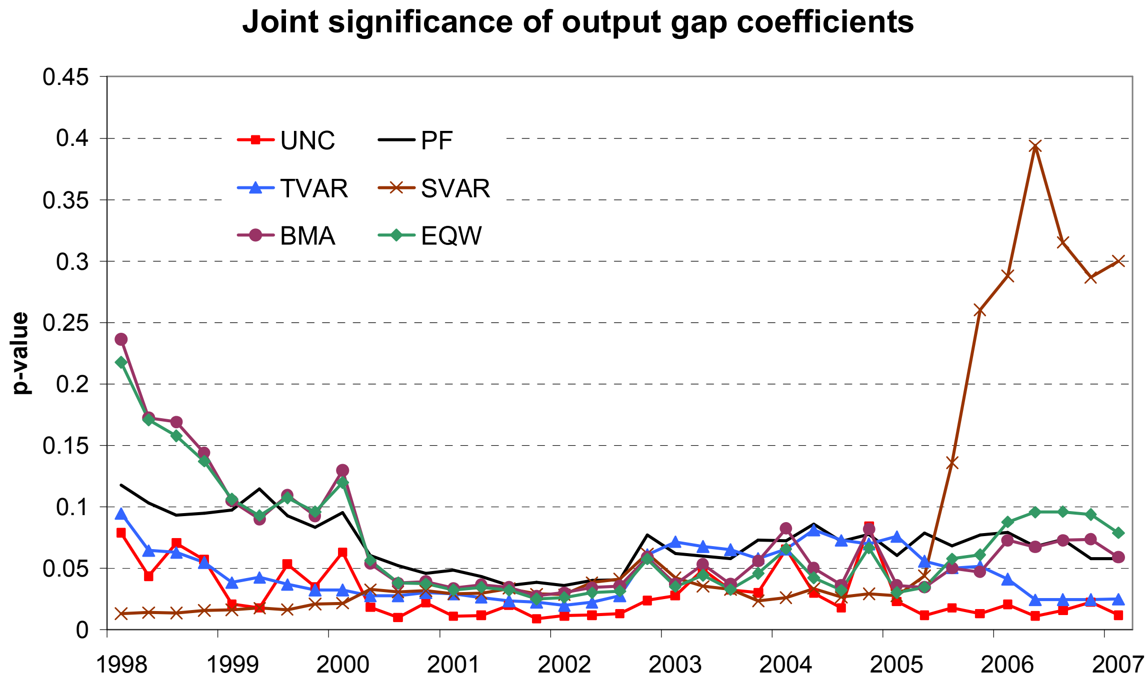
with A = *BMA, EQW.*

In order to have some benchmarks for comparison, we also estimate (i) a simple autoregressive model of order four for inflation (AR henceforth) and (ii) a Phillips curve using, as an explanatory variable, the q-o-q rate of growth of real GDP, rather than the output gap (GDP henceforth). Orphanides and Van Norden (2005) shows that the latter model delivers sistematically better forecasts when compared with models based on the output gap estimated in real time and is therefore a suitable benchmark.

All models are estimated by OLS, using 'quasi real-time' data over an expanding window; accordingly, the first regression is estimated on the 1983Q1-1999Q1 sample, the second on the 1983Q1-1999Q2 interval and so on until 2007Q4. In more formal terms, we estimate the Phillips curves over the sequence of 'quasi real-time' output gap series $\left\{y_{t_0+s}^{T} - E\left(\bar{y}_{t_0+s}\middle|^{t_0+r}\Im^{T}\right)\right\}_{s=0}^{r}$, with 1998Q4 $\leq t_0+s \leq$ 2008Q2.

In order to assess the relevance of the competing output gap measures as explanatory variables for predicting inflation, we compute for each sample period an *F*-test of joint significance of the lagged output gap coefficients. The evolution of the *F*-tests over time confirms that all output gap measures have non-negligible explanatory power in most estimation windows (see Fig. 6): for the UNC, TVAR and SVAR models the *F*-tests rejects the null hypothesis of no significance at the 5% level most of the times, while both the PF output gap and those computed by averaging over all methods are significant at the 10% level for all samples from 2000Q2 onwards. As the chart shows, the null hypothesis is in general rejected; when it is not, as in the case of the SVAR indicator, this is arguably due to outlying observations, rather than to regime shifts. There is no evidence that the impact of economic slackness on inflation has decreased over time.

Figure 6



4.2 Forecast comparison

The predictive content of the output gap measures is assessed 1, 4 and 8 steps ahead, finding in all cases strong supporting evidence. Table 9 reports the average Theil's *U* statistics[25] for the Phillips curve equations (2)-(6), as well as for the benchmark specifications. In addition we show the performance of a forecast obtained by combining the individual forecasts of the four methods, using both predictive weights (fBMA) and equal weights (fEQW). Since the forecasting accuracy of each model is assessed in 'quasi real-time', each entry in the table is in fact the average of the U statistics obtained on a set of 36 regressions, estimated using observations up to, respectively, 1999Q1, 1999Q2, …, 2007Q4. Among single-method indicators, the UNC model seems to be the best-performing specification at all forecast horizons, with the TVAR model raking second and the SVAR

[25] The Theil's U statistics is the ratio of the RMSE of the specification being tested and that of a random-walk model.

ranking last; both weighted forecasts and forecasts delivered by Phillips curves using combined output gap measures generally perform better than single-method indicators (with the exception of the UNC and TVAR indicators at the 8-steps ahead horizon). Besides, all specifications overachieve compared with both the AR and GDP models, especially at the shorter forecast horizons.

Table 9

Theil's U statistics for the Phillips curve

(time range: 1999.1 - 2007.4)

	1 step ahead	4 steps ahead	8 steps ahead
Unobserved components	0.67	0.79	0.81
Production function	0.69	0.83	0.85
Structural VAR	0.70	0.83	0.89
TVAR	0.70	0.80	0.81
BMA	0.65	0.79	0.82
EQW	0.65	0.79	0.82
fBMA	0.68	0.80	0.82
fEQW	0.68	0.80	0.82
GDP	0.75	0.85	0.86
AR	0.72	0.80	0.85

Note: ratios of RMSE to those implied by a random walk forecast.

While most of the differences in RMSEs are not statistically significant, combining forecasts can be shown to be potentially relevant in improving the forecasting performance of the Phillips curve. Table 10 reports the results of a battery of pairwise forecast encompassing tests as suggested by Harvey *et al.* (1998), computed for the Phillips curves based on the four original output gaps and for the benchmark models. A forecasting model is said to encompass a competing alternative if the latter contains no additional information that can be used to enhance predictive accuracy. Let's e_{jt}, $j=1,2$, be the forecast error of model *j*; Harvey *et al.* write the null hypothesis of forecast encompassing as $H_0{:}E(f_t)=0$, where $f_t = e_{1t}(e_{1t} - e_{2t})$, and construct a t-test based on the statistics $\bar{f} = \frac{1}{H}\sum_{t=1}^{H} f_t$, where H is the forecast horizon: model 1 is said to (forecast) encompass model 2 if the null hypothesis is not rejected. Table 10 must be read as follows: if the cell corresponding to model A (in row) and model B (in column) reports a significant test statistics, a combination between the forecasts delivered by the two models improves predictive accuracy with respect to model A; if instead the cell is empty, model A encompasses model B.

Table 10

Forecast Encompassing Tests

			1 step ahead			
Method	**UNC**	**PF**	**SVAR**	**TVAR**	**GDP**	**AR**
Unobserved components				**		
Production function						
Structural VAR	*					
TVAR	***					
GDP	***	***	**	***		*
AR	**	*				
			4 steps ahead			
Method	**UNC**	**PF**	**SVAR**	**TVAR**	**GDP**	**AR**
Unobserved components					**	
Production function	**			*		**
Structural VAR	*					*
TVAR	*					
GDP	***	*		***		***
AR						
			8 steps ahead			
Method	**UNC**	**PF**	**SVAR**	**TVAR**	**GDP**	**AR**
Unobserved components						
Production function				**		
Structural VAR	**	*		***		**
TVAR						
GDP	*			**		
AR						

Note: *** denotes test statistics corresponding to the 1% significance level, ** to the 5% level, and * to the 10% level.

The table shows that no model is uniformly superior and that each of them contains information that cannot be replicated by the others - maybe because of better predictive properties in particular historical episodes - which gives support forecast combination as a way to enhance efficiency.

In Table 11 the results of a battery of Diebold-Mariano tests of equal predictive accuracy between different models are presented:[26] whenever a model listed in a row outperforms significantly one of the models listed in columns, the corresponding cell is marked with one or more asterisks, depending on the significance level (10, 5, 1%).

At the 1-step-ahead horizon the UNC model delivers a better forecasting performance than the TVAR and GDP counterparts, the latter being outperformed by the PF model as well; at the 4-step-ahead horizon the UNC estimator still improves upon the GDP's, which scores quite poorly also compared with the TVAR and AR specifications. At the 8-step-ahead horizons no method clearly stands out, as they all exhibit similar predictive accuracy.

It can be easily noted that all forecasts implying some form of averaging are generally more accurate than those delivered by the underlying models, regardless of the weighting scheme. The most accurate predictions are those computed using as a regressor the average output gaps: at the one-step-ahead horizon, they improve upon simple forecast combinations, while at longer horizons they are not statistically different.

This finding is not surprising: though the reasons underlying the success of simple combination schemes are poorly understood, there is ample evidence that their effectiveness is wide ranging. Numerous arguments in favour of using forecast combinations have been advanced in the literature. First, it is convenient to pool forecasts rather than the information sets underlying them, because in so doing one reduces the computation burden and saves degrees of freedom. Second, individual forecasts may be heterogeneously affected by structural breaks and may adapt at different speed to the new regime: it is plausible that, on average, combinations of forecasts from models with different degrees of adaptability will outperform predictions from individual models. Third, since no model is true and all are, at best, local approximations, it is implausible that the same model outperforms all

[26] The Diebold-Mariano tests shown in Table 11 are based upon differences in squared forecast errors; using absolute errors, the findings are nearly identical, the only test delivering a significant difference being that between the TVAR and UNC models (in favour of the latter) at the 1 step-ahead horizon.

others at all points in time: combining forecasts can be viewed as a way to get results that are more robust against misspecification bias and measurement errors.

Table 11

Pairwise Diebold-Mariano tests of equal predictive accuracy

1 step ahead

Method	UNC	PF	TVAR	VAR	BMA	EQW	fBMA	fEQW	GDP	AR
Unobserved components			**						*	
Production function									*	
Structural VAR										
TVAR										
BMA		***	**				*		***	
EQW		***	**				*	*	***	
fBMA			**						**	
fEQW			**						**	
GDP										
AR										

4 steps ahead

Method	UNC	PF	TVAR	VAR	BMA	EQW	fBMA	fEQW	GDP	AR
Unobserved components		*							***	
Production function										
Structural VAR									**	
TVAR										
BMA		**							**	
EQW		**							**	
fBMA		*							***	
fEQW		*							***	
GDP										
AR									*	

8 steps ahead

Method	UNC	PF	TVAR	VAR	BMA	EQW	fBMA	fEQW	GDP	AR
Unobserved components										
Production function										
Structural VAR										
TVAR										
BMA										
EQW										
fBMA										
fEQW				*						
GDP										
AR										

Note: *** denotes a difference in predictive accuracy significant at the 1% level, ** a difference significant at the 5% level, * a difference significant at the 10% level.

5. Summary and conclusions

Due to unobservability, the estimate of potential output (and accordingly of the output gap) is plagued by difficulties. The performance of competing estimators can be appraised only indirectly, by testing their compliance with business cycle indicators and their accuracy in predicting output or inflation. Besides, the period-by-period revisions of the statistical sources from which potential output is filtered, reduce the real-time utility for policy purposes of the estimates. Nonetheless, the importance of measuring supply capacity to appraise medium-run growth prospects and debt sustainability on the one hand, and the need to monitor economic slackness in order to preserve price stability on the other, provide a challenge to search for better and better techniques.

We consider four different methods to compute potential output: two based on predominantly statistical methods (TVAR and UNC), one using a production function approach (PF), one relying on structural VAR modelling (SVAR); we also evaluate the performance of linear combinations of the previous 4 estimates. We focus mostly on univariate methods and, as a result, our analysis is not well suited to study how potential output varies when the economic or institutional environment changes.

Using Bayesian model averaging to rank the output gap indicators, we find that the UNC and PF measures are superior to the TVAR and SVAR methods; no method receives a zero weight, suggesting that all indicators bring about relevant pieces of information. This interpretation is confirmed by the analysis carried out in section 3, which shows that no indicator is uniformly superior at tracking the business cycle. For this reason, we choose to merge all indicators in a synthetic measure, which improve along several dimensions over single indicators.

The main finding documented in this paper are the following:

1. all methods generate gaps that properly describe the Italian business cycle and are consistent with the output gap measures published by international organisations, the OECD in particular, the only institution that provides quarterly estimates of the degree of economic slackness. The performance of the 4 indicators is not entirely satisfactory in the early 1980s, due to the beginning-of-sample problem, but improves substantially in the following quarters. All competing approaches have no

mean bias (the TVAR gap excluded), exhibit volatility that resemble that of other similar indicators (with the exception of the SVAR output gap) and detect with an acceptable degree of precision the peaks and troughs of the business cycles;

2. the revisions to output gap estimates, implemented as new observations become available under the maintained assumption that historical data do not change, are non-negligible, though not larger than those shown by other indicators. The revisions show high serial autocorrelation, are at times large (especially for the TVAR indicator) and tend to be positively biased (for three gap output measures out of four). The 'quasi real-time' estimates are nonetheless informative, being strongly correlated with the 'final' outturn and only rarely having an opposite sign with respect to the 'final' estimate. Most of the revisions takes place in the 1st year;
3. unlike the evidence reported in most of the literature, we found that 'quasi real-time' estimates of the output gap help predicting inflation and improve upon the performance of simple benchmark specifications, such as the random-walk or the autoregressive model: F-tests of joint significance of the lagged output gap coefficients exhibit a very low p-value, regardless of whether one focuses on the whole sample periods or on shorter time frames. Besides, Phillips curve equations featuring the output gap outperform by far specifications using the growth rate of GDP as a regressor. The predictive performance is higher for shorter-forecast horizons (1 and 4 step-ahead), but remains satisfactory also when the time frame is extended to 8 quarters. Tough the Diebold-Mariano tests show that no indicator is uniformly more efficient, the UNC estimator stands out compared with either the benchmark models or with the TVAR, PF and SVAR indicators;
4. output gaps obtained as linear combinations of the 4 basic indicators turn out to be (i) less sensitive to revisions, (ii) at least as good at tracking business cycle fluctuations and (iii) better inflation predictors. We use two different aggregation procedures to combine estimates: a) equal weights for all models and b) weights proportional to the predictive posterior density and adjusted to take into account the correlation across in-sample residuals. Regardless of the choice of the weighting scheme, the combinations improve in many respects upon the single indicators: in

the majority of cases, they are less biased, less volatile and have smaller ranges. Both the mean and the amplitude of the revisions are much smaller than those of the constituent series, though their persistence is by and large the same; moreover, the number of quarters in which the 'quasi real-time' estimate has opposite sign of the 'final' one is much smaller. The combined measures detect the peaks and troughs with sharp precision, showing neither a systematic lead nor a systematic lag at turning points; the percentage of inconsistent signals is non-negligible, in particular for the equally-weighted output gap;

5. the specifics of the weighting scheme do not affect the performance of the combined output gap measures. Additional experiments made with alternative weights confirm that pooled indicators ensure superior outcomes in terms of either (i) minimising the impact of revisions, or (ii) detecting turning points or (iii) predicting inflation. Apparently, averaging removes the noise inherent in each estimate and enhances performance. Accordingly, on efficiency grounds, the equally-weighted combination is to be preferred.

References

Altissimo F., Cristadoro R., Forni M., Lippi M. and G. Veronese (2007), *New Eurocoin: Tracking economic growth in real time*, Banca d'Italia, Temi di discussione n. 631.

Altissimo F., Marchetti D. J. and G. P. Oneto (2000), *The Italian Business Cycle: Coincident and Leading Indicators and Some Stylized Facts*, Banca d'Italia, Temi di discussione n. 377.

Bassanetti A., Döpke J., Torrini R. and R. Zizza (2006), *Capital, labour and productivity: what role do they play in the potential GDP weakness of France, Germany and Italy?*, in Convergence or Divergence in Europe?, O. De Bandt, Hermann H. and G. Parigi eds., Springer

Benati L., *Drift and Breaks in Labour Productivity*, Journal of Economic Dynamics and Control, 31, January 2007, pp. 2847-2877.

Blanchard O. and D. Quah (1989), *The Dynamic Effects of Aggregate Demand and Supply Disturbances*, The American Economic Review, vol. 79 no. (4).

Bry G. and C. Boschan (1971), *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, NBER, Technical Paper, no. 20.

Casella, G. and E. I. George (1992), *Explaining the Gibbs Sampler*, The American Statistician, Vol. 46, No. 3., pp. 167-174.

Chib, S. and E. Greenberg (1995), *Understanding the Metropolis-Hastings algorithm*, The American Statistician, Vol. 49, No. 4., pp. 327-335

Christiano L.J. and T.J. Fitzgerald (2003), *The Band Pass Filter*, International Economic Review, vol. 44, No. 2

Denis C., Grenouilleau D. McMorrow K. and W. Roeger (2006), *Calculating potential growth rates and output gaps – A revised production function approach*, European Commission Economic Paper n. 247

De Masi P.R. (1997), *IMF estimates of potential output: theory and practice*, IMF Working Paper n. WP/97/177

Eklund J. and S. Karlsson (2007), *Forecast Combination and Model Averaging Using Predictive Measures*, Econometric Reviews, vol. 26 no. 2, pp. 329-363.

Gaiotti E. (2008), *Has globalisation changed the Phillips curve? Firm-level evidence on the effect of activity on prices*, Banca d'Italia, Temi di Discussione, no. 676.

Giorno C., Richardson P., Roseveare D. and P. van den Noord (1995), *Estimating potential output gaps and structural budget balances*, OECD Economic Department Working Paper n. 152

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter* Cambridge University Press, Cambridge.

Harvey, A. C. and A. Jaeger (1993). *Detrending, Stylized Facts and the Business Cycle*, Journal of Applied Econometrics 8, pp. 231-47.

Harvey, C. Trimbur, T. M., and H. K. Van Dijk, *Trends and cycles in economic time series: A Bayesian approach*, Journal of Econometrics, vol. 140 (2), October 2007, pp. 618-649.

Harvey D.I., S.J. Leybourne and P. Newbold (1998), *Tests for Forecast Encompassing*, Journal of Business & Economic Statistics, vol. 16, no. 2, pp. 254-259.

IMF (2006), *How globalisation affected inflation?*, IMF World Economic Outlook, April.

Koske I. and N. Pain (2008), *The Usefulness of Output Gaps for Policy Analysis*, OECD, Economics Department Working Paper, no. 621.

Mönch E. and H. Uhlig (2004), *Towards a Monthly Business Cycle Chronology for the Euro Area*, Centre for Economic Policy Research, Discussion Paper No. 4377.

Orphanides A. and S. van Norden (2002), *The Unreliability of Output-Gap Estimates in Real Time*, The Review of Economics and Statistics, November 2002, 84(4), pp. 569-583.

Orphanides A. and S. van Norden (2005), *The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time*, Journal of Money, Credit and Banking, 2005, vol. 37, issue 3, pp. 583-601.

Pain N., I. Koske and M. Sollie (2006), *Globalisation and inflation in the OECD economies*, OECD, Economics Department Working Paper, no. 524.

Stock, J. and Watson, M. *Asymptotically Median Unbiased Estimation of Coefficient Variance in a Time Varying Parameter Model*, Journal of the American Statistical Association, vol. 93, No. 441, march 1998, pp 349-358.

Tosetto E. (2008), *Revisions of Quarterly Output Gap Estimates for 15 OECD Member Countries*, OECD, mimeo.

Watson M.W. (2007), *How Accurate Are Real-Time Estimates of Output Trends and Gaps?*, Federal Reserve Bank of Richmond Economic Quarterly, vol. 93, no. 2.

Appendix 1: methods for estimating potential output

A PRODUCTION FUNCTION APPROACH FOR POTENTIAL OUTPUT

International institutions have widely adopted the production function method in order to estimate the level of potential output; among them, the European Commission, the International Monetary Fund and the OECD.[27] Following closely the specification set out by the European Commission, we resort to the same approach to estimate potential output and the output gap for Italy; differently from the above mentioned institutions, however, we employ quarterly data instead of annual time series.

In the production function method, actual GDP (Y) is obtained through a combination of labour (L) and capital (K), both corrected for the degree of capacity utilization (U_L and U_K, respectively), multiplied by the technological level (or total factor productivity, TFP). More formally, adopting a Cobb-Douglas functional form:

$$Y = (U_L \cdot L)^{\alpha} \cdot (U_K K)^{1-\alpha} \cdot TFP$$

Labour input is measured by data on employment (number of employees); capital input is a comprehensive measure which includes spending on structures and equipment. When adopting this functional specification, a few assumptions are made for simplicity, the most important being those of constant returns to scale and perfect competition. Under these assumptions, the output elasticities of labour and capital, represented respectively by α and $(1-\alpha)$, can be easily estimated from the wage share in the National Accounts.[28] TFP is measured by the Solow residual:

$$TFP = Y - (U_L \cdot L)^{\alpha} \cdot (U_K K)^{1-\alpha}$$

Moving from actual to potential GDP (Y*) requires the estimation of the potential use of the production factors (L* and K*) and of the trend level of efficiency (TFP*). Once these estimates are obtained, they can be inserted into equation (1) to get the potential output level.

As regards capital input, its potential use is simply given by the full utilization (U_K=1) of the existing capital stock (K) in the economy; there is no need to smooth this series since it can be interpreted as an indicator of the overall capacity in the country.[29]

As for potential labour input, the first step is to define its maximum level, which is given by the working age population (PWA).[30] Then trend labour force (LF*) is obtained multiplying PWA by the trend participation rate, which in turn is the outcome of the application of a statistical filter to the

---

27 See Denis et al. (2006), Giorno et al. (1995) and De Masi (1997).

28 Specifically, we adopted the average wage share over the sample period.

29 Alternatively, we also tried to correct the capital stock for its degree of utilisation, without obtaining significant improvements on final outcomes.

30 In the following we assume the unavailability of a measure of labour capacity utilization U_L. In this case, this cyclical component of the labour input ends up in the Solow residual and will be filtered out once the TFP is smoothed through a statistical filter.

ratio PR=LF/PWA.[31] Finally, potential labour input is calculated consistently with a non accelerating inflation rate of unemployment (NAIRU): L* = LF*·(1-NAIRU).[32]

Also the trend (i.e. “normal”) level of efficiency of factor inputs (TFP*) is obtained through the statistical filtering of the Solow residual.

Once potential output is calculated, the output gap is given as usual by:

$$OG = (Y/Y^* - 1)\cdot 100$$

Besides its simplicity, transparency is one of the main advantages of the production function approach allowing, through a growth accounting exercise, to break down the dynamics of potential GDP into the contributions coming from the production factors (potential labour and capital) and from the trend level of TFP.

On the other hand, just like all the methods that recur to some sort of statistical filtering, a drawback of this approach is the end-of-sample bias. In fact most of the filters employed to extract the trend component from a time series are based on two-sided weighted moving averages, implying the need to extend the series at the end of the sample through forecasts. As new and updated actual data becomes available, thus replacing previously forecasted figures, filtered series - and therefore potential output - may be revised.

A BAYESIAN APPROACH TO ESTIMATING POTENTIAL OUTPUT

The unobserved component approach first pioneered by Harvey (1989) and Harvey and Jaeger (1993) provides a very flexible method to estimate the trend and cyclical components of GDP. In particular, unobserved component models can accommodate a variety of frameworks (both univariate and multivariate), within which both statistical and economic restrictions can be exploited in order to identify the long run behaviour and the cyclical fluctuations of a macroeconomic time-series, as in Harvey et al. (2007), who specify a multivariate model for output and inflation using the restrictions implied by a Phillips curve. All these frameworks share two key identifying assumptions: (i) the trend component follows a non-stationary process, e.g. a random walk with a (possibly) time-varying drift, and (ii) the cyclical component is a stationary ARMA process. Following this approach, we assume that the Italian GDP can be decomposed as follows:

$$y_t = \mu_t + c_{n,t} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$$

where y_t denotes output, μ_t the trend component and $c_{n,t}$ the cycle, while ε_t is an idiosyncratic shock capturing the high frequency variability of the series.

Trend output follows an integrated process of the second order:

$$\mu_t = \mu_{t-1} + \beta_t + u_t, \qquad u_t \sim N\left(0, \sigma_u^2\right)$$

$$\beta_t = \beta_{t-1} + v_t, \qquad v_t \sim N\left(0, \sigma_v^2\right)$$

[31] We do not resort to the widely used Hodrick-Prescott filter, rather to the filter proposed by Christiano-Fitzgerald that argue in favour of a better performance in real time. However, results would be basically similar should we use the Hodrick-Prescott filter. See Christiano and Fitzgerald (2003).

[32] The NAIRU is obtained through a bivariate unobserved component method that includes unemployment and GDP, embedding Okun’s Law. Details of the method can be found in Bassanetti *et al.* (2006).

where β_t represents the slope of the trend component and u_t and v_t are random shocks. When the slope β_t is deterministic (i.e., when $\sigma_v^2 = 0$), the trend is a simple random walk with drift, while if both $\sigma_v^2 = 0$ and $\sigma_u^2 = 0$, the trend is a deterministic function of time.

For the cyclical component we consider a so-called first order stochastic cycle, evolving according to the following process:

$$\begin{bmatrix} c_t \\ c_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} \begin{bmatrix} c_{t-1} \\ c_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \qquad \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \sim N\left(0, \sigma_\kappa^2 I_2\right)$$

where the parameter λ denotes the frequency of the fluctuations, ρ is a damping factor and κ_t and κ_t^* are stochastic disturbances. For values of ρ inside the unit circle, the cycle is stationary and it can be represented as an ARMA process. If λ is such that $0 < \lambda < \pi$, the spectrum of $c_{1,t}$ displays a peak around frequency λ, thus implying (pseudo) cyclical behaviour.

The model defined by the previous equations can be cast in state-space form, with the first equation representing the observation equation and the remaining ones the transition equations. For given values of the parameter vector $\theta \equiv [\rho, \lambda, \sigma_\varepsilon^2, \sigma_u^2, \sigma_v^2, \sigma_\kappa^2]$, the extraction of the unknown states μ_t, β_t, c_t, c_t^* can be straightforwardly implemented by a Kalman filter recursion.

When the parameters of the model are unknown, maximum likelihood estimates can be obtained by using e.g. the well-known EM-algorithm. However, the maximisation of the likelihood function can lead to imprecise estimates, given the often poor identification of some of the parameters. In addition, should the likelihood display more than one peak, a maximum likelihood estimator could get trapped in a small region of the parameter space, so delivering fragile or implausible results. The reliability of the estimates could be improved by imposing variance restrictions, which are indeed needed in order to select key properties of the extracted components, as for instance the amplitude and persistence of the cycle and the smoothness of the trend component. However, using exact restrictions may shut out relevant information coming from the data, thus significantly affecting the precision of the estimates. For this reason, we use a Bayesian approach, which enables us to impose stochastic restrictions over the parameter space, thus letting the data speak whenever they contain enough information.

In formal terms, we define a prior distribution $p(\theta)$ for the parameter vector θ and combine the prior with the likelihood $L(\theta \mid Y)$ (with Y representing the sample of actual data), according to Bayes' rule, to get the posterior distribution of the parameters of interest:

$$p(\theta|Y) = \frac{L(\theta|Y)p(\theta)}{\int L(\theta|Y)p(\theta)d\theta} \qquad (1)$$

Our assumptions about prior information follow basically those employed in Harvey et al. (2007). We assume that the parameters in θ are distributed independently of each other, implying that $p(\theta) = p(\rho)p(\lambda)p(\sigma_\varepsilon^2)p(\sigma_u^2)p(\sigma_v^2)p(\sigma_\kappa^2)$. The prior distributions of ρ and λ are informative: specifically, we assume that ρ follows a uniform distribution and must be larger than 0.5 and smaller than 1, while λ is distributed according to a Beta(2,5.1), which implies a mode corresponding to a cycle of 32 quarters. In addition, λ is restricted in order to exclude too high and too low frequencies (up to 12 and over 48 quarters, respectively). For all remaining parameters we choose uninformative priors.

In practice, (1) cannot be directly computed and has to be simulated by Markov Chain Monte Carlo techniques. We use the Gibbs sampler[33] to simulate random draws from the conditional posterior distributions of the variance parameters – namely $p\left(\sigma_\varepsilon^2 \middle| \theta_{-\sigma_\varepsilon^2}, Y\right)$, $p\left(\sigma_u^2 \middle| \theta_{-\sigma_u^2}, Y\right)$, $p\left(\sigma_v^2 \middle| \theta_{-\sigma_v^2}, Y\right)$, $p\left(\sigma_\kappa^2 \middle| \theta_{-\sigma_\kappa^2}, Y\right)$ – and of the parameter measuring the persistence of the cycle – i.e. $p\left(\rho \middle| \theta_{-\rho}, Y\right)$.[34] A conditional posterior distribution for λ cannot be simulated, as this parameter enters nonlinearly in the model; for this reason we resort to a Metropolis step nested into the Gibbs sampling algorithm.[35]

POTENTIAL OUTPUT IN A TIME VARYING PARAMETER FRAMEWORK

An alternative approach of estimating potential output relies upon the use of time-varying-parameter unobserved-component models. Following Benati (2007), the growth rate of GDP (y_t) can be estimated by using a time varying autoregressive specification:

$$y_t = \mu_t + \varphi_{1t} y_{t-1} + \varphi_{2t} y_{t-2} + \ldots + \varphi_{pt} y_{t-p} + u_t$$

$$\begin{bmatrix} \mu_t \\ \varphi_{1t} \\ \varphi_{2t} \\ \varphi_{3t} \\ \ldots \\ \varphi_{pt} \end{bmatrix} = \begin{bmatrix} \mu_{t-1} \\ \varphi_{1t-1} \\ \varphi_{2t-1} \\ \varphi_{3t-1} \\ \ldots \\ \varphi_{pt-1} \end{bmatrix} + \xi_t$$

where: $u_t \sim \mathrm{N}(0, \sigma_\mathrm{u})$, $\xi_t \sim \mathrm{N}(0, \Sigma_\xi)$.

The model can be rewritten more compactly by stacking the constant and the p[36] lagged values of y_t into a single vector Z_t :

$$y_t = \Phi'_t Z_t + u_t$$
$$\Phi_t = \Phi_{t-1} + \xi_t$$

where $Z_t \equiv \begin{bmatrix} y_t & y_{t-1} & \ldots & y_{t-p+1} \end{bmatrix}'$ and $\Phi_t \equiv \begin{bmatrix} \mu_t & \varphi_{1t} & \varphi_{2t} & \varphi_{3t} & \ldots & \varphi_{pt} \end{bmatrix}'$. The model is linear and Gaussian and the unknown parameters (σ_u and the covariance *(p+1)x(p+1)* matrix Σ_ξ) could in principle be estimated by maximum likelihood, using the Kalman filter. When the variance of the random walk component is low compared with that of the noise (σ_u), the maximum likelihood estimator turns out to be biased towards zero. The problem can be circumvented using an alternative estimation method, the Median Unbiased Estimator (MUB), originally developed by Stock and Watson (1998). This method basically consists of inverting the quantiles of the test statistics that the parameter governing the time variation of the coefficients is zero. To understand how this estimation

[33] See Casella and George (1992) for more details on the Gibbs sampler.

[34] Here the notation θ_{-i} denotes the parameters in θ excluding parameter i.

[35] Chib and Greenberg (1995) show that Gibbs sampling is a particular case of the more general Metropolis-Hastings algorithm, to which we resort for the simulation of posterior distributions for which the Gibbs sampler is not usable.

[36] The lag length is determined with conventional Information criteria on the time constant parameters model.

method works take the simple local level model where a series y_t is modelled as a random walk plus noise:

$$y_t = \mu_t + u_t$$
$$\mu_t = \mu_{t-1} + \eta_t$$

where $u_t \sim \mathrm{N}(0,\sigma_u)$ and $\eta_t \sim \mathrm{N}(0,\sigma_\eta)$ and consider the case in which the signal to noise ratio σ_η / σ_u is small. A break test on the mean of y_t is an implicit test of the hypothesis $\sigma_\eta = 0$. Therefore a constant mean model ($y_t = \mu + u_t$) can be estimated via OLS and a break test can be performed on the estimated mean $\hat{\mu}$. This test has a known distribution from which we can compute the p-value associated with the test statistics ($\hat{L}$). A p-value lower than the chosen confidence interval (say lower than 0.05) provides some evidence that $\sigma_\eta > 0$.[37] The variance σ_η can be estimated as follows:

1. define a grid of possible values of σ_η in $(0,\bar{\sigma}_\eta]$; then for each value in the grid (call it $\sigma_\eta^{\,j}$) simulate *N* times (for *N* as large as 10000) the time varying mean model[38];
2. for each replication, perform the break test and save the median of these *N* break tests. Repeat this process for all the $\sigma_\eta^{\,j}$;
3. select as the MUB estimate of σ_η the $\sigma_\eta^{\,j}$ in the grid whose median of the break test statistics is the closest to $\hat{L}$, the value estimated on the real data.

The more complicated case of a time varying autoregressive model sketched above can be simplified by assuming that the covariance matrix of the coefficients Σ_ξ has the form

$$\Sigma_\xi = (Z_t' Z_t)^{-1} \lambda \sigma_u$$

where λ is a scalar parameter that needs to be estimated and σ_u is estimated under the null hypothesis of no time variation in the parameters (i.e. it is the error variance of a constant parameters AR model). Estimation can be performed as in the previous simpler case by simulating the model conditional on different values of $\lambda \in (0,\bar{\lambda}]$ and choosing the one ($\hat{\lambda}$) that reproduces the amount of time variation in the coefficients that is the closest to that found in the data.

Equilibrium growth $\hat{y}_t$ is then defined as:

$$\hat{y}_{t/T} = \frac{\mu_{t,T}}{1 - \varphi_{1,t/T} - \varphi_{2,t/T} - \ldots - \varphi_{p,t/T}}$$

where the time varying parameters are estimated with the Kalman smoother.[39]

[37] Actually, considering the fact that the time invariant model is a *restricted* version of the time variant model, using confidence levels as low as 0.05 is equivalent to giving the restricted model a privileged status. Benati (2007) considers p-values as high as 0.4 as indicating the presence of time variance.

[38] This is done by drawing *T* random shocks from a normal with zero mean and variance $\sigma_h^{\,j}$, *T* random shocks from a normal with zero mean and variance σ_u (where T is the sample size and σ_u has been estimated under the hypothesis of constant parameters) and then building a simulated series using the Kalman filter.

A STRUCTURAL VAR APPROACH Á LA BLANCHARD-QUAH

One of the main drawbacks of the purely statistical or "direct" techniques generally adopted to estimate potential output growth is that they do not allow capturing and identifying the variety of factors that may have affected it over time. Understanding the sources of potential output growth changes - and their sustainability - is indeed important also to analyse the conjunctural situation. One step in this direction is the decomposition of potential output into its main determinants (TFP, labour and capital) i.e. the explicit estimation of a production function. This approach, that has the well known drawback of deriving TFP growth as a residual, can be usefully complemented by a more structural analysis of the sources of potential growth in terms of shocks and propagating mechanisms.

The empirical exercise aims at carrying out this analysis within a structural vector autoregression (sVAR) framework, which explicitly takes into account the institutional settings that characterise the Italian product and labour market, their modifications over time and their interactions with a number of shocks that have hit the economy in the last few decades.

More in detail, the quantitative analysis is focused on developing a suitable framework aimed at disentangling the different sources of shocks that have driven output fluctuations from the effects of structural changes affecting the way product and labour markets operate. We specify a model that allows the identification of four different shocks: i) aggregate demand disturbances, that can be traced back to impulses generated by monetary and fiscal policy; ii) productivity shocks, representing the forces that affect the permanent component of output; iii) labour supply shocks, corresponding to exogenous movements in the labour force due to demography and changes in participation behaviour; and finally iv) changes in the institutional setting of the wage bargaining system. We impose a priori specific restrictions to identify structural components, on the basis of theoretical considerations that distinguish shocks that have permanent effects from those that have only a transitory influence.

After having investigated the effects and the propagation patterns of such shocks, we provide a breakdown of the Italian GDP growth rate into its cyclical and structural (potential) component, by identifying the latter as that part of observed output growth driven by non-demand shocks. This approach allows the derived measure of potential output growth to be broken down further into the effects of various supply side shocks identified in the sVAR.

---

[39] In practice the procedure is a little more complicated as one has to consider the uncertainty surrounding the estimate of $\hat{\lambda}$. The empirical probability distribution of $\hat{\lambda}$ can be computed as $\Phi_{\hat{\lambda}}(\lambda_j) = P(L > \hat{L} \mid \lambda_j)$ where $\hat{L}$ is the break test on the actual data, L is the test on data simulated conditional on the λ_j and $P(L > \hat{L} \mid \lambda_j)$ is the fraction of the *N* break tests performed in the simulation conditional on the λ_j that give a p-value higher than $\hat{L}$. Given this empirical probability distribution $\Phi_{\hat{\lambda}}$, one can draw a λ^i from $\Phi_{\hat{\lambda}}$ and run the Kalman smoother therefore getting an estimate of $\mu^i{}_{t/T}, \varphi^i{}_{1,t/T}, \varphi^i{}_{2,t/T}, \ldots, \varphi^i{}_{p,t/T}$ and compute $\hat{y}^i{}_{t/T}$. Repeating this *R* times one obtains an empirical distribution of the $\hat{y}_{t/T}$ from which one can consider the relevant quantile (for example the median or the 25th and 75th percentiles if one wants some confidence bands). See Benati (2007) for a more complete procedure that allows also for breaks in the variance σ_u .

As said, the method has a number of advantages, namely it allows i) to recover the shocks impinging on the economy; ii) to disentangle them from their propagation and amplification mechanisms working through the functioning of markets; iii) to avoid measurement difficulties related to the estimation of TFP in the standard production function approach. Clearly, it has also shortcomings, the most relevant ones being the over-parametrisation of the reduced form model, that affects the precision of the estimates, and the identifying assumptions, that can be quite controversial especially if the model is a large one.

The model is an extension of the framework set out in Blanchard and Quah (1989), augmented to allow for a richer variety of shocks and in particular with a wage-price block building upon the Layard-Nickell framework. It can be described by the following set of equations:

- an aggregate demand equation, supplemented by the law of motion of the demand and productivity shocks, both modelled as random walks:

$$y_t = \phi(d_t - p_t) + a\vartheta_t \quad \text{with } d_t = d_{t-1} + \varepsilon_t^d \text{ and } \vartheta_t = \vartheta_{t-1} + \varepsilon_t^s$$

- a constant return to scale production function, where capital is substituted out under the assumption that in the long run it is a constant fraction of output:

$$y_t = n_t + \vartheta_t$$

- a price setting equation where firms have market power and set prices on the basis of unit labour costs and unemployment conditions:

$$p_t = w_t - \vartheta_t + \beta u_t$$

- a labour supply equation that depends on the difference between real wages and productivity and on demographic factors:

$$l_t = \alpha E_{t-1}(w_t - p_t - \vartheta_t) + \tau_t \quad \text{with} \quad \tau_t = \tau_{t-1} + \varepsilon_t^l$$

- a wage setting schedule, according to which unions bargain so as to tie real wages to expected productivity increases. Compensations move pro-cyclically and depend also on an exogenous variable that represents wage-push shocks:

$$w_t = E_{t-1}(p_t + \vartheta_t) + k_t - \sigma E_{t-1} u_t \quad \text{with} \quad k_t = \rho k_{t-1} + \varepsilon_t^w$$

- an unemployment equation:

$$u_t = l_t + n_t$$

Under the assumption that wage bargaining shocks have only a temporary effect (i.e. $|\rho|<1$), solving out the model for real wage growth, output, price inflation and the rate of unemployment yields:

$$\begin{bmatrix} \Delta(w_t - p_t) \\ \Delta y_t \\ \Delta p_t \\ u_t \end{bmatrix} = \begin{bmatrix} c_{11}(1) & 0 & 0 & 0 \\ c_{21}(1) & c_{22}(1) & 0 & 0 \\ c_{31}(1) & c_{32}(1) & c_{33}(1) & 0 \\ c_{41}(1) & c_{42}(1) & c_{43}(1) & c_{44}(1) \end{bmatrix} \begin{bmatrix} \varepsilon_t^s \\ \varepsilon_t^l \\ \varepsilon_t^d \\ \varepsilon_t^w \end{bmatrix} + C^*(L) \begin{bmatrix} \Delta\varepsilon_t^s \\ \Delta\varepsilon_t^l \\ \Delta\varepsilon_t^d \\ \Delta\varepsilon_t^w \end{bmatrix}$$

The unemployment rate is determined by the interaction between wage bargaining and price setting and reflects shifts in the bargaining power between unions and firms; besides, nominal shocks as well as productivity and labour supply disturbances may push it temporarily above or below its equilibrium value. The real wage is permanently affected only by technology shocks, while output is ultimately only a function of productivity developments and demographics.

The model is estimated on quarterly Italian data, on a sample covering the period 1970-2008; potential output is obtained by setting the innovations attributed to demand and wage bargaining to zero, i.e. by omitting the effect of disturbances deemed to have only a temporary effect on economic activity.

Appendix 2: model averaging

In this section we provide more details about the model averaging exercise carried out in section 5. Let's denote by x_t^m the output gap estimated by method *m* (*m = UNC, PF, TVAR, SVAR*); we have to find a set of weights for aggregating the output gap measures as follows:

$$x_t = \omega_1 x_t^{UNC} + \omega_2 x_t^{PF} + \omega_3 x_t^{ARV} + \omega_4 x_t^{VAR}$$

Ideally, this combination should be able to achieve a) a performance of the aggregate measure in forecasting inflation which overcomes those of the individual components and b) a stable and reliable indicator of cyclical conditions. For x_t to satisfy these requirements, weights ω_i should be estimated having an eye to both the predictive accuracy of models using individual output gap measures and to the correlation among them. In order to measure these features, we estimate a Phillips curve relating the rate of change of domestic inflation (measured by the private sector value added deflator) to the output gap for each of the methods. In particular, we estimate Phillips curves of the following form:

$$\pi_t = \alpha + \sum_{s=1}^{p} \delta_s \pi_{t-s} + \sum_{j=1}^{L} \gamma_j x_{t-j}^m + \varepsilon_t \qquad (1)$$

where π_t denotes inflation, x_t the output gap, *p* and *L* denote the maximum lag order of inflation and output gap respectively, and ε_t is an i.i.d. disturbance term. Each of the estimated Phillips curves represents a model which will be given a weight on the basis of its forecasting performance and of its correlation with the remaining models. Then, the weights will be used to aggregate the different output gap measures.

The procedure we use to estimate the weights ω_i hinges on Bayesian methods. In general, in a Bayesian framework, the weight of a model *M* can be determined on the basis of its posterior probability $p(M|Y)$. Assuming we have *n* models M_i ($i = 1,\ldots,n$), $p(M_i|Y)$ can be computed by applying the Bayes rule:

$$p(M_i|Y) = \frac{p(Y|M_i)p(M_i)}{\sum_{i=1}^{n} p(Y|M_i)p(M_i)} \qquad (2)$$

where $p(Y|M_i) = \int_{\theta_i} L(Y|\theta_i, M_i)p(\theta_i|Y, M_i)d\theta_i$ is the marginal likelihood of model M_i, θ_i denotes the parameter vector associated with model M_i (with posterior distribution $p(\theta_i \mid M_i, Y)$), $L(Y \mid \theta_i, M_i)$ is the density function of the observed sample *Y*, conditional on model M_i and parameter vector θ_i, and $p(M_i)$ is the prior distribution of model M_i.

Posterior distributions are based on the marginal data density $p(Y \mid M_i)$, which tend to penalize models with poor in-sample fit and can deliver a biased weighting scheme when some models display sensible overfitting. In addition, posterior weights are not designed to take into account the fact that different models may display a certain degree of correlation. An alternative weighting scheme, first proposed by Eklund and Karlsson (2007), allows to overcome the former shortcoming by using predictive out-of-sample densities rather than marginal likelihoods. Following Eklund and Karlsson, we implement the method by partitioning the data sample *Y* in two subsamples: a training sample Y^*, which we use to estimate the parameter vector θ_i, and a hold-out sample Y_f , used to assess the out-of-sample predictive performance of model *i*. The latter is measured by the *predictive posterior density* associated with future observations Y_f, conditional on the data Y^*, which we compute as follows:

$$p(Y_f | Y^*, M_i) = \int_{\theta_i} L(Y_f | Y^*, \theta_i, M_i) p(\theta_i | Y^*, M_i) d\theta_i \qquad (3)$$

The weights ω_i are computed by reformulating (2) in terms of posterior predictive densities:

$$w(M_i | Y_f, Y^*) = \frac{p(Y_f | Y^* M_i) p(M_i)}{\sum_{i=1}^{n} p(Y_f | Y^* M_i) p(M_i)} l(\rho(M_i, \overline{M})), \qquad (4)$$

where, $l(\rho(M_i, \overline{M}))$ is a correction factor depending on the correlation $\rho(M_i, \overline{M})$ between model M_i and a benchmark model $\overline{M}$.

We assume that all models are equally likely a-priori, so that the prior distribution is a constant. For simplicity, we also assume that $l(\rho(M_i, \overline{M})) = \rho(M_i, \overline{M})^{-1}$; $\rho(M_i, \overline{M})$ is computed as $\sigma_{i,\overline{M}}^{-1} = \frac{1}{\sum_t (\hat{\varepsilon}_t^i - \hat{\varepsilon}_t^{\overline{M}})^2}$, where $\hat{\varepsilon}_t^i$ is the estimated residual of the Phillips curve M_i and $\hat{\varepsilon}_t^{\overline{M}}$ is the error term of a simple autoregressive model, labelled $\overline{M}$:

$$\pi_t = \alpha + \sum_{s=1}^{p} \delta_s \pi_{t-s} + \varepsilon_t^{\overline{M}} \qquad (5)$$

The residuals are estimated over a training sample ranging from 1982Q4 to 1986Q4.

The reason underlying the choice of a correction factor $l(\rho(M_i, \overline{M})) = \sigma_{i,\overline{M}}$ is the following: if all output gaps measures provide useful information for inflation, then equation (5) is misspecified as it omits relevant variables; therefore, its residual, $\varepsilon_t^{\overline{M}}$, should be correlated with the omitted variables (the output gaps), or, in other words, be a function of the average output gap x_t. For the same reason, if x_t^i is uncorrelated with x_t^j , $j \neq i$, the residual ε_t^i should be a function of output gaps x_t^j , $j \neq i$ (and of the average gap x_t as well) and it should be correlated with $\varepsilon_t^{\overline{M}}$. Therefore, the smaller the correlation of x_t^i with other output gaps, the larger $\sigma_{i,AR}$. In the limit, if x_t^i is perfectly correlated with other output gaps, $\sigma_{i,AR} = 0$, which implies $w(M_i | Y_f, Y^*) = 0$.

For the computation of predictive densities in (3) we follow Eklund and Karlsson (2007). The prior distributions of the model parameters are calibrated on the basis of an OLS estimate over a training sample 1982Q4 1986Q4, while the posterior is estimated over the period 1987Q1-1993Q4, so that the resulting posterior distributions are of the form $p(\theta_i | Y^*, M_i) \propto N(\hat{\theta}_i, V_i)$, with $\hat{\theta}_i$ being the OLS estimate over the period 1982Q4-1993Q4 and V_i being proportional to the OLS estimated dispersion of the parameters. The hold-out sample over which the predictive density is evaluated ranges from 1994Q1 to 1998Q4.

RECENTLY PUBLISHED "TEMI" (*)

N. 749 – *An analysis of the determinants of credit default swap spread changes before and during the subprime financial turmoil*, by Antonio Di Cesare and Giovanni Guazzarotti (March 2010).

N. 750 – *Estimating DSGE models with unknown data persistence*, by Gianluca Moretti and Giulio Nicoletti (March 2010).

N. 751 – *Down the non-linear road from oil to consumer energy prices: no much asymmetry along the way*, by Fabrizio Venditti (March 2010).

N. 752 – *Information technology and banking organization*, by Sauro Mocetti, Marcello Pagnini and Enrico Sette (March 2010).

N. 753 – *The pricing of government-guaranteed bank bonds*, by Aviram Levy and Andrea Zaghini (March 2010).

N. 754 – *Misure di valore aggiunto per le scuole superiori italiane: i problemi esistenti e alcune prime evidenze*, by Piero Cipollone, Pasqualino Montanaro and Paolo Sestito (March 2010).

N. 755 – *Asset-based measurement of poverty*, by Andrea Brandolini, Silvia Magri and Timothy M. Smeeding (March 2010).

N. 756 – *Credit supply, flight to quality and evergreening: an analysis of bank-firm relationships after Lehman*, by Ugo Albertazzi and Domenico J. Marchetti (April 2010).

N. 757 – *A note on rationalizability and restrictions on belief*, by Giuseppe Cappelletti (April 2010).

N. 758 – *A non-parametric model-based approach to uncertainty and risk analysis of macroeconomic forecasts*, by Claudia Miani and Stefano Siviero (April 2010).

N. 759 – *Financial innovation and risk: the role of information*, by Roberto Piazza (June 2010).

N. 760 – *Switching costs in local credit markets*, by Guglielmo Barone, Roberto Felici and Marcello Pagnini (June 2010).

N. 761 – *The determinants of teacher mobility. Evidence from a panel of Italian teachers*, by Gianna Barbieri, Claudio Rossetti and Paolo Sestito (June 2010).

N. 762 – *A public guarantee of a minimum return to defined contribution pension scheme members*, by Giuseppe Grande and Ignazio Visco (June 2010).

N. 763 – *Debt restructuring and the role of lending technologies*, by Giacinto Micucci and Paola Rossi (June 2010).

N. 764 – *Disentangling demand and supply in credit developments: a survey-based analysis for Italy*, by Paolo Del Giovane, Ginette Eramo and Andrea Nobili (June 2010).

N. 765 – *Information uncertainty and the reaction of stock prices to news*, by Paolo Angelini and Giovanni Guazzarotti (July 2010).

N. 766 – *With a little help from abroad: the effect of low-skilled immigration on the female labor supply*, by Guglielmo Barone and Sauro Mocetti (July 2010).

N. 767 – *Real time forecasts of inflation: the role of financial variables*, by Libero Monteforte and Gianluca Moretti (July 2010).

N. 768 – *The effect of age on portfolio choices: evidence from an Italian pension fund*, by Giuseppe G. L. Cappelletti, Giovanni Guazzarotti and Pietro Tommasino (July 2010).

N. 769 – *Does investing abroad reduce domestic activity? Evidence from Italian manufacturing firms*, by Raffaello Bronzini (July 2010).

N. 770 – *The EAGLE. A model for policy analysis of macroeconomics interdependence in the euro area*, by Sandra Gomes, Pascal Jacquinot and Massimiliano Pisani (July 2010).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2008

P. ANGELINI, *Liquidity and announcement effects in the euro area*, Giornale degli Economisti e Annali di Economia, v. 67, 1, pp. 1-20, **TD No. 451 (October 2002).**

P. ANGELINI, P. DEL GIOVANE, S. SIVIERO and D. TERLIZZESE, *Monetary policy in a monetary union: What role for regional information?*, International Journal of Central Banking, v. 4, 3, pp. 1-28, **TD No. 457 (December 2002).**

F. SCHIVARDI and R. TORRINI, *Identifying the effects of firing restrictions through size-contingent Differences in regulation*, Labour Economics, v. 15, 3, pp. 482-511, **TD No. 504 (June 2004).**

L. GUISO and M. PAIELLA,, *Risk aversion, wealth and background risk*, Journal of the European Economic Association, v. 6, 6, pp. 1109-1150, **TD No. 483 (September 2003).**

C. BIANCOTTI, G. D'ALESSIO and A. NERI, *Measurement errors in the Bank of Italy's survey of household income and wealth*, Review of Income and Wealth, v. 54, 3, pp. 466-493, **TD No. 520 (October 2004).**

S. MOMIGLIANO, J. HENRY and P. HERNÁNDEZ DE COS, *The impact of government budget on prices: Evidence from macroeconometric models*, Journal of Policy Modelling, v. 30, 1, pp. 123-143 **TD No. 523 (October 2004).**

L. GAMBACORTA, *How do banks set interest rates?,* European Economic Review, v. 52, 5, pp. 792-819, **TD No. 542 (February 2005).**

P. ANGELINI and A. GENERALE, *On the evolution of firm size distributions*, American Economic Review, v. 98, 1, pp. 426-438, **TD No. 549 (June 2005).**

R. FELICI and M. PAGNINI, *Distance, bank heterogeneity and entry in local banking markets*, The Journal of Industrial Economics, v. 56, 3, pp. 500-534, **No. 557 (June 2005).**

S. DI ADDARIO and E. PATACCHINI, *Wages and the city. Evidence from Italy*, Labour Economics, v.15, 5, pp. 1040-1061, **TD No. 570 (January 2006).**

S. SCALIA, *Is foreign exchange intervention effective?,* Journal of International Money and Finance, v. 27, 4, pp. 529-546, **TD No. 579 (February 2006).**

M. PERICOLI and M. TABOGA, *Canonical term-structure models with observable factors and the dynamics of bond risk premia*, Journal of Money, Credit and Banking, v. 40, 7, pp. 1471-88, **TD No. 580 (February 2006).**

E. VIVIANO, *Entry regulations and labour market outcomes. Evidence from the Italian retail trade sector*, Labour Economics, v. 15, 6, pp. 1200-1222, **TD No. 594 (May 2006).**

S. FEDERICO and G. A. MINERVA, *Outward FDI and local employment growth in Italy*, Review of World Economics, Journal of Money, Credit and Banking, v. 144, 2, pp. 295-324, **TD No. 613 (February 2007).**

F. BUSETTI and A. HARVEY, *Testing for trend*, Econometric Theory, v. 24, 1, pp. 72-87, **TD No. 614 (February 2007).**

V. CESTARI, P. DEL GIOVANE and C. ROSSI-ARNAUD, *Memory for prices and the Euro cash changeover: an analysis for cinema prices in Italy*, In P. Del Giovane e R. Sabbatini (eds.), The Euro Inflation and Consumers' Perceptions. Lessons from Italy, Berlin-Heidelberg, Springer, **TD No. 619 (February 2007).**

B. H. HALL, F. LOTTI and J. MAIRESSE, *Employment, innovation and productivity: evidence from Italian manufacturing microdata*, Industrial and Corporate Change, v. 17, 4, pp. 813-839, **TD No. 622 (April 2007).**

J. SOUSA and A. ZAGHINI, *Monetary policy shocks in the Euro Area and global liquidity spillovers,* International Journal of Finance and Economics, v.13, 3, pp. 205-218, **TD No. 629 (June 2007).**

M. DEL GATTO, GIANMARCO I. P. OTTAVIANO and M. PAGNINI, *Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms,* Journal of Regional Science, v. 48, 1, pp. 97-129, **TD No. 635 (June 2007).**

P. DEL GIOVANE, S. FABIANI and R. SABBATINI, *What's behind "inflation perceptions"? A survey-based analysis of Italian consumers,* in P. Del Giovane e R. Sabbatini (eds.), *The Euro Inflation and Consumers' Perceptions. Lessons from Italy*, Berlin-Heidelberg, Springer, **TD No. 655 (January 2008).**

R. BRONZINI, G. DE BLASIO, G. PELLEGRINI and A. SCOGNAMIGLIO, *La valutazione del credito d'imposta per gli investimenti,* Rivista di politica economica, v. 98, 4, pp. 79-112, **TD No. 661 (April 2008).**

B. BORTOLOTTI, and P. PINOTTI, *Delayed privatization,* Public Choice, v. 136, 3-4, pp. 331-351, **TD No. 663 (April 2008).**

R. BONCI and F. COLUMBA, *Monetary policy effects: New evidence from the Italian flow of funds,* Applied Economics , v. 40, 21, pp. 2803-2818, **TD No. 678 (June 2008).**

M. CUCCULELLI, and G. MICUCCI, *Family Succession and firm performance: evidence from Italian family firms,* Journal of Corporate Finance, v. 14, 1, pp. 17-31, **TD No. 680 (June 2008).**

A. SILVESTRINI and D. VEREDAS, *Temporal aggregation of univariate and multivariate time series models: a survey,* Journal of Economic Surveys, v. 22, 3, pp. 458-497, **TD No. 685 (August 2008).**

2009

F. PANETTA, F. SCHIVARDI and M. SHUM, *Do mergers improve information? Evidence from the loan market*, Journal of Money, Credit, and Banking, v. 41, 4, pp. 673-709, **TD No. 521 (October 2004).**

M. BUGAMELLI and F. PATERNÒ, *Do workers' remittances reduce the probability of current account reversals?,* World Development, v. 37, 12, pp. 1821-1838, **TD No. 573 (January 2006).**

P. PAGANO and M. PISANI, *Risk-adjusted forecasts of oil prices*, The B.E. Journal of Macroeconomics, v. 9, 1, Article 24, **TD No. 585 (March 2006).**

M. PERICOLI and M. SBRACIA, *The CAPM and the risk appetite index: theoretical differences, empirical similarities, and implementation problems,* International Finance, v. 12, 2, pp. 123-150, **TD No. 586 (March 2006).**

U. ALBERTAZZI and L. GAMBACORTA, *Bank profitability and the business cycle,* Journal of Financial Stability, v. 5, 4, pp. 393-409, **TD No. 601 (September 2006).**

S. MAGRI, *The financing of small innovative firms: the Italian case*, Economics of Innovation and New Technology, v. 18, 2, pp. 181-204, **TD No. 640 (September 2007).**

V. DI GIACINTO and G. MICUCCI, *The producer service sector in Italy: long-term growth and its local determinants,* Spatial Economic Analysis, Vol. 4, No. 4, pp. 391-425, **TD No. 643 (September 2007).**

F. LORENZO, L. MONTEFORTE and L. SESSA, *The general equilibrium effects of fiscal policy: estimates for the euro area*, Journal of Public Economics, v. 93, 3-4, pp. 559-585, **TD No. 652 (November 2007).**

R. GOLINELLI and S. MOMIGLIANO, *The Cyclical Reaction of Fiscal Policies in the Euro Area. A Critical Survey of Empirical Research*, Fiscal Studies, v. 30, 1, pp. 39-72, **TD No. 654 (January 2008).**

P. DEL GIOVANE, S. FABIANI and R. SABBATINI, *What's behind "Inflation Perceptions"? A survey-based analysis of Italian consumers,* Giornale degli Economisti e Annali di Economia, v. 68, 1, pp. 25-52, **TD No. 655 (January 2008).**

F. MACCHERONI, M. MARINACCI, A. RUSTICHINI and M. TABOGA, *Portfolio selection with monotone mean-variance preferences,* Mathematical Finance, v. 19, 3, pp. 487-521, **TD No. 664 (April 2008).**

M. AFFINITO and M. PIAZZA, *What are borders made of? An analysis of barriers to European banking integration*, in P. Alessandrini, M. Fratianni and A. Zazzaro (eds.): The Changing Geography of Banking and Finance, Dordrecht Heidelberg London New York, Springer, **TD No. 666 (April 2008).**

A. BRANDOLINI, *On applying synthetic indices of multidimensional well-being: health and income inequalities in France, Germany, Italy, and the United Kingdom,* in R. Gotoh and P. Dumouchel (eds.), Against Injustice. The New Economics of Amartya Sen, Cambridge, Cambridge University Press, **TD No. 668 (April 2008).**

G. FERRERO and A. NOBILI, *Futures contract rates as monetary policy forecasts,* International Journal of Central Banking, v. 5, 2, pp. 109-145, **TD No. 681 (June 2008).**

P. CASADIO, M. LO CONTE and A. NERI, *Balancing work and family in Italy: the new mothers' employment decisions around childbearing,* in T. Addabbo and G. Solinas (eds.), Non-Standard Employment and Qualità of Work, Physica-Verlag. A Sprinter Company, **TD No. 684 (August 2008).**

L. ARCIERO, C. BIANCOTTI, L. D'AURIZIO and C. IMPENNA, *Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG,* Journal of Artificial Societies and Social Simulation, v. 12, 1, **TD No. 686 (August 2008).**

A. CALZA and A. ZAGHINI, *Nonlinearities in the dynamics of the euro area demand for M1,* Macroeconomic Dynamics, v. 13, 1, pp. 1-19, **TD No. 690 (September 2008).**

L. FRANCESCO and A. SECCHI, *Technological change and the households' demand for currency*, Journal of Monetary Economics, v. 56, 2, pp. 222-230, **TD No. 697 (December 2008).**

G. ASCARI and T. ROPELE, *Trend inflation, taylor principle, and indeterminacy*, Journal of Money, Credit and Banking, v. 41, 8, pp. 1557-1584, **TD No. 708 (May 2007).**

S. COLAROSSI and A. ZAGHINI, *Gradualism, transparency and the improved operational framework: a look at overnight volatility transmission*, International Finance, v. 12, 2, pp. 151-170, **TD No. 710 (May 2009).**

M. BUGAMELLI, F. SCHIVARDI and R. ZIZZA, *The euro and firm restructuring*, in A. Alesina e F. Giavazzi (eds): Europe and the Euro, Chicago, University of Chicago Press, **TD No. 716 (June 2009).**

B. HALL, F. LOTTI and J. MAIRESSE, *Innovation and productivity in SMEs: empirical evidence for Italy*, Small Business Economics, v. 33, 1, pp. 13-33, **TD No. 718 (June 2009).**

2010

A. PRATI and M. SBRACIA, *Uncertainty and currency crises: evidence from survey data*, Journal of Monetary Economics, v, 57, 6, pp. 668-681, **TD No. 446 (July 2002).**

S. MAGRI, *Debt maturity choice of nonpublic Italian firms* , Journal of Money, Credit, and Banking, v.42, 2-3, pp. 443-463, **TD No. 574 (January 2006).**

R. BRONZINI and P. PISELLI, *Determinants of long-run regional productivity with geographical spillovers: the role of R&D, human capital and public infrastructure,* Regional Science and Urban Economics, v. 39, 2, pp.187-199, **TD No. 597 (September 2006).**

E. IOSSA and G. PALUMBO, *Over-optimism and lender liability in the consumer credit market*, Oxford Economic Papers, v. 62, 2, pp. 374-394, **TD No. 598 (September 2006).**

S. NERI and A. NOBILI, *The transmission of US monetary policy to the euro area,* International Finance, v. 13, 1, pp. 55-78, **TD No. 606 (December 2006).**

A. CIARLONE, P. PISELLI and G. TREBESCHI, *Emerging Markets' Spreads and Global Financial Conditions*, Journal of International Financial Markets, Institutions & Money, v. 19, 2, pp. 222-239, **TD No. 637 (June 2007).**

M. IACOVIELLO and S. NERI, *Housing market spillovers: evidence from an estimated DSGE model,* American Economic Journal: Macroeconomics, v. 2, 2, pp. 125-164, **TD No. 659 (January 2008).**

A. ACCETTURO, *Agglomeration and growth: the effects of commuting costs,* Papers in Regional Science, v. 89, 1, pp. 173-190, **TD No. 688 (September 2008).**

A. B. ATKINSON and A. BRANDOLINI, *On analysing the world distribution of income*, World Bank Economic Review , v. 24, 1 , pp. 1-37, **TD No. 701 (January 2009).**

R. CAPPARIELLO and R. ZIZZA, *Dropping the Books and Working Off the Books,* Labour, v. 24, 2, pp. 139-162 ,**TD No. 702 (January 2009).**

V. DI GIACINTO, G. MICUCCI and P. MONTANARO, *Dynamic macroeconomic effects of public capital: evidence from regional Italian data,* Giornale degli economisti e annali di economia, v. 69, 1, pp. 29-66, **TD No. 733 (November 2009).**

F. COLUMBA, L. GAMBACORTA and P. E. MISTRULLI, *Mutual Guarantee institutions and small business finance,* Journal of Financial Stability, v. 6, 1, pp. 45-54, **TD No. 735 (November 2009).**

A. GERALI, S. NERI, L. SESSA and F. M. SIGNORETTI, *Credit and banking in a DSGE model of the Euro Area,* Journal of Money, Credit and Banking, v. 42, 6, pp. 107-141, **TD No. 740 (January 2010).**

S. FEDERICO, *Outsourcing versus integration at home or abroad and firm heterogeneity,* Empirica, v. 37, 1, pp. 47-63, **TD No. 742 (February 2010).**

V. DI GIACINTO, *On vector autoregressive modeling in space and time*, Journal of Geographical Systems, v. 12, 2, pp. 125-154, **TD No. 746 (February 2010).**

A. DI CESARE and G. GUAZZAROTTI, *An analysis of the determinants of credit default swap spread changes before and during the subprime financial turmoil,* Journal of Current Issues in Finance, Business and Economics, v. 3, 4, pp., **TD No. 749 (March 2010).**

A. BRANDOLINI, S. MAGRI and T. M SMEEDING, *Asset-based measurement of poverty,* Journal of Policy Analysis and Management, v. 29, 2 , pp. 267-284, **TD No. 755 (March 2010).**

FORTHCOMING

L. MONTEFORTE and S. SIVIERO, *The Economic Consequences of Euro Area Modelling Shortcuts,* Applied Economics, **TD No. 458 (December 2002).**

M. BUGAMELLI and A. ROSOLIA, *Produttività e concorrenza estera*, Rivista di politica economica, **TD No. 578 (February 2006).**

G. DE BLASIO and G. NUZZO, *Historical traditions of civicness and local economic development,* Journal of Regional Science, **TD No. 591 (May 2006).**

G. FERRERO, A. NOBILI and P. PASSIGLIA, *Assessing Excess Liquidity in the Euro Area: The Role of Sectoral Distribution of Money,* Applied Economics, **TD No. 627 (April 2007).**

Y. ALTUNBAS, L. GAMBACORTA and D. MARQUÉS, *Securitisation and the bank lending channel*, European Economic Review, **TD No. 653 (November 2007).**

F. BALASSONE, F. MAURA and S. ZOTTERI, *Cyclical asymmetry in fiscal variables in the EU*, Empirica, **TD No. 671 (June 2008).**

M. BUGAMELLI and F. PATERNÒ, *Output growth volatility and remittances*, Economica, **TD No. 673 (June 2008).**

F. D'AMURI, O. GIANMARCO I.P. and P. GIOVANNI, *The labor market impact of immigration on the western german labor market in the 1990s,* European Economic Review, **TD No. 687 (August 2008).**

C. NICOLETTI and C. RONDINELLI, *The (mis)specification of discrete duration models with unobserved heterogeneity: a Monte Carlo study,* Journal of Econometrics, **TD No. 705 (March 2009).**

L. FORNI, A. GERALI and M. PISANI, *Macroeconomic effects of greater competition in the service sector: the case of Italy,* Macroeconomic Dynamics, **TD No. 706 (March 2009).**

Y. ALTUNBAS, L. GAMBACORTA, and D. MARQUÉS-IBÁÑEZ, *Bank risk and monetary policy,* Journal of Financial Stability, **TD No. 712 (May 2009).**

L. FORNI, A. GERALI and M. PISANI, *The macroeconomics of fiscal consolidations in euro area countries,* Journal of Economic Dynamics and Control, **TD No. 747 (March 2010).**

A. DI CESARE and G. GUAZZAROTTI, *An analysis of the determinants of credit default swap spread changes before and during the subprime financial turmoil,* in C. V. Karsone (eds.), Finance and Banking Developments, Nova Publishers, New York., **TD No. 749 (March 2010).**