II. An assessment of the relative quality of the EU output gap estimates

This focus section assesses the performance of the EU's production function methodology for quantifying output gaps since its introduction in EU policy surveillance procedures in 2002. In particular, we assess how the methodology has performed compared with that used previously (the Hodrick-Prescott filter) in terms of gauging the euro-area business cycle. We also compare it with the equivalent OECD and IMF methodologies in terms inter alia of stability, real-time reliability and financial crisis performance. The analysis shows that the PF methodology is superior to both the HP filter and the methods used in other international organisations. This vindicates the decision to adopt it for estimating output gaps as the 'commonly agreed' reference method to be used in EU fiscal surveillance procedures. Nevertheless, while it has clearly done well in relative terms since it was first used, the analysis also recognises the extent of the output gap errors made in the pre-crisis period. This stresses the importance of continuing to improve the EU's commonly agreed methodology, with a particular focus on attenuating procyclicality risks in the upswing phase of the business cycle.

II.1. Introduction(25)

Real-time estimates of the output gap (26) are inevitably surrounded by a large element of uncertainty, since potential output is not directly observable and actual GDP is subject to significant *ex post* revisions. Many studies have documented the extent of the uncertainty, highlighting the fact that the sign, as well as the magnitude, of output gaps estimated in real time are subject to large revisions when new information becomes available. (27) Policy-makers are aware of these uncertainties but nevertheless accept that estimates of potential growth and the output gap are indispensable in assessing the cyclical position of the economy and its productive capacity. Accordingly, output gap indicators have been used as an operational surveillance tool in the Stability and Growth Pact (SGP), since its inception in the second half of the 1990s, for calculating indicators such as the structural (i.e. cyclically adjusted) fiscal

balance. The doubt surrounding underlying output gap calculations means that the main focus has been on changes in, rather than the level of, the structural balance.

Initially, a purely statistical detrending method, the Hodrick-Prescott (HP) filter, was used to calculate output gaps in the SGP context. However, following sustained criticism of the HP filter, especially its end-point bias problems, EU policy-makers adopted the production function (PF) approach for surveillance purposes. This is based on a more comprehensive analytical framework and alleviates the risk of end-point biases.

In the aftermath of the financial crisis and following the strengthening of EU policy surveillance procedures, the growing importance of the output gap estimates produced by the PF method for all of the EU's 28 Member States (due to their usage in the calculation of structural budget balances), has been accompanied by sustained criticism from academics, policy think-tanks and policy-makers. The method's relative stability and real-time reliability have been called into question, with criticism focusing on three areas:

- It has been argued that the real-time reliability gains from moving to the PF method have not been worth the **additional complexity** involved (with greater intricacy linked to the growing preference for embedding multivariate estimation approaches in the PF framework). In

---

(24) Section prepared by Kieran Mc Morrow, Rafal Raciborski, Werner Roeger and Valerie Vandermeulen.

(25) Whilst this analysis focusses only on the euro area, the conclusions also apply to the majority of the EU's 28 Member States.

(26) The output gap is defined as the difference between actual and potential GDP.

(27) See Orphanides A. and S. van Norden (2002), 'The unreliability of output gap estimates in real time', *The Review of Economics and Statistics*, Vol.84, Issue 4, pp. 569-583; Nelson E. and K. Nikolov (2003), 'UK inflation in the 1970s and 1980s: the role of output gap mismeasurement', *Journal of Economics and Business*, Vol. 55, Issue 4, pp. 353-370; Cayen J.-P. and S. van Norden (2005): 'The reliability of Canadian output-gap estimates', *The North American Journal of Economics and Finance*, Vol.16, Issue 3, pp. 373-393; Marcellino M. and A. Musso (2011), 'The reliability of real-time estimates of the euro area output gap', *Economic modelling*, Vol.28, Issue 4, pp. 1842-1856; *Deutsche Bundesbank* (2014):'On the reliability of international organisations' estimates of the output gap', *Monthly Report*, Vol.66, No 4, pp. 13-37.

addition, it has been suggested that the HP filter may actually provide more reliable results; [28]

- With the entry into force of the European Fiscal Compact in January 2013 (and its requirement that the structural deficit of the euro-area Member States be less than 0.5 %), the questioning of the **stability** and **real-time reliability** of the output gaps produced by the PF method has intensified considerably. These estimates are regularly compared with (and criticised as being inferior to) those produced by other international organisations such as the OECD and the IMF; and

- Most importantly, it is argued that the EU's method and those of the OECD and the IMF do a particularly poor job in the **upswing phase** of cycles, where most fiscal policy errors occur. This was dramatically demonstrated in the run-up to the financial crisis (2006-2008). The procyclicality evident in this period can be explained only partially by a systematic optimistic bias in output gap methodologies. The scale of the *ex post* output gap revisions for the pre-crisis period points to a more fundamental weakness in the way the methods handle investment in the boom phase of cycles. This point has been raised by Bank of International Settlements (BIS) economists suggesting that the conventional output gap calculation should be supplemented with information on the financial cycle to identify investment booms and so produce 'finance-neutral' output gaps (see Box II.1 for details of the BIS work). [29]

Against this background, we try here to assess the performance of the EU's PF method since its introduction in 2002, using the output gap estimates for the euro area as the focus of the analysis. [30] More specifically, we address two interrelated questions:

- How has the methodology performed, in terms of its **revisions** record, compared with the previous HP filter approach? And

- What is its **real-time reliability** compared with the OECD and IMF methodologies?

II.2. Comparison between the PF methodology and the HP filter approach

EU policy-makers moved from the HP filter to the PF approach in the autumn of 2002. The current section assesses whether this was a prudent decision. More specifically, it compares the **revision** properties of the PF and HP methods since 2002 by:

- looking at the short-term stability of the estimates, i.e. how much they are revised from one forecast to the next;

- comparing the long-term real-time reliability of the methods;

- assessing the performance of the methods during the financial crisis; and

- examining the economic plausibility of the estimates, in particular whether the optimism in 2002 as to the ability of indicators such as capacity utilisation to reduce the cyclicality of trend total factor productivity (TFP) estimates was justified.

Short-term stability of PF and HP estimates

Given their central role in EU fiscal surveillance procedures, the relative stability of output gap estimates is an important input into the policy making process. Large short-term revisions in estimates undermine the credibility of a method, with significant knock-on implications for crucial policy target variables such as the change in the structural fiscal balance.

[28] This latter viewpoint is very much the conclusion of an April 2014 *Bundesbank* analysis which showed, using a HP filter with a smoothing parameter of 6.25 (compared with a parameter of 100 used in the EU's equivalent HP method up to 2002), more reliable output gap estimates than those of the OECD and the IMF. While the procyclicality issues with a HP 6.25 renders it highly problematic for fiscal policy surveillance purposes, nevertheless the *Bundesbank* analysis suggests that, at least for monetary policy purposes, the choice between a HP filter and the PF approach is far from clear-cut.

[29] See: Borio C., P. Disyatat and M. Juselius (2013): 'Rethinking potential output: embedding information about the financial cycle', *BIS Working Papers,* No 404.

[30] As it is universally accepted that output gap uncertainty is a fact of life for all estimation methods and that output gap estimates are inevitably subject to large revisions, the relative revisions performance of the EU's PF method, rather than the absolute size of those revisions, is considered in this focus section.

Box II.1: Finance-neutral output gaps: a useful complement to the EU's production function method?

In a series of papers, Borio et al. (2013, 2014) make a case against the use of inflation as the conventional gauge of the cyclical stance of an economy. Instead, they propose using information about the financial cycle to produce 'finance-neutral' output gaps, which they argue could help in alerting policy makers to the role of asset prices in the dynamics of business cycles. This approach distinguishes itself from more conventional methods not only by its replacement of cyclical indicators (variation of unit labour costs, capacity utilization etc.) with financial indicators (real interest rates, credit growth and property prices), but also by its reliance on a relatively simple, reduced-form econometric model.

Borio et al. apply the following model, which derives from the Hodrick-Prescott (HP) statistical filter, but extend it with a set of financial variables:

(1) $y_t = p_t + c_t$
(2) $p_t = 2p_{t-1} - p_{t-2} + \varepsilon_t^p$
(3) $c_t = \phi c_{t-1} + \gamma z_t + \varepsilon_t^c$

Above y_t is output in t, p_t is potential output, c_t is the output gap, $\varepsilon_t{}^p$ and $\varepsilon_t{}^c$ are white noise processes with zero mean and variances, V^p and V^c, respectively, and γ is a vector of coefficients on a vector of contemporaneous and lagged exogenous financial variables z_t, respectively. Equation (3) allows for the output gap to be an AR(1) process with AR coefficient ϕ. Potential output process (2) is an integrated process of order 2. The vector of exogenous financial variables, z_t, includes credit growth, house price inflation and the real interest rate and their lags. In the estimation some lags and/or variables may be dropped due to a lack of significance.

The finance-neutral approach attempts to capture the effects of asset price booms on the real economy and in turn on public finances. These effects can include weakened financial and supply constraints, for instance through mortgage finance and misallocation of resources, for example towards the construction sector.

However, in its original form, it does not embed any description of the supply-side structure of the economy, in contrast with the production function used in EU fiscal surveillance. In effect, by definition, it can only be useful for those national economies which tend to experience asset bubbles or large credit fluctuations. In the euro area, the obvious candidates would be Spain and Ireland. Indeed, the approach has been found to be less useful for countries like France and Greece, for which the financial variables typically turn out to be non-significant. However, the finance-neutral output gap approach can only be viewed as complementary to more structural approaches, even for countries where the financial cycle has historically had a large amplitude. Tests performed for Spain and Ireland reveal that Borio et al.'s approach predicts a large positive (finance-neutral) gap before the 2009 crisis, a development that is not fully captured by the conventional models. However, Borio et al.'s approach tends to produce much more persistent cycles, which results in a much 'coarser' view on the cyclical stance of an economy, even for those countries which have been experiencing large financial cycles. This is because it ignores cyclical indicators, which capture important economic developments (e.g. unemployment movements) which are typically only weakly (or not at all) correlated with financial variables. For example, the finance-neutral gaps tend to be strongly positive during the whole period of the 2000s, hence clearly missing the downturns of most European economies after the dot-com bubble burst. Overall, the finance-neutral output gap concept could become a useful tool to complement the more conventional, structural output gaps. This is particularly so in view of the last financial and economic crisis, which according to most economists had its origins in the financial sector. However, they cannot be considered the only, or even the main, instrument for assessing the cyclical position of national economies.

References:

Borio, C., P. Disyatat and M. Juselius (2013), 'Rethinking potential output: Embedding information about the financial cycle' Bank for International Settlements Working Paper No 404

Borio, C., P. Disyatat and M. Juselius (2014), 'A parsimonious approach to incorporating economic information in measures of potential output' Bank for International Settlements Working Paper No 442

Consequently, forecast-to-forecast stability in estimates is an important criterion when evaluating a method's performance. Graph II.1 summarises the forecast-to-forecast revisions for the PF and HP methods for the 2002-2014 period.

The graph shows that both methods produce estimates that are relatively stable in the short term (with average revisions, over all European Commission forecast vintages from 2002-2014, of only 0.06 pp). While we do see an improvement in the relative stability performance of the PF method over the entire period, we must conclude that the methods do not differ greatly in this respect.

Graph II.1: **Short-term (forecast-to-forecast) stability of output gap estimates, euro area (1)(2)**
(pps)
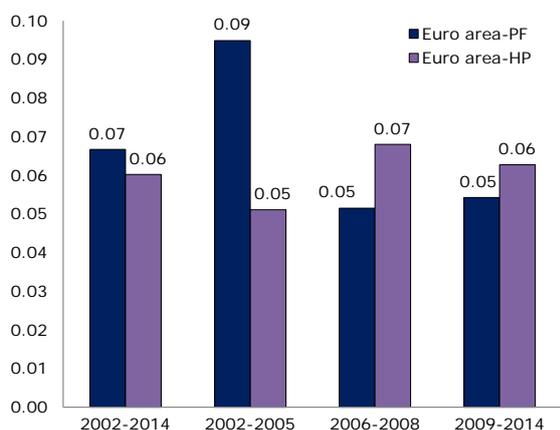
(1) Produced by the PF and HP filter methods.
(2) Period average revisions to estimates.
Source: DG ECFIN calculations.

Long-term real-time reliability of PF and HP estimates

While relatively stable short-term output gap estimates are important, stability should not come at the expense of long-term real-time reliability. Here we compare real-time and *ex post* PF and HP filter estimates for the euro area for 2002-2014 using the Commission's autumn forecast vintages. The autumn 2014 vintage is used as the *ex post* reference, with the scale of the differences between real-time and *ex post* estimates used as an indication of the methods' relative reliability. The results are shown in Graph II.2. The main conclusions are:

- Over the period as a whole, revisions under the methods differ relatively little overall.

Graph II.2: **Long-term reliability of output gap estimates, euro area (1)(2)**
(pps)
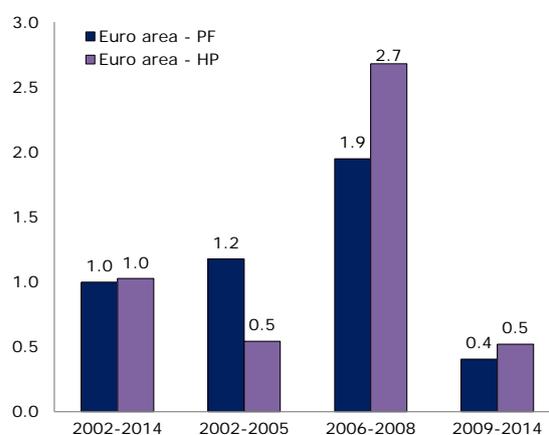
(1) Produced by the PF and HP filter methods.
(2) Period average absolute revisions to real-time vs ex post estimates.
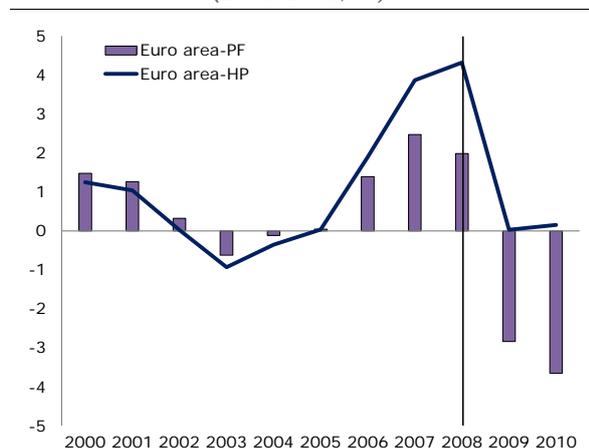***Source:*** DG ECFIN calculations.

- While the average 2002-2014 revisions under the methods may be similar, this hides a much more interesting and nuanced dynamic picture when one looks at trends over time. Graph II.2 suggests that the relative reliability of the PF method has gradually improved over time. While the HP filter outperformed the PF method in the early years of its existence (2002-2005), the relative performance of the latter improved dramatically in the run-up to the crisis (2006-2008), partly perhaps thanks to the introduction in autumn 2005 of the 'hours worked' factor. As regards the post-crisis period (2009-2014), the PF method outperformed the HP filter (due *inter alia* to the introduction of the TFP method in autumn 2010) and

- While a method's relative revisions performance is important, policy-makers should not lose sight of absolute revisions. Graph II.2 shows that both methods made big mistakes in calculating euro-area output gaps in the pre-crisis period, with extremely large average annual *ex post* revisions of 2 pps for the PF method and 2.75 pps for the HP filter. Consequently, any future research agenda should focus on how the methods can be adapted to reduce revisions in the upswing stage of cycles by addressing the optimistic bias inherent in the potential or trend growth rates produced in both cases. Once the extent of the growth optimism had been exposed with the Lehman Brothers default in September 2008,

the backward smoothing of revisions to the level of potential output in the pre-crisis years led directly to revisions in those years that were multiples of those made in 2002-2005 and 2009-2014.

Performance of methods around the turning point of the financial crisis

Graph II.3 shows enormous differences between the output gap estimates produced by the two methods around the turning point of the crisis, with the HP filter pointing to a zero output gap for 2009 and 2010 in the spring 2009 forecast vintage, compared with an average of roughly -3.25 % for the PF method. Following the economic turmoil provoked by the onset of the financial crisis in September 2008, it is not credible that about eight months later the HP filter was estimating a zero output gap for the post-crisis years, 2009 and 2010. The PF method's estimate of -3.25 % for the same years was undoubtedly more consistent with the economic conditions in the euro area at the time, with the 'hours worked' change in 2005 contributing strongly to its performance around this crucial cyclical turning point.

Graph II.3: **PF vs HP output gaps, euro area (1)(2)**
(2000-2010, %)

(1) Spring 2009 forecast exercise.
(2) Turning-point of financial crisis: 2008-2010.
Source: DG ECFIN calculations.

Economic plausibility of PF and HP estimates

In addition to the introduction of hours worked in 2005, EU policy-makers included a new TFP estimation method in the PF framework in

2010. [31] This helped further enhance the credibility of the overall methodology. For example, a comparison of the real-time and *ex post* output gap estimates for the euro area under the HP and Kalman Filter (KF) TFP approaches since the crisis shows clearly that the introduction of the latter did lead to significant reliability gains. For the 2010-2014 period as a whole, the average reduction in revisions with the KF method was of the order of one third.

The impact of the KF TFP method is particularly visible in terms of the overall output gap. Graph II.4 shows output gaps for the euro area from the autumn 2014 forecast using the following three methods:

- the official PF methodology, which includes the KF TFP method;

- a version of the PF methodology which replaces the KF TFP method with the old HP filtered TFP approach used up to autumn 2010; and

- the HP filter methodology on actual GDP, which was the official method used up to autumn 2002.

- Graph II.4 shows that, for 1996-2006, the shift to the KF TFP method explains almost all of the difference in output gaps estimated under the old HP method and the new PF method. From 2006, the HP output gap is always higher than the PF output gap, with the PF output gap using the HP TFP component somewhere in the middle. Graph II.4 supports the view that introducing the KF TFP method in 2010 has further improved the economic plausibility of the estimates produced by the PF method, in that it led directly to a larger negative output gap in all of the years since the crisis. The gain

---

[31] The old HP filtered TFP approach was replaced by a Kalman Filter (KF) approach. This decision was based on the evidence in the literature that multivariate methods (i.e. KF) lead to improved real-time output gap estimates compared with univariate filters (i.e. HP). According to D'Auria F., et al. (2010) on the EU's PF methodology, the change towards a bivariate method for the extraction of trend TFP was expected to help in avoiding both an overestimation of trend TFP in 'good' times and an underestimation in 'bad' times.
D'Auria F., C. Denis, K. Havik, K. McMorrow, C. Planas, R. Raciborski, W. Röger and A. Rossi (2010), 'The production function methodology for calculating potential growth rates and output gaps', DG ECFIN, *European Economy, Economic Papers*, No 420,

is particularly evident in 2011, when the alternative HP filtered TFP method would have produced an overall output gap of zero for the euro area as a whole. The PF estimates – substantially more negative than the HP filter estimates for each of the years since the start of the crisis – are much more consistent with the evidence from other cyclical indicators for this period.

Graph II.4: **Average size of output gap of total economy, euro area (1)**
(1991-2016, %)
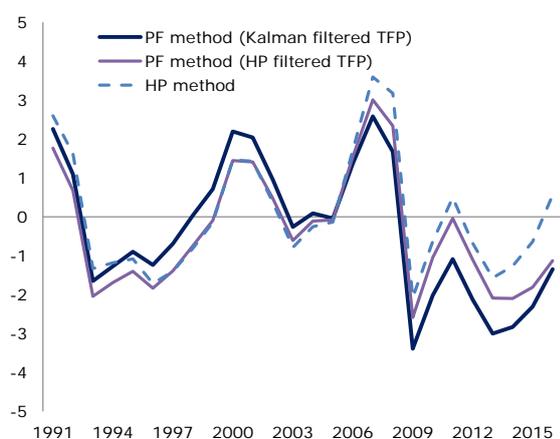
(1) Autumn 2014 vintage estimated with the old HP method (used up to 2002) and two variants of the PF method (one using HP filtered TFP; the other using KF TFP).
Source: DG ECFIN calculations.

Overall evaluation of relative performances of the PF and HP methodologies: was the shift to the PF method justified?

The previous sections have provided a significant amount of evidence to vindicate the decision at EU level to shift to the PF method in 2002. Apart from the obvious advantage of providing policy-makers with a more comprehensive framework for evaluating structural and fiscal policies and for analysing economic trends, the PF method has outperformed the HP filter method in a number of other important respects:

- At the level of the euro area as a whole, it has a consistently **better real-time reliability** record since 2006, producing substantially lower absolute revisions than the HP filter over the pre-crisis (2006-2008) and post-crisis (2009-2014) periods;

- As predicted in 2002, it has proven itself at important cyclical turning points by **alleviating**

the risk of end-point biases, with this gain dramatically demonstrated in the spring 2009 forecast, which was dominated by the crisis. It produced more intuitive output gap levels for 2009-2010 (-3 %/-3.5 %) than the end-point-bias-afflicted HP filter estimate of zero to slightly positive; and

- The adoption of the new TFP methodology in autumn 2010, with trend TFP estimates corrected with capacity utilisation, has **reduced its overall cyclicality** as compared with that of the HP filter. As the TFP gap constitutes a major component of the overall output gap, it is not surprising that any improvements from using the bivariate TFP method would translate into more intuitive overall output gaps.

While the PF method has clearly done well in relative terms, this section also stresses the importance of recognising the extent of output gap errors made in the pre-crisis period.

### II.3. Comparison with the equivalent OECD and IMF methodologies

The previous section stresses the better performance of the EU's PF method over the 2002-2014 period, as compared with the HP filter. Nevertheless, a more pertinent issue is how it has performed compared with the equivalent OECD and IMF methods. Here we assess the respective performances of the EU, OECD and IMF methods over the period as a whole (2004-2014), and in the pre-crisis (2006-2008) and post-crisis (2009-2014) sub-periods. For a longer-term perspective, we go on to examine the results from an equivalent comparative revisions exercise published by the German *Bundesbank* in April 2014, which covered the 1980-2010 period. The *Bundesbank* compared the output gap revisions from the IMF and the OECD methodologies with those using a HP filter. Since it did not include the EU methodology in its comparison, we have applied the same approach to the EU estimates for the same period, in order to compare the degrees of uncertainty surrounding estimates from the EU, OECD and IMF methods.

Short-term stability of EU-PF vs OECD and IMF estimates

As with the PF/HP comparison, this section starts with an evaluation of the relative short-term stability of EU, OECD and IMF output gap

estimates. Graph II.5 summarises forecast-to-forecast revisions for the three institutions for 2004-2014. Unlike the PF/HP comparison, where stability differences were relatively small, Graph II.5 shows that the forecast-to-forecast revisions for the EU method are substantially smaller than both the IMF's and the OECD's for the period as a whole. This outperformance is particularly striking for the 2009-2014 sub-period, where EU revisions are roughly a third of the IMF's and a quarter of the OECD's. This could have non-negligible implications for policy-relevant fiscal indicators such as the change in the structural fiscal balance. While more research is needed to explain the source of these post-2008 differences, one possible explanation is that they are linked to the introduction of the multivariate TFP method in 2010.

Graph II.5: **Short-term (forecast-to-forecast) stability of output gap estimates, euro area, (1),(2)**
(pps)
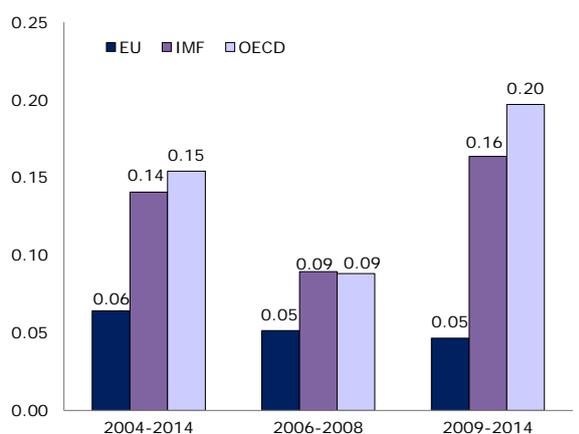
(1) Produced by the EU-PF vs the IMF and OECD methods.
(2) Period average revisions to estimates.
Source: IMF, OECD, DG ECFIN calculations.

### Long-term real-time reliability of EU-PF vs OECD and IMF estimates

Graph II.6 compares real-time and *ex post* (i.e. autumn 2014) output gap estimates for the euro area as a whole for 2004-2014 (autumn vintages), as produced by the EU, IMF and OECD methods. It shows the average absolute revisions for the three sets of estimates for the period as a whole and for the pre-crisis (2006-2008) and post-crisis (2009-2014) sub-periods. The autumn 2014 vintage is used as the *ex post* reference in all three cases. The scale of the differences between the

real-time and *ex post* estimates provides an indication of the relative reliability of the IMF, OECD and EU approaches.

The key conclusions to be drawn from Graph II.6 are as follows:

- For the **period as a whole**, the real-time reliability of the estimates produced by the EU method is significantly better than in the case of the IMF and OECD methods. The average absolute revisions for the EU method are less than half those of the OECD method (0.9 vs 2.0) and significantly smaller than those of the IMF method (0.9 vs 1.3);

- While errors in the **pre-crisis period** were significantly higher for all three institutions (than for the period as a whole and the post-2008 period), the EU method is much more reliable than that of the IMF and especially that of the OECD; and

- For the **2009-2014 period**, the real-time reliability performances of the EU and IMF methods converge substantially, but the OECD's performance remains very much an outlier, with revisions roughly three times greater. The EU and IMF real-time and *ex post* estimates are strikingly similar for each year of this period.

Graph II.6: **Long-term reliability of output gap estimates, euro area (1),(2)**
(pps)
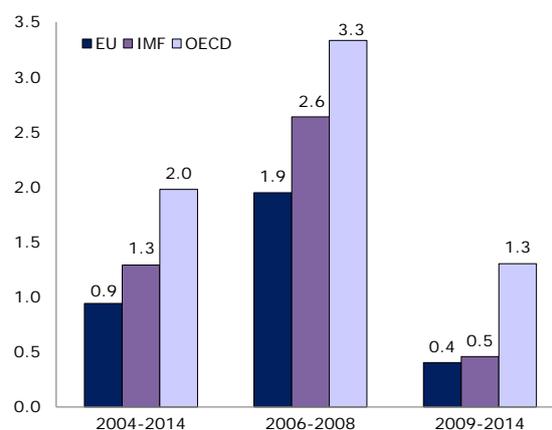
(1) Produced by the EU-PF vs the IMF and OECD methods.
(2) Period average absolute revisions to real-time vs ex post estimates.
Source: IMF, OECD, DG ECFIN calculations.

Performance of EU vs IMF and OECD methods around the turning point of the financial crisis

As to relative performances around the time of the crisis, Graph II.7 shows the real-time (spring 2009) and *ex post* (autumn 2014) output gap estimates for 2009 from the EU (PF and HP filter), IMF and OECD methods.

Graph II.7: **Real-time (spring 2009 forecast) and ex post (autumn 2014 forecast) output gap estimates, euro area (1)**
(pps)
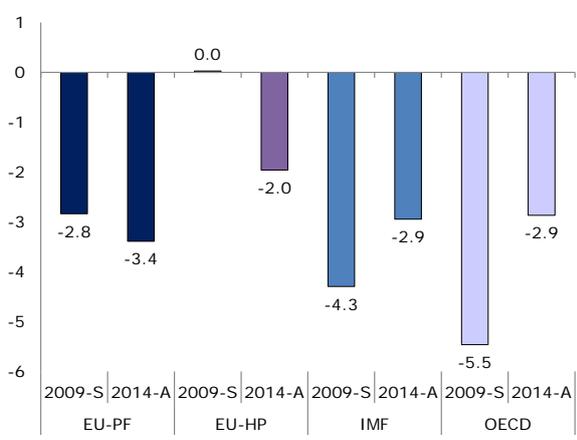
(1) EU (PF and HP), IMF and OECD methods.
Source: IMF, OECD, DG ECFIN calculations.

The graph shows that, in spring 2009, forecasts for the 2009 euro-area output gap ranged very widely, from a zero forecast from the HP filter (implying that most of the effects of the crisis were structural) to -4.3 % and -5.5 % from the IMF and the OECD respectively (implying the opposite, i.e. that most of the effects were cyclical), with the EU PF method in the middle (-2.8 %). Five and a half years later (in autumn 2014), the revisions for 2009 suggest that not only did the PF method do significantly better than the HP filter, but also that it did much better than the IMF's and, especially, the OECD's. In fact, the *ex post* estimates produced by the IMF and OECD methods (-2.9 %) are almost identical to the EU's initial (i.e. real-time) spring 2009 estimate (-2.8 %).

Average spread of estimates and number of years in which the sign of the output gap changes

One of the most recent attempts to evaluate the reliability of international organisations' output gap

estimates was published by the *Bundesbank* in April 2014. It focuses on the output gaps produced by the OECD and the IMF methods, but not those using the EU PF methodology. [32] The main criteria it uses to evaluate the real-time reliability of the OECD and IMF estimates are the **average spread** of the estimates and the **number of years in which the sign of the output gap changes**.

The purpose of the current section is to extend the *Bundesbank*'s analysis to include the output gap results from the EU's PF methodology for the 1980-2010 period and then to assess the relative real-time reliability of the output gaps produced by all three institutions. We replicated the *Bundesbank* approach for the output gaps produced by the EU's common methodology using the bi-annual Commission forecast vintages for 2004-2014 (a total of 19 vintages). The analysis is restricted to the three G-7 countries in the euro area: Germany, France and Italy. For these three countries, we examined the output gap estimates for each year in the two periods covered in the *Bundesbank* analysis, II.e. 1980-1997 and 1998-2010. The output gap spread for a given year is calculated as the difference between the maximum and minimum values and the change in the sign of the output gap over the different vintages is identified. The first part of Table II.1 shows the average spread for the two periods and the second part gives the number of times that an estimate changed its sign, at least once, for a given year in each period. Table II.1 shows that applying the *Bundesbank* approach to the EU's methodology would lead to the following conclusions:

- On the first criterion of real-time reliability (the **average spread** of output gaps), the EU's methodology is consistently and, in a significant number of countries/periods, substantially better than the equivalent IMF and OECD methodologies; and

- On the second criterion of real-time reliability (the **number of years in which the sign of the output gap changed**), the EU's methodology is at least as good or substantially better for the three euro-area countries in five of the six periods in question (the exception being 1980-1997 for Italy).

Table II.1: **Assessment of real-time reliability of OECD, IMF and EU PF output gap methodologies (1)**

		Average spread of output gaps (maximum value for a year over the 19 different forecast vintages less the minimum value)			Number of years in which the sign of the output gap changed		
		IMF	OECD	EU PF (ECFIN calculations) (2)	IMF	OECD	EU PF (ECFIN calculations) (2)
Germany	1980-1997	3.6	3.1	1.1	7	8	4
	1998-2010	1.9	1.9	1.2	5	4	4
France	1980-1997	1.6	2.2	1.4	5	5	2
	1998-2010	2.6	2.7	2.1	11	8	7
Italy	1980-1997	4.2	2.0	1.8	10	5	7
	1998-2010	3.1	3.3	1.9	10	11	7

(1) Two assessment criteria: average spread of output gap estimates and number of years in which the sign of the output gap changes.
(2) Calculations apply the methodology used in the study on the reliability of international organisations' estimates of the output gap published in April 2014, in the Bundesbank Monthly Report.
Source: DG ECFIN calculations.

II.4. Concluding remarks

This focus section provides evidence that vindicates the decision to adopt the PF methodology for estimating output gaps as the 'commonly agreed' reference method to be used in EU fiscal (and, by default, structural policy) surveillance procedures. While uncertainty will always be a feature of output gap calculations for the euro area, we have seen that in relative terms the stability, real-time reliability and financial crisis performance of the PF methodology has been superior to the HP filter, OECD and IMF methodologies.

Not only does the EU's methodology do better over the 2004-2014 period in terms of relative stability and reliability, but it crucially outperforms the HP, OECD and IMF methods around the turning point of the crisis. In spring 2009, estimates for the euro area's output gap in 2009 ranged very widely, from an economically implausible zero estimate from the HP filter to -4.3 % and -5.5 % respectively from the IMF and the OECD, with the EU PF method in the middle (-2.8 %). Five and a half years later (in autumn 2014), the IMF and OECD *ex post* estimates for 2009 for the euro area (-2.9 %) were almost identical to the real-time spring 2009 estimate produced by the EU's methodology (-2.8 %).

In addition, the relatively large OECD and IMF revisions for 2009, compared with the EU's, suggest that the respective methodologies led to a fundamentally different assessment of the impact

of the financial crisis. The extremely large negative output gaps for 2009 predicted by the IMF and the OECD in spring 2009 were consistent with a view at that time that the effects of the crisis on potential output would be relatively limited and temporary in nature. The much smaller negative gaps produced by the EU's methodology suggested a less benign interpretation, namely that the impact on potential would be much more significant and prolonged (a view subsequently confirmed by economic developments in the post-2008 period). This interpretation is supported by a recent *Bundesbank* analysis which stresses that the IMF and the OECD initially interpreted the drop in actual output post-2009 as a cyclical phenomenon. It was not until the economic recovery proved weak that the preceding upward movement in potential output was seen to be unsustainable.

While the relative performance of the EU's methodology suggests that it is well-designed, unfortunately its absolute performance, especially in the pre-crisis period (2006-2008), leaves a lot to be desired. Although it outperforms the HP, OECD and IMF methods in terms of reliability over the pre-crisis period, this provides little comfort from a fiscal surveillance perspective, since it is now clear that big output gap (and consequently structural budget balance) errors were made over this period. Consequently, we must conclude that excessive optimism in the pre-crisis period with respect to underlying growth trends in the EU underlines, yet again, that handling the upswing stage of cycles remains the Achilles heel of all mainstream output gap estimation methods. It

explains why commentators are right to issue regular warnings that most significant fiscal and structural policy errors are made in the 'good times'.

Finally, one can legitimately argue that the pre-crisis period was a once-in-a-generation financial shock and that the real-time reliability performance of the commonly agreed EU method was exceptionally good around the turning point of the crisis and in the subsequent post-crisis years. Whilst this is a valid line of argumentation, nevertheless it is important to continue to improve the EU's commonly agreed methodology, with a particular focus on attenuating the procyclicality risks in the upswing phase of cycles.

In this regard, the annual Work Programme of the Economic Policy Committee's (EPC) Output Gap Working Group (OGWG) is the vehicle via which the EU Member States can bring forward suggestions for further improvements to the method. The current 2015 Work Programme focusses on areas such as: the working age population; refinements of the NAWRU and TFP calculations; exploring the possibility of integrating recent structural reforms into the method; and including additional explanatory macro variables in the methodology.