An assessment of the real time reliability of the Output Gap estimates produced by the EU's Production Function Methodology

(K Mc Morrow, W Roeger, V Vandermeulen & K. Havik)*

An assessment of the real time reliability of the Output Gap estimates produced by the EU's Production Function Methodology

Introductory Remarks

Real time estimates of the output gap (the difference between actual & potential GDP) are surrounded by a large element of unavoidable uncertainty since potential output is not directly observable whilst actual GDP is subject to significant historical / forecast revisions. Many studies (e.g. Orphanides & van Norden, 2002; Nelson & Nikolov, 2003; Cayen & van Norden, 2005; Marcellino & Musso (2011); Bundesbank (2014)) have documented the extent of the uncertainty, highlighting the fact that the sign, as well as the magnitude, of the output gaps estimated in real time are subject to large revisions as new information becomes available with the passage of time. EU policy makers are also very well aware of these output gap uncertainties but nevertheless accept that estimates of potential growth & the output gap are indispensable in assessing the cyclical position of the economy & its productive capacity. In this context, output gap indicators have been used as an operational surveillance tool in the Stability & Growth Pact (SGP), since its inception in the second half of the 1990's, for calculating indicators of the structural (i.e. cyclically adjusted) fiscal balance, with the main focus on the change in, rather than the level of, the structural balance given the doubt surrounding the underlying output gap calculations.

Initially, a purely statistical detrending method, the HP filter, was used to calculate the output gaps used in the SGP. However, following sustained criticism of the HP filter, especially its end-point bias problems, the EU's Economic Policy Committee (EPC) set up an Expert Working Group in 1999 to review the advantages & disadvantages of the most widely used statistical (e.g. HP filter) & economic (e.g. the Production Function – PF) estimation methods used by national & international policy making institutions. Whilst the revisions properties of the HP & PF estimation methods turned out to be not that different, nevertheless the Expert Working Group recommended in 2001, by a large majority, that the EU should adopt the PF approach for surveillance purposes given its relative advantages compared with the HP filter, both in terms of its more comprehensive analytical framework & its benefits in alleviating the risk of end-point biases. Following EPC approval, the ECOFIN Council duly adopted the PF approach in July 2002 & it has been used ever since for operational policy surveillance purposes in the EU.

In the aftermath of the financial crisis, & following the strengthening of the EU's policy surveillance procedures, the growing importance of the output gap estimates produced by the PF method has been accompanied by a sustained level of critical commentary, from academics, policy think-tanks, as well as policy makers, with the PF method's relative real time reliability performance regularly being called into question. There have been three broad strands to the criticisms levelled at the method's real time performance :

- Firstly, there has been a significant group of commentators over the years which have questioned the EPC / ECOFIN Council decision to move away from the simple HP filter method in 2002. Some have argued that the real time reliability gains from moving to the more complicated PF method have not been worth the additional complexity involved (with the greater level of intricacy essentially linked to the steady shift over time towards embedding multivariate estimation approaches inside the PF framework). In addition, others have suggested that the HP filter may actually provide more reliable results. This latter viewpoint is very much the conclusion of an April 2014 analysis by the Bundesbank 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 & the IMF. Whilst the procyclicality issues with a HP6.25 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 & the PF approach is far from clearcut.

- Secondly, with the entering into force of the European Fiscal Compact in January 2013 (& its requirement that the structural deficit of the Euro Area's Member States be less than 0.5%), the questioning of the real time reliability of the output gaps produced by the EU's method has intensified considerably, with these EU estimates regularly being compared with (& criticized as being inferior to) those produced by other international organisations such as the OECD & the IMF. For example, Pier Carlo Padoan, Italy's Economy Minister, openly questioned the measurement of the Italian output gap by the EU's method in an interview with the Financial Times in November 2014. Mr Padoan noted that the EU's estimate of the Italian output gap in 2014 was -3.5% of GDP which was significantly lower than the -5.1% from the OECD. Mr Padoan suggested that if the latter number was applied, Italy "would be in structural surplus now … we would be in a different world… we would have to do nothing (in terms of fiscal actions)". Other criticisms of the EU's methodology have been put forward by Cottarelli et al. (2014); by the CPB in the Netherlands in a policy brief by Hers & Suyker (2014); & by the economic think tank Bruegel, both in a conference which Bruegel organized in 2014 on "Assessing the European Fiscal Framework" & in its ongoing "Blogs Review" on output gaps & the structural balance.

- Finally, & in methodological terms the strand of criticism with the most objective validity, there is a widely held belief that the output gap estimation method currently used by the EU, as well as those of the OECD & 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 pre-financial crisis period from 2006-2008. The extent of the procyclicality evident in this period can only be partially explained by a systematic optimistic bias in the output gap methodologies of the EU, IMF & OECD. The size of the ex post output gap revisions for the pre-crisis period point to a more fundamental weakness in such methods in the way they handle investment in the boom, not the bust, phase of cycles. This point has been raised by BIS economists (Borio et al. 2014), with their suggested solution being to augment the conventional output gap calculation methods by using information about the financial cycle to identify investment booms & in this way to produce "finance-neutral" output gaps.

Against the above background of sustained criticism, & given the necessity of retaining an ongoing scepticism as to the PF's underlying performance, the present paper tries to objectively assess the real time performance of the EU's PF method over the 13 years since it was first introduced in Autumn 2002. Given that it is universally accepted that output gap uncertainty is a fact of life for all estimation methods & that output gap estimates are inevitably subject to large revisions, **what we are therefore focussing on in this paper is the relative revisions performance of the EU's PF method**, not the absolute size of those revisions. More specifically, the paper will try to answer three fundamental, interrelated, questions :

- Firstly, how has the EU's PF methodology performed, in terms of its revisions record, compared with the previously used HP filter approach ? In essence, was the judgement of the EPC / ECOFIN Council in 2002 to move towards the economics inspired, PF method, the correct one in hindsight ? (Section 1).

- Secondly, how do the output gap revision properties of the EU's method compare with those of the equivalent methods used by other international organisations, such as the OECD & the IMF ? Is it actually true, as many commentators have suggested, that the EU's real time output gap estimates are more unreliable than those of the OECD & the IMF ? (Section 2) &

- Finally, how can the EU's common methodology be further improved, most notably with respect to avoiding estimation errors in the upswing phase of cycles (with the 2006-2008 pre-crisis, investment boom, period being a spectacular example of such errors) ? Would the suggestion to exploit the information content of financial cycle variables such as real interest rates, credit growth & property prices provide a more robust approach to the decomposition of actual time series into their trend & cyclical components ? (Section 3)

Section 1 : How does the real time reliability of the Output Gaps produced by the EC's PF methodology compare with those of its predecessor, the HP Filter ?

The EU moved to the Production Function (PF) approach to estimating output gaps in Autumn 2002, following an EPC / ECOFIN Council evaluation of the performance of the previously used HP filter approach (see Box 1). The current section assesses whether the 2002 decision was a prudent one or not. More specifically, it compares the revision properties of both the PF & HP methods by firstly, comparing the real time reliability of the PF & HP methods at the level of the Euro area as a whole since the PF method's inception in 2002 up to 2014 (section 1.1); & then examining whether the optimism in 2002 regarding the ability of indicators such as capacity utilisation to reduce the cyclicality of trend TFP estimates was justified (section 1.2).

1.1 An assessment of the real time reliability of the PF & HP filter (with λ=100) methods : Aggregate Euro Area Analysis 2002-2014 : Whilst there are a number of different ways to measure the real time reliability of output gap estimation methods, the current section evaluates the results using just one, widely used, approach[1]. It compares the real time & ex post output gap estimates for the Euro Area produced by the PF & HP filter methodologies respectively for the period 2002-2014, using the Commission services Autumn forecast vintages. The autumn 2014 vintage is used as the ex post reference (leading to no revisions for the year 2014), with the size of the differences between the real time & ex post estimates being used as an indication of the relative reliability of the PF & HP filter methods. The results are shown in Graph 1a-1d, with the main conclusions being the following :
- Firstly, if one was to compare the respective performances of both the PF & HP methods based on the evidence in Graphs 1a & 1b, one would conclude that apart from around the crisis, where the PF method clearly did better, the overall degree of

---

[1] Annex 1 provides the results from an alternative approach where the focus is on comparing output gap revisions over a relatively short time horizon, i.e. from one forecast vintage to the next. In terms of the EU's fiscal policy surveillance exercises conducted under the Stability & Growth Pact, this "forecast to forecast" revisions approach is possibly the best one to use if one is solely interested in assessing the implications for the change in the structural fiscal balance.

revisions differences between both methods was relatively small over the period as a whole. This would also be the conclusion if one looks at the evidence for the average absolute revisions for both methods for the period as a whole shown in graph 1c, with the first two columns in graph 1c suggesting that both methods had broadly similar revisions records for 2002-2014, with a slight overall advantage for the PF method (note : to construct graph 1c we simply sum the absolute value of the output gap revisions for each of the years 2002-2014 & then present an average of those revisions – we also provide the average absolute revisions for the sub-periods 2002-2005, 2006-2008 & 2009-2014 respectively in the same graph).

- Secondly, whilst the average 2002-2014 revision properties of both methods may be similar, this average performance hides a much more interesting & nuanced dynamic picture when one looks at the evolution of revisions over time, as shown in graphs 1a & 1b. These graphs suggest that the PF method has gradually improved its relative reliability performance, with this picture confirmed in the average absolute revisions for the different sub-periods shown in graph 1c. From graph 1c one sees that whilst initially the HP filter outperformed the PF in the early years of its existence (i.e. the period 2002-2005), there was a dramatic improvement in the relative performance of the PF method in the run up to the crisis (i.e. 2006-2008), with the introduction of hours worked into the PF method in Autumn 2005 being perhaps one of the drivers of this enhanced revisions performance. Regarding the post-crisis period (2009-2014), graph 1c shows that again the PF method outperformed the HP filter. In addition, graphs 1a & 1b indicate that the level of the PF output gaps were substantially more negative (& therefore more economically plausible) than those of the HP filter for each of the post-crisis years from 2010 onwards.

- Thirdly, around the turning point of the crisis, graph 1d shows an enormous difference in the output gap estimates of both methods, with the HP filter pointing to a zero output gap for 2009 and 2010 in the Spring 2009 forecast vintage, compared with -3 1/4% for the PF method. Following the economic turmoil provoked by the onset of the financial crisis in September 2008, it is not credible that roughly 8 months later, the HP filter was estimating a zero output gap for the post crisis years 2009 & 2010. The PF method's estimate of -3 ¼% for the same years was undoubtedly more consistent with the economic conditions prevailing in the Euro area at that time, with the hours worked change in 2005 contributing strongly to the PF's performance around this crucial cyclical "turning point".

- Finally, whilst the relative revisions performance of both methods is important, policy makers should not lose sight of the absolute level of those revisions. Graph 1c shows that both methods got the Euro Area's output gaps badly wrong in the pre-crisis period, with extremely large average annual ex post revisions of 2% points for the PF method & 2 ½% points for the HP filter. Consequently, any future research agenda for output gap calculation methods should focus on how the methods can be adapted to reduce the revisions in the upswing stage of cycles by addressing the optimistic bias inherent in the potential / trend growth rates produced by both estimation approaches. Once the extent of the growth optimism was spectacularly exposed with the Lehman crash in September 2008, the backward smoothing of the revisions to the level of potential output in the pre-crisis years led directly to output gap revisions in those years which were multiples of the revisions made at other periods of the cycle.
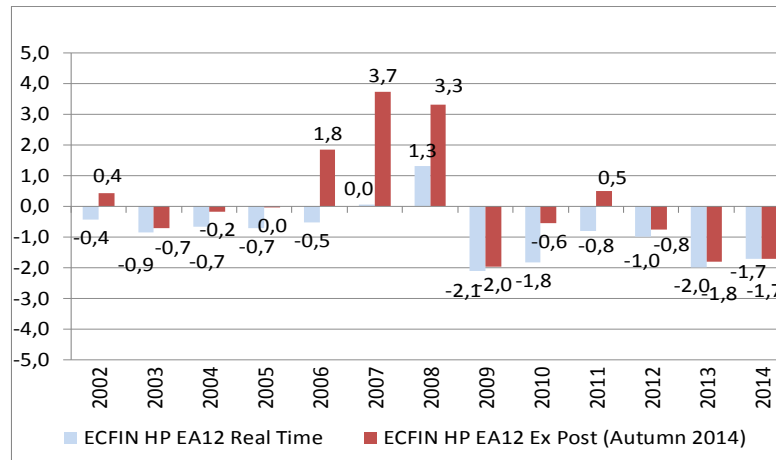
Graph 1 : Reliability of Output Gap estimates for EA12
Average Absolute Differences– Real Time versus Ex Post (Autumn 2014) Estimates : HP vs PF methodologies (2002-2014)
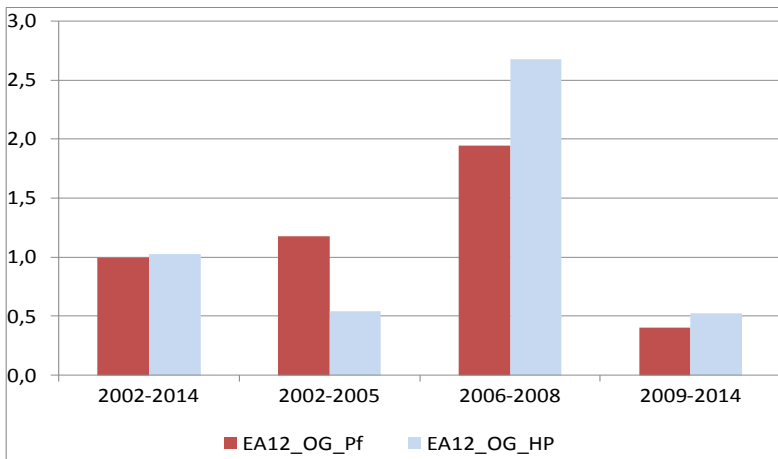
Graph 2A. PF: Real Time vs Ex Post OG estimates

ECFIN Pf EA12 Real Time ■ ECFIN Pf EA12 Ex Post (Autumn 2014)
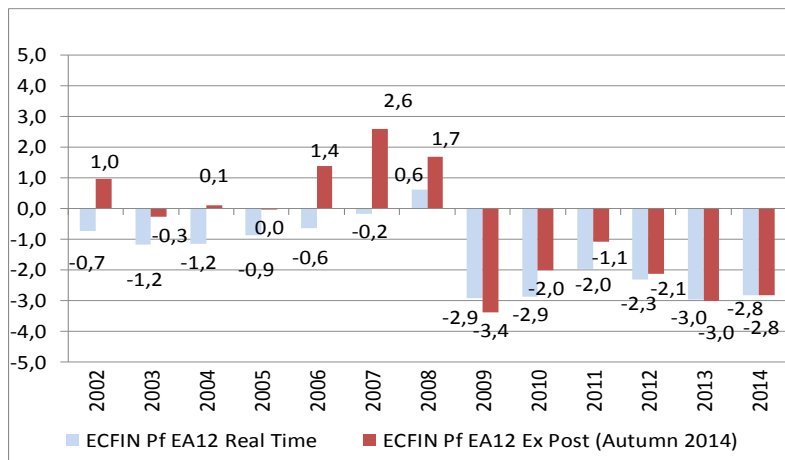
Graph 2B. HP: Real Time vs Ex Post OG estimates

ECFIN HP EA12 Real Time ■ ECFIN HP EA12 Ex Post (Autumn 2014)

Graph 2C. Real Time vs Ex Post: Real Time Reliability of PF & HP methods
(Annual Avg Revisions in % points)

■ EA12_OG_Pf ■ EA12_OG_HP

Graph 2D. PF vs HP: Spring 2009 Forecast
(Turning Point of Financial Crisis: 2008-2010)

ECFIN EA12_OG_PF ---- ECFIN EA12_OG_HP

6
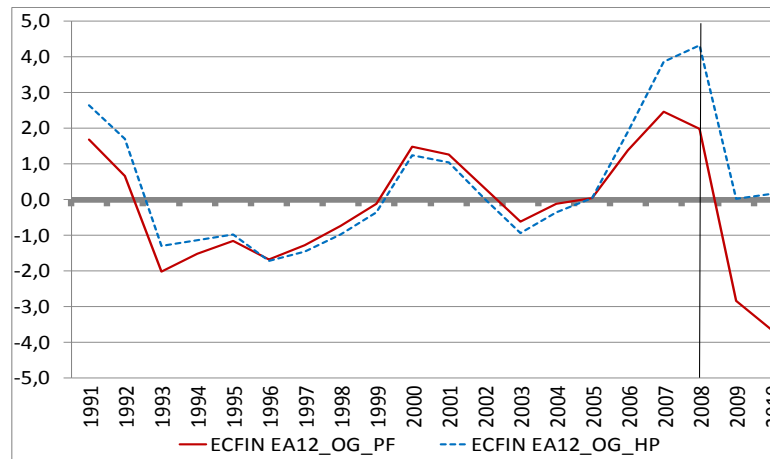
1.2 What has been the impact from the replacement of the HP filtered trend TFP approach by a Kalman filter (KF) approach in Autumn 2010 ? Do multivariate methods (i.e. KF) lead to improved real-time Output Gap estimates compared with univariate filters (i.e. HP) ?: According to the 2010 ECFIN Economic Paper 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. Planas, Roeger and Rossi (2010) showed that for all Member States, the bivariate method outperformed the HP method for real-time TFP gap estimates. They documented large reductions in the magnitude of the revisions when the bivariate method for extracting the TFP cycle was compared with the univariate HP method used in the production function methodology prior to 2010. As a result of this research, sizeable reductions in revisions were expected with the introduction of the new TFP trend estimation methodology.

What has been the impact of the introduction of the new TFP methodology in 2010 in terms of Output Gap revisions for the Euro Area as a whole : Graph 2 compares the real time & ex post output gap estimates for the Euro Area produced by the production function (PF) using an HP filtered TFP (with λ=100) and a Kalman filtered TFP approach respectively. It covers the post crisis years 2010-2014, using Autumn vintages, with the autumn 2014 vintage used as the ex post reference. As was the case with the analysis in graph 1, the size of the differences between the real time & ex post estimates is used as an indication of the relative reliability of the methods. The graph shows that for most of the post-crisis years, the revisions between the real time and ex post output gaps are larger when trend TFP is estimated using an HP filter. However, this conclusion only really applies to the early post-crisis years. Graphs 2a & 2b show that in the first years after the crisis (2010-2011), the introduction of the Kalman filter TFP did lead to significant reliability gains for the EA as a whole, with the evidence for the more recent years pointing to average revisions which are broadly similar between the HP and KF TFP methods. As shown in graph 2c, for the period as a whole, the average reduction in revisions with the KF method is of the order of 1/3, an impressive outcome especially since all of the gains were made in just two years i.e. 2010 & 2011.

The impact from the introduction of the KF TFP method is particularly visible in terms of the level of the overall output gap : Graph 3 shows the output gaps for the Euro Area from the Autumn 2014 forecast produced using the following three methods:
- Firstly, the official version of the PF methodology which includes the KF TFP method;
- Secondly, a version of the PF methodology which replaces the KF TFP method with the old HP filtered TFP approach used up to Autumn 2010; &
- Finally, the HP filter methodology on actual GDP which was the official method used up to Autumn 2002.

Graph 3 shows that in the earlier reported years (1996-2006), the shift towards using the KF TFP method explains almost all of the difference in the levels of the output gaps produced by the old HP method and the new PF method. Since 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 of the overall set of results. Introducing the KF TFP method has thus led to a further improvement in the gap, in the sense that it leads to a larger negative output gap in all of the post crisis years. The gain is particularly evident in 2011 where the alternative HP filtered TFP method would have produced an overall output gap of zero for the Euro Area as a whole.

Graph 2 : Output Gap estimates for EA12 - Real Time versus Ex Post (Autumn 2014) Estimates : HP filtered TFP vs KFiltered TFP (Post Crisis Period – 2010-2014) + Average Absolute Revisions 2010-2014 & KF & HP filtered TFP growth Rates
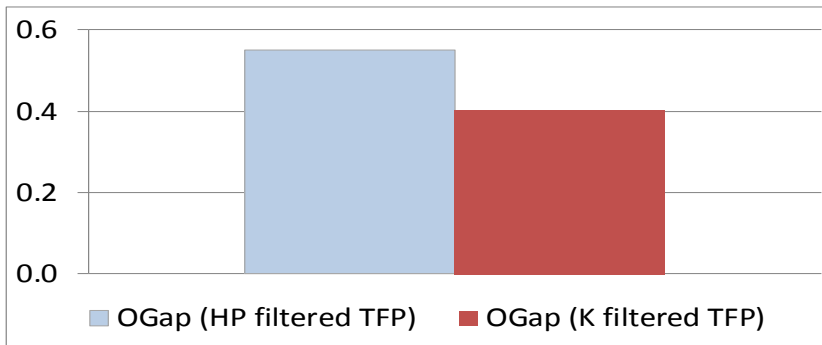
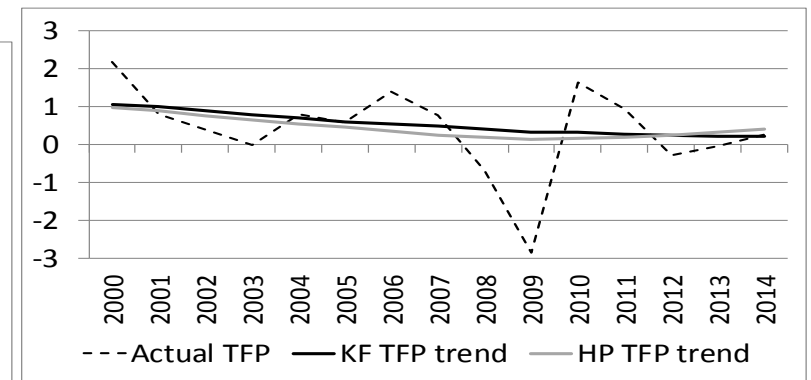A - Output Gap Estimated with HP filtered TFP : Real Time vs Ex Post Estimates
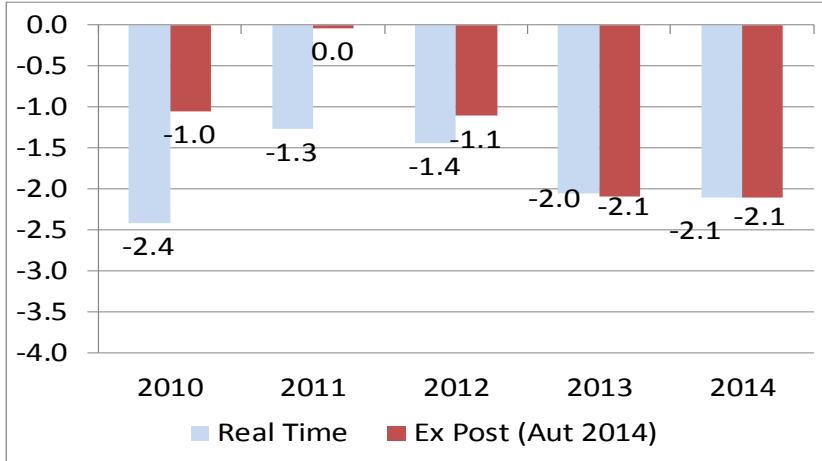
B - Output Gap Estimated with Kfiltered TFP : Real Time vs Ex Post Estimates
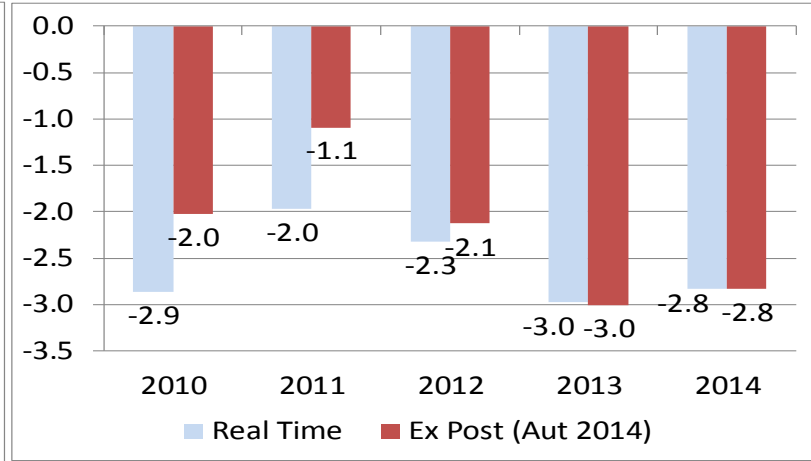

C - Real Time Reliability of Output Gaps Estimated with HP and Kfiltered TFP (Annual Average Revisions 2010-2014: %points)

D - TFP Growth Rates (Actual vs HP Trend vs KF Trend Rates)

Graph 3 : Average Size of Euro Area's Total Economy Output Gap (Autumn 2014 vintage) estimated with the old HP method (used up to 2002) & Two Variants of the PF method (i.e. one using HP Filtered TFP & the other using Kalman filtered TFP)
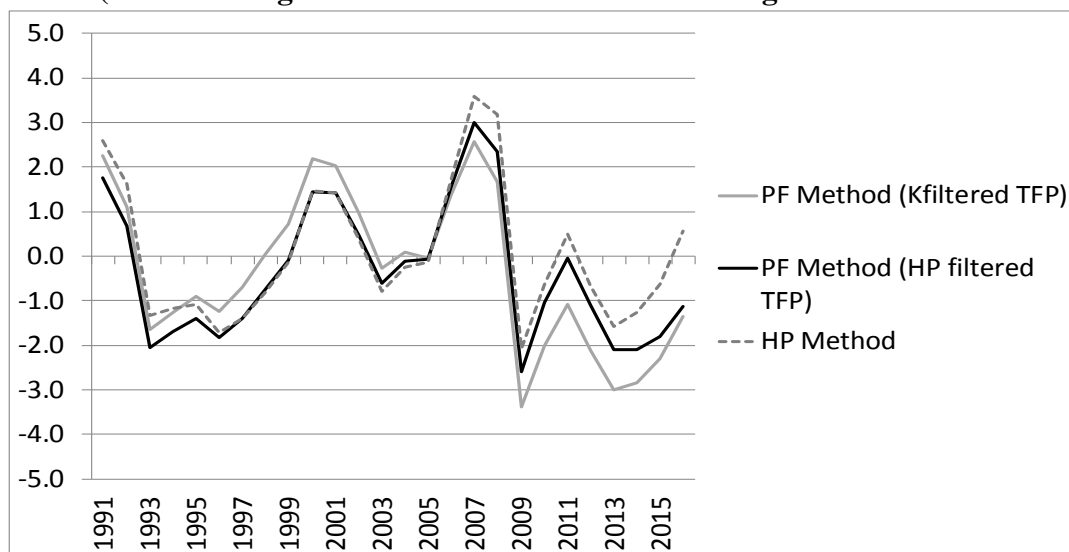
Overall evaluation of the relative performances of the PF & HP filter methodologies over the period 2002-2014 – was the shift to the PF method justified or not ? This section has provided a significant amount of evidence in support of the EPC / ECOFIN Council decision to shift to the PF method in 2002. Apart from the obvious PF advantage of providing policy makers with a more comprehensive framework for evaluating policies & analysing economic trends, the PF method has outperformed the HP filter method in a number of other important respects :

- Firstly, at the level of the Euro area as a whole, since 2006, the PF method has a consistently better real time reliability record, producing substantially lower absolute revisions than that of the HP filter over the pre-crisis (2006-2008) & post-crisis (2009-2014) periods.

- Secondly, as predicted in the 2001 EPC report, the PF method has proven itself at important cyclical turning points by alleviating the risk of end-point biases, with this gain dramatically demonstrated in the financial crisis dominated Spring 2009 forecast, where the PF method produced more economically intuitive output gap levels for 2009-2010 (-3% / -3 ½%) compared with the end-point bias afflicted HP filter estimate of zero to slightly positive for the same years, with the latter HP estimates running counter to the evidence from other indicators of economic activity.

- Thirdly (again as predicted by the 2001 EPC report), the adoption of the new TFP methodology in Autumn 2010, leading to the production of capacity utilisation corrected trend TFP estimates, has resulted in a reduction in the overall cyclicality of the PF method relative to 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 obtained by adopting the bivariate TFP method would translate into more economically intuitive overall output gaps, with the better isolation of cyclical TFP yielding significant benefits for the EU's common methodology.

- Finally, whilst the PF method has clearly done well in relative terms, this section also stressed the importance of recognising the extent of the absolute level of the output gap errors made in the pre-crisis period.

BOX 1 : KEY POINTS FROM THE EPC's REPORT ON POTENTIAL OUTPUT AND THE OUTPUT GAP (October 2001)

1. Introduction : Potential output and output gap indicators have acquired an "operational" status in the Stability and Growth Pact (SGP), as they provide an essential input for calculating indicators of structural (i.e. cyclically adjusted) fiscal balance, which are used in turn for assessing the progress made by countries towards achieving the goal of medium-term fiscal balance. Estimates of potential output and the output gap are known to be particularly uncertain, as different approaches provide estimates, which may differ significantly from each other. Given their reliance on the output gap, estimates of the structural fiscal balance are, in turn, rather uncertain, while measures of its changes over time are generally considered to be more robust (although annual changes in the structural fiscal balance should also be interpreted with caution since they do not fully reflect discretionary fiscal policy changes). Against this background, the Economic Policy Committee (EPC) entrusted an ad-hoc working group of experts, chaired by Mr. Jean Philippe Cotis, to review the estimation methods used by the European Commission (EC) and other national and international institutions, with a view to strengthen the understanding and broaden the consensus on the EC estimates that are used in the surveillance procedures. The group was composed of experts from the EPC and EU Member States, the EC and the ECB, as well as international organizations such as the IMF and the OECD.

2. Assessing the different estimation methods : Members of the group assessed the respective advantages and drawbacks of purely statistical detrending methods, such as the HP filter, and economic methods, such as the production function approach or semi-structural approaches that had been developed more recently. *__Members of the group unanimously stressed the uncertainty surrounding any estimates of the level of potential output and the output gap. Measures of potential output growth, and associated changes in the output gap, appear however much less uncertain as estimates from conventional methods generally show no sizable differences__*.

3. Conclusions and recommendations : The traditional approach used by the EC, based on the HP filter, has served reasonably well in the past. However, the group considers that a simple production function (PF) approach would provide a more comprehensive and adequate framework for assessing the economic outlook as well as the macroeconomic policies of Member countries for EU surveillance procedures implemented by the EC. The new PF approach devised by the EC strikes an appropriate balance between the objective of strengthening the underlying economic analysis and the requirement of maintaining transparent and equal treatment of member countries. The group noted that the new PF estimates were generally relatively close to those provided by the HP filter. Although the resulting potential output growth estimates display similar cyclical patterns as those obtained from the HP filter for past years, this approach can help to alleviate the risks of end-point biases. In addition, some members of the group noted that the remaining cyclicality of trend total factor productivity could be reduced by appealing to a smoother trend or by extracting additional information from other cyclical indicators, such as capacity utilization rates. The group would recommend, by a large majority, that the EC adopt this new production function approach for surveillance procedures, whilst keeping estimates from the standard HP filter approach as an additional reference. (**Note** : *Following presentations by the Chairman of the EPC & by Commissioner Solbes, the ECOFIN Council adopted the PF approach at its meeting on 15 July 2002*)

Section 2 : Real Time Reliability of the EU's Production Function methodology for calculating Output Gap estimates compared with the equivalent OECD & IMF Methodologies

The last section has stressed the better performance, based on a number of criteria, of the EU's PF method over the period 2002-2014 when compared with the method it replaced in 2002, namely the HP filter. Whilst the overall relative performance of the PF method may have been better, a more pertinent current issue for policy makers is how the PF method has been doing relative to the equivalent OECD & IMF methods (since Member State policy makers are much more interested in this comparison rather than one with the HP filter). This section now

addresses this question by assessing the respective performances of the EU, OECD & IMF methods over the period as a whole (2004-2014), as well as in the pre-crisis (2006-2008) & post-crisis (2009-2014) sub-periods[2]. In addition, in order to provide a longer run perspective, the section goes on to examine the results from an equivalent comparative revisions exercise published by the Bundesbank in April 2014 which covered the 30 year period stretching from 1980-2010. The Bundesbank research compared the output gap revisions from the IMF & the OECD methodologies with those of a HP filter. Since the Bundesbank research did not include the EU's methodology in its comparison, we have applied the Bundesbank's approach to the EU's output gap estimates for the equivalent 1980-2010 period, with the objective of assessing the relative degree of uncertainty surrounding the estimates from the EU, OECD & IMF methods respectively.

2.1 EU vs OECD & IMF methods – Euro Area as a whole (2004-2014) : Graph 4 compares the real time & ex post (i.e. Autumn 2014) output gap estimates for the Euro area as a whole for the years 2004-2014 (autumn vintages), produced by the EU (4a), IMF (4b) & OECD (4c) methods, with the last graph in the set of four (4d) showing the average absolute revisions for the three sets of output gap estimates for the 2004-2014 period as a whole, as well as for the pre-crisis (2006-2008) & post-crisis periods (2009-2014) respectively. The autumn 2014 vintage is used as the ex post reference in all 3 cases, explaining why the real time & ex post Output Gap figures for 2014 are identical (i.e. no revisions). The size of the differences between the real time & ex post estimates provides an indication of the relative reliability of the IMF, OECD & EU approaches.

The key conclusions to be drawn from graph 4 are as follows :

- Firstly, regarding the period as a whole (2004-2014), the EU's method was significantly more reliable than those used by the IMF & OECD methods, with the EU's method outperforming that of the OECD in every single year of the 2004-2014 period & outperforming the IMF in all years except one. The average absolute revisions for the EU's method were less than half those of the OECD (0.9 vs 2.0) & were significantly smaller relative to those of the IMF method (0.9 vs 1.3).

- Secondly, whilst the degree of errors in the pre-crisis period 2006-2008 were significantly higher for all three institutions (compared with the period as a whole & the post-crisis period), nevertheless the EU's PF method was, in relative terms, much more reliable than that of the IMF, & especially relative to the OECD.

- Finally, regarding the post-crisis period 2009-2014, whilst the real time reliability performance of the EU & IMF methods converged substantially, the OECD's performance remained very much an outlier, with revisions which were roughly three times greater than those of the EU & IMF methods. Regarding the EU & IMF methods, it is quite striking how similar the real time & ex post output gap estimates have been for the Euro area in each year of the post crisis period.

---

[2] In addition, Annex 1 shows the "forecast-to-forecast" performance of the EU, OECD & IMF methods.

Graph 4 : Reliability of Output Gap estimates for Euro Area
(Average Absolute Differences - Real Time versus Ex Post (Autumn 2014) Estimates : OECD, IMF & EU methodologies & summary indicator of relative real time reliability)

Graph 4A. ECFIN: Real Time vs Ex Post OG Estimates
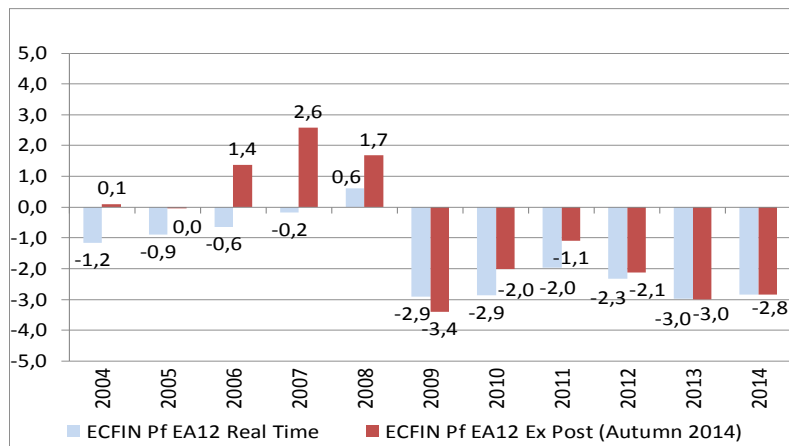
Graph 4B. IMF: Real Time vs Ex Post OG Estimates
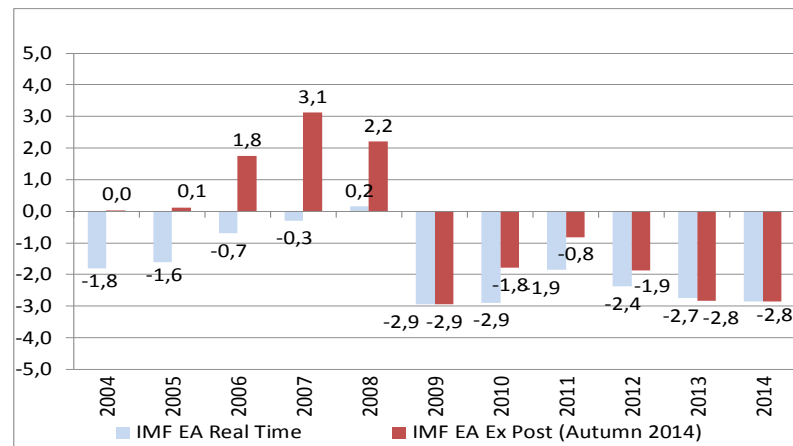
Graph 4C. OECD: Real Time vs Ex Post OG Estimates

Graph 4D. Real Time vs Ex Post: Real Time Reliability of EU, IMF & OECD methods (Annual Avg Absolute Revisions in % points)

Regarding the relative performances of the different institutions around the time of the crisis, graph 5 shows the real time (Spring 2009)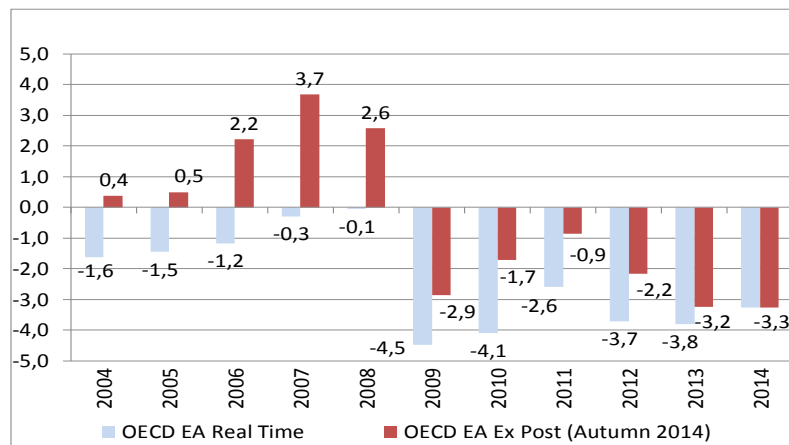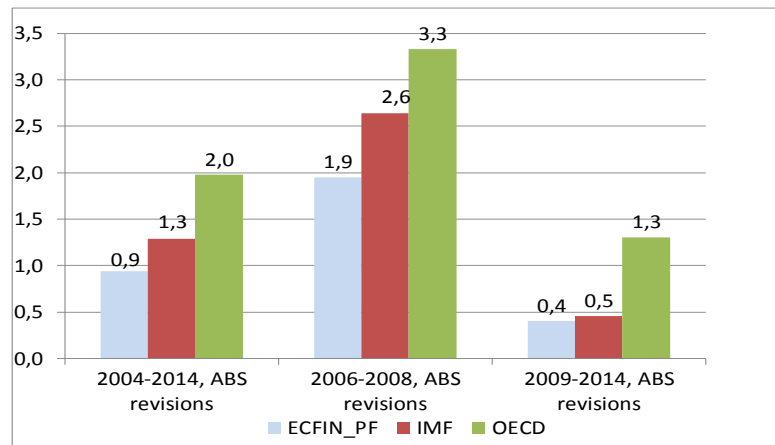 & ex post (Autumn 2014) output gap estimates for the year 2009 from the EU (for both the PF & HP filter methods), IMF & OECD methods plus the associated output gap revisions for all four methods. The graph shows that in Spring 2009, there was a very wide range of forecasts for the Euro Area's output gap in the year 2009, ranging from a zero forecast from the HP filter to -4.3% / -5.5% from the IMF / OECD, with the EU's PF method in the middle of these estimates (-2.8%). Five and a half years later in Autumn 2014, the revisions shown in graph 5b for the year 2009 suggest that not only did the PF method do significantly better compared with the HP filter in terms of revisions, it also did much better relative to the IMF &, particularly so, relative to the OECD. In fact, the ex post output gap estimates for 2009 for the Euro Area produced by the IMF & OECD methods are almost identical to the initial (i.e. real time) Spring 2009 EU estimate for 2009 i.e. -2.9% versus -2.8%.

Graph 5 : Euro Area - Real time (Spring 2009 Forecasting Exercise) & Ex Post (Autumn 2014 Forecasting Exercise) Output gap estimates for the year 2009 from the EU (PF & HP), IMF & OECD methods plus the associated output gap revisions
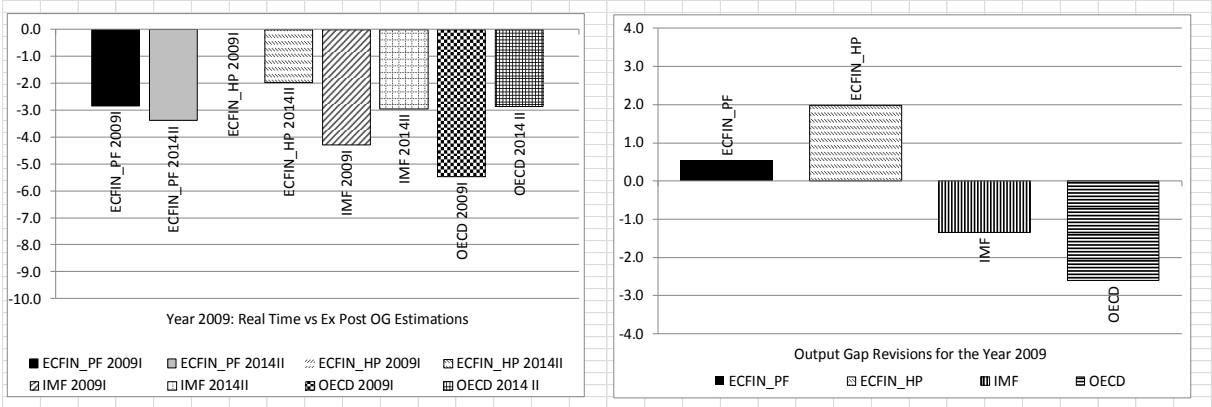
2.2 Assessing the real time reliability of the OECD, IMF & EU output gap methodologies for the period 1980-2010 based on the average spread of the output gap estimates & the number of years in which the sign of the output gap changes : Whilst an exhaustive account of all the different methods which can be used to evaluate the reliability of the output gap estimates produced by national & international organisations is beyond the scope of this paper, what is possible is to use the evaluation criteria utilised by other researchers in carrying out similar comparison exercises to the one being carried out here. One of the most recent attempts to evaluate the reliability of the output gap estimates of international organisations was published by the Bundesbank in April 2014. It focussed on the output gaps produced by the OECD & the IMF methods, but not those of the EU's PF methodology. The two central criteria used by the Bundesbank analysis to evaluate the real time reliability of the OECD & IMF estimates were firstly the average spread of the output gap estimates & secondly the number of years in which the sign of the output gap changed. The Bundesbank analysis was carried out for the G7 economies (US, Japan, Germany, France, UK, Italy & Canada).

The purpose of the current section is to extend the Bundesbank's OECD & IMF analysis to include the output gap results from the EU's PF methodology for the 30 year period 1980-2010 & to use this analysis to assess the relative real time reliability of the output gaps produced by the 3 institutions. In replicating the Bundesbank's approach, for the output gaps produced by the EU's common methodology, we have used the bi-annual Commission

forecast vintages from 2004 until 2014 (a total of 19 vintages). We have restricted the analysis to the three members of the G7 which are also Euro area members, namely Germany, France & Italy. For these three countries, we have looked at the output gap estimates for each year over the two periods covered in the Bundesbank analysis, namely 1980 until 1997 and 1998 until 2010. We calculate the output gap spread for a given year as being the difference between the maximum and minimum value of the output gap & we define whether the output gap has changed signs over the different vintages. In the first column of table 1, the average output gap spread for the two time periods is reported, with the second column giving the number of times that an output gap estimate has changed its sign, at least once, for a certain year in each period. Table 1 shows that if one applies the Bundesbank's approach to the EU's methodology, one would draw the following two main conclusions :

- Firstly, based on the first criterion of real time reliability (i.e. the average spread of the Output Gaps), the EU's methodology is consistently better, & in a significant number of countries / time periods, substantially better than the equivalent IMF or OECD methodologies; &

- Secondly, based on the second criterion of real time reliability (i.e. the number of years in which the sign of the Output Gap changed), the EU's methodology is at least as good or is substantially better for the three Euro Area countries & in 5 of the 6 time periods covered in Table 1 (the only exception being the period 1980-1997 for Italy).

Table 1 : Assessment of the real time reliability of the OECD, IMF & EU PF Output Gap Methodologies (Two assessment criteria : Average spread of Output Gap estimates plus number of years in which the sign of the Output Gap changed)

		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*)	IMF	OECD	EU PF (ECFIN calculations*)
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

* Calculations made using the Bundesbank approach

Section 3 : How can the EU's Production Function approach be further improved to avoid the mistakes of the pre-crisis period (the upswing stage of cycles) ?

Sections 1 & 2 have shown that the EU's method has performed at least as good, & arguably much better, with respect to the real time reliability criterion, when compared with the most widely used alternatives, namely the OECD, IMF & HP filter methods. However, an additional insight from the previous sections is that all methods, including the EU's PF method, did a poor job in the pre-crisis, boom, period, leading to the pursuit of inappropriate pro-cyclical fiscal policies in many Euro area countries in the run up to the crisis. Linked with this failure to spot the significant cyclical excess in pre-crisis growth patterns, the current section looks at three issues with respect to the pre-crisis period :

- Firstly, section 3.1 examines if the method could have done a better, real time, job in picking up the unsustainable nature of the pre-crisis rise in actual GDP – firstly, could the method have produced more realistic estimates of cyclical pressures in the Euro Area's labour & product markets by, for example, producing less pro-cyclical NAWRU estimates or by using more robust capacity utilisation indicators for extracting the trend component of TFP; or secondly, could the method have done better in warning about the unsustainable nature of the pre-crisis investment boom by using trend, as opposed to actual, measures of the capital stock to reflect the cyclical capital stock component of the overall output gap (i.e. capital gaps).

- Secondly, section 3.2 then goes on to try to explain, on the basis of a regression analysis, why the cyclical indicators which the method depends on to make its trend / cycle decomposition were not able to pick up the cyclical excess. It poses the question, was the method unable to fully exploit the information content in these indicators ? In addition, it asks if the method would have performed better, in terms of its short-term reliability, with an augmented basket of cyclical indicators, for example using variables such as changes in the current account balance; the capital gap; real interest rates; or credit growth ?

- Finally, section 3.3 attempts to bring all of the different strands of the debate together in order to come to a balanced interpretation of the lessons which need to be drawn from this period for the EU's PF methodology. In particular, it asks the question whether methodological changes are necessary to avoid such enormous real time revisions in future and, if so, is the way forward the development of better capacity utilisation indicators or the incorporation of financial market variables such as credit growth etc in order to be able to pick up asset driven cycles or "capital gaps". Alternatively, would the correct conclusion be that methodological changes would not have helped given the asymmetric nature of the shock in the pre- & post-crisis periods. One of the hallmarks of the pre-crisis period was the flexible nature of the supply side response (both with respect to capital and labour flows) which easily accommodated (i.e. without wage / price pressures emerging) the boost to demand (most notably from the housing boom) which occurred in various EU economies. This absence of capacity constraints in the pre-crisis period, contrasted with the constrained supply side conditions which emerged in the post-crisis era, provoked by the housing-bust induced breakdown in large parts of the systemically important banking sector (with devastating effects on real economy variables such as output & employment).

3.1 Could the EU's PF method have done better at picking up the unsustainable nature of the demand boom in the pre-crisis period ? The real time performance of the PF method in the pre-crisis period was not very good since the method produced a pre-crisis trend output path which pointed to the Euro Area having a roughly zero output gap whereas subsequent revisions now show a relatively large positive output gap in the years preceding the crisis. Whilst the dramatic nature of the forecast errors with respect to actual GDP was the primary factor driving the poor real time performance of the method in the pre-crisis period, nevertheless it is important to assess whether the PF methodology could have done better in terms of questioning the structural underpinnings of the Euro Area's pre-crisis GDP performance. In particular, did the method do a reasonable job in extracting labour and TFP trends in the run up to the crisis & would the use of a trend measure of the capital stock (rather than the method's use of the actual capital stock) have been more effective in alerting

policy makers to dangerous over-investment patterns emerging in some of the countries in the European periphery.

Did the method do a good job in extracting labour & TFP trends ? : Graph 6 provides a decomposition of the overall output gap for the year 2009 into its labour and TFP elements, with the graph indicating that the shift in the aggregate output gap picture for 2009 between the Spring 2008 and Spring 2009 forecast vintages was due to revisions to both the labour and TFP gaps (with changes to the labour gap explaining about 1/3 of the overall revision & those to the TFP gap the remaining 2/3).

Graph 6 : Decomposition of the Euro Area's Output Gap for 2009 into its Labour & TFP components (Based on the Spring 2008 & Spring 2009 Forecast Vintages)
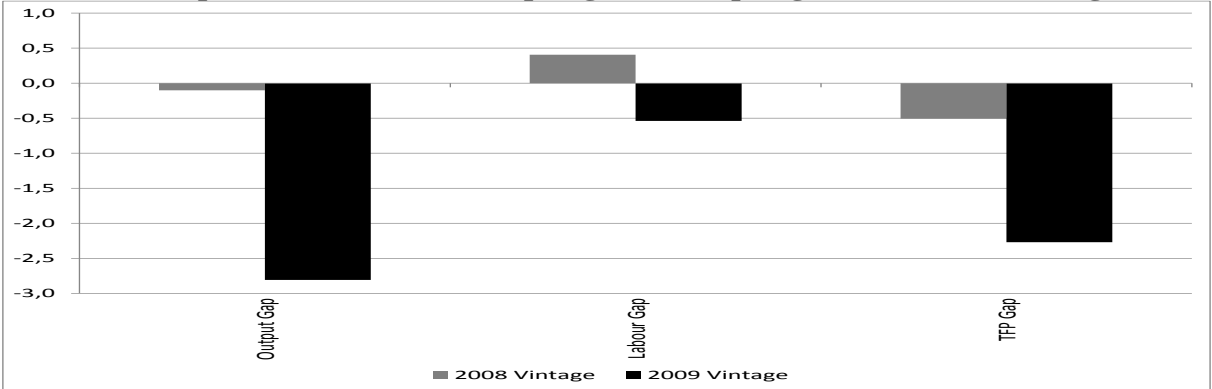
With respect to the labour market gap, the key driver of changes in the overall labour market gap in the 2008 / 2009 period was the unemployment gap (rather than participation rates or hours worked). Many commentators have suggested that perhaps part of the errors made were due to a procyclicality problem with the NAWRU estimates (i.e. changes in actual unemployment & NAWRU rates were tracking each other too closely), with excessively low NAWRU's boosting potential growth rates & wrongly suggesting that the EU's economy was on a sustainable growth path. Given that the NAWRU method used in the pre-crisis period has subsequently been revised to help address this pro-cyclicality issue, an inevitable question to ask is whether the new method could offer some hope that some of the pre-crisis problems could be avoided in the future. In the revised methodology, introduced in Spring 2014, the growth in real unit labour costs –RULC, rather than the change in nominal unit labour costs (NULC) is now being used as the primary labour cost indicator for identifying the NAWRU. However, it turns out that even if the PF method had been focussing on the RULC indicator, it would not have helped in the pre-crisis period. In fact, in some countries, it could have led to even lower NAWRU / higher potential growth rates. Consequently, one must conclude that the new NAWRU methodology would not have led to a reduced level of unemployment gap revisions in the pre-crisis period and one is consequently left with the troubling question as to why the boom period did not show up in either the RULC or NULC labour cost indicators ?

Regarding the TFP gap, most of the real time revisions to output gaps in the pre-crisis period were linked to revisions to the TFP gap. In Spring 2008, ECFIN was using an HP filter to extract trend TFP which may have induced a pro-cyclical bias to the trend TFP estimates. Since Autumn 2010, as explained earlier, ECFIN has been using a new Kalman Filter (KF) based methodology which uses capacity utilisation (CU) indicators to extract the trend TFP component. This new approach has been shown to have better real time properties than the HP filter, with less trend TFP revisions reflecting the reduced risk of a pro-cyclical bias. We saw in section 1 how the new method has helped in the post-crisis period. Since the post-crisis

evidence suggests that using information on CU to control for cyclical factors helped to stabilise trend TFP growth over this period, an obvious question is whether it would have helped to give a more accurate picture of trend TFP developments in the pre-crisis period ? Unfortunately the answer to this question is not very comforting, with the results from applying the new TFP methodology to the Spring 2008 forecast vintage, indicating that we would have got broadly the same assessment of the Euro Area's cyclical position as given by the HP filter method. This disappointing conclusion is not because the new methodology was not working in the way it should have. The real problem was that the CU indicators, on which the accuracy of the new methodology depends, did not indicate any serious overheating in the economy (in fact, the CU indicator for the Euro Area was giving readings in the pre-crisis period which were lower than those indicated for the previous cyclical high around the year 2000). Again, as with the conclusion on the labour cost indicators, the key unanswered question is why the CU indicators were not indicating greater levels of overheating in the Euro Area's economy in the pre-crisis period ? In addition, why do we see such a large asymmetry in the size of the TFP shock experienced in the pre- & post-crisis periods, with the CU indicator falling significantly more in the post-crisis period than it rose in the pre-crisis period ? These are questions which we will return to in section 3.3.

Would a trend measure of the capital stock have been better than the actual capital stock for measuring potential in the pre-crisis period ? Could the method have done better in warning about the unsustainable nature of the pre-crisis investment boom by using trend, as opposed to actual, measures of the capital stock to reflect the cyclical capital stock component of the overall output gap (i.e. capital gaps) ? Since the capital stock is an indicator of overall capacity in an economy and since net investment in any given year is only a tiny fraction of the overall capital stock figures, the PF method does not smooth this series. This approach is soundly based since the method's agreed definition for potential is the amount of output which could be produced if all factors of production were utilised at average or normal capacity rates. Using this definition of capital, the method does not correct for possible deviations of the capital stock from its long run trend level due, for example, to an asset bubble. The method does however fully take into account the slowdown in post-boom potential capital growth because of low investment rates. In addition, the method has the property that the investment rate returns to average sample levels over the medium term (an important assumption for crisis economies) and consequently will pick up the normalisation of investment after the boom-bust cycle.

Given the advantages of the current approach in "normal" times as well as in the bust phase of cycles, the key question is whether we can do anything to improve the way the method handles investment in the boom phase of cycles in order to avoid the type of problems experienced in the pre-crisis period. Since there is no breakdown of the investment to GDP ratio into a structural and cyclical component, there is always the risk that the method will over- or under-estimate potential, especially in extreme asset driven, rather than income driven, cycles. To help alleviate this risk, DG ECFIN has recently developed a type of structural model of investment which establishes an explicit link between investment and its fundamental, long run, determinants, namely changes in trend labour supply and trend TFP. Whilst it is clear that this more structural approach will help anchor long run investment trends, unfortunately use of such an approach over short to medium term time horizons, & more specifically in the pre-crisis period, would not have helped in alerting policy makers to the type of problems provoked by the asset boom experienced in those years.

Why does the trend capital / "capital gaps" approach not help very much over a short to medium term time horizon ? The problem with this approach is not that it does not indicate

the over-investment pattern of the pre-crisis boom period - graph 7 clearly shows that investment was being driven by factors other than the fundamentals of trend employment and trend TFP from 2005 onwards, resulting in a pre-crisis potential growth path which was exaggerated by a number of unproductive, leverage-induced, investments. The problem is that although the capital gap was clearly large, since the weight of capital in overall output is relatively small, this ensures that its impact on the Euro Area's overall output gap in the pre-crisis period is also correspondingly small. Even if we had been using a trend capital measure in Spring 2008 & we had estimated the capital "gap", we would still not have produced an overall output gap for 2008 which is substantially different from the one actually published. We should perhaps not be surprised by this conclusion since the OECD was using a trend capital measure in the pre-crisis period and their measure of the "capital gap" did not show up "overheating" issues.

Graph 7 : Pre-crisis divergence between the growth rates of potential output and of the capital stock
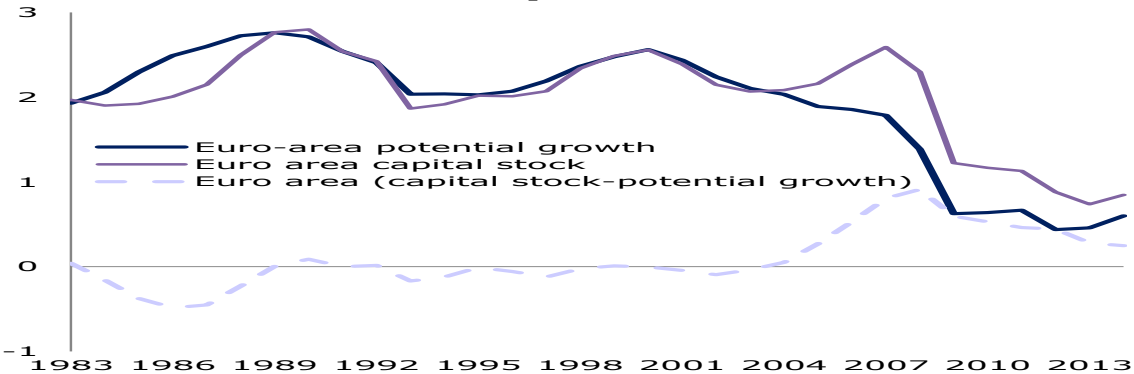
3.2 Why were the cyclical indicators which the method depended on in the pre-crisis period to make its trend / cycle decomposition not able to pick up the cyclical excess ? Could the use of an augmented basket of indicators have helped in improving the method's short run reliability ? : The current section examines the overall fragility of the EU's method, most notably over the pre-crisis period, & tries to answer two fundamental questions :

- Firstly, in overall terms, is the PF method failing to fully exploit the information content in the set of cyclical indicators which it currently uses & would an augmented basket of indicators lead to greater real time reliability ?
- Secondly, in the years preceding the crisis, would the method have performed better (in terms of reducing year-to-year revisions) if it had been using an augmented basket of indicators, including for example changes in the current account balance; an indicator of the capital / investment gap; real interest rates; or credit growth.

In order to answer these questions we have studied the year-to-year revisions in the output gap calculations in more detail. We have analysed whether the revisions in the real time output gap can be linked to some financial and/or economic indicators. We focus on the Spring forecasts, so that all of the information used is final for the year being analysed (which would not be the case if we used the Autumn forecasts for that year) and, as shown below, we compare each year's output gap estimate to the estimation of the output gap for that same year in the forecast of the following year :

$$d(YGAP)_t = YGAP_{t|Spring(t+2)} - YGAP_{t|Spring(t+1)} = \alpha + \beta * IND_{t|Spring(2015)} + \varepsilon_t$$

The indicators are taken from the Spring 2015 forecast, or from the available data in June 2015. If a link can be found, this suggests that if we would have added these indicators into the output gap calculations, revisions could have been smaller in the past (and consequently might be smaller in the future). In a subsequent step we recalculate the output gap after adding in the estimated revisions which could be expected and compare this with the output gap as it turned out to be in the following year. The selected indicators are shown in table 2. In addition we include two binary variables which reflect the impact of the two major changes in the methodology since its inception in 2002 (i.e. the shift towards using hours worked in 2005 and the introduction of a Kalman Filtered based TFP approach in 2010).

Table 2 : Indicators possibly related to the revisions

Indicator	Calculation
Set of indicators currently used (included in Model 1)	
Growth in GDP	Growth rate of GDP
Capacity utilisation (TFP)	Economy wide capacity utilisation rate
Wage indicator (NAWRU)	Growth rate of real unit labour costs (RULC)
Augmented basket of indicators	
- Set of financial indicators (included in Model 2)	
Growth in credit (total)	Growth rate in total loans
Interest rates	Long term interest rate
- Set of other additional indicators (included in Model 3)	
Current account balance	CAB_t/GDP_t
Capital gap	Calculated based on the PF method*

* The capital gap is defined as the difference between the actual capital stock and a capital stock consistent with TFP and employment trends (and the assumption of constant capital costs)

Q1 Is the PF method failing to exploit the existing set of cyclical indicators & would an augmented basket of indicators help ? : Table 3 shows the results of the models that have been tested. Whenever an indicator is not significant, (with p>0.10) for a certain model, n.s. (not significant) is included in the table cell. A blank cell means that this variable was not included in that specific model. All the models are estimated using Panel OLS with White corrected standard errors to deal with possible period autocorrelation.

The first model estimates the revisions based on a constant, the two binary variables and those economic indicators which are already included in the output gap calculations, namely actual GDP & the two cyclical indicators used in calculating the TFP & NAWRU components of potential GDP (i.e. the capacity utilisation (CUBS) indicator & the wage indicator (change in RULC's)). We can see in Table 3 that all of the variables in Model 1 can collectively only explain about 21% of the year-to-year revisions over the period 2001-2013. In addition, a sensitivity analysis shows that if Model 1 only included the actual GDP indicator, this single indicator would explain 16% of the overall, year-to-year, 2001-2013 revisions (equivalent to ¾ of the total explanatory power of Model 1). It is interesting that whilst the results for Model 1 show that the actual GDP variable is significant, this only applies when the actual GDP growth rate is positive i.e. if the actual growth in GDP is positive, the output gap estimate is revised upwards in the following year. This is consistent with the view that the current method underestimates positive output gaps in boom periods (i.e. it produces excessively optimistic potential growth rate estimates in "good economic times"), with no such bias evident in "bad economic times". Regarding the indicators used in the potential growth calculations, changes in the RULC indicator (which has only been exploited in the method since Spring 2014) is not significantly related to the output gap revisions. The CUBS

indicator has been used to estimate trend TFP since 2010 and is shown to be significant. The significance of the RULC & CUBS indicators are however sensitive to the list of indicators included in the overall model. The only consistently robust result, from the model 1 simulations, is that the actual GDP indicator can explain a significant part of the year-to-year output gap revisions. Consequently, one can conclude that less historical revisions to the actual GDP growth rate figures, as well as more accurate actual GDP forecasts would help in reducing real time output gap revisions.

Model 2 augments Model 1 by adding financial variables to the current basket of indicators. Adding these variables makes the wage indicator (i.e. the change in RULCs) significant. Both financial indicators (growth in total credit and long term interest rates) are significant, but their explanatory power is limited since the adjusted R^2 only increases by 3ppt. One could hypothesize that financial indicators might be more important in explaining the revisions in the output gaps of those countries that suffered the most from the crisis. Although Rabanal and Sanjani (2015) suggest that financial frictions in Greece, Ireland, Italy, Portugal & Spain amplified fluctuations in the measurement of the output gap, we find that for this group of countries both financial variables become insignificant. When we confine the selection to Ireland and Spain, the long term interest rate remains significant but the growth in credit indicator becomes insignificant. This result also holds if only financial variables are used in the model.

Model 3 completes the analysis by adding a number of additional economic variables which could theoretically have some explanatory value but which are currently not used in the output gap calculations. The model shows a further improvement in the adj. R^2 of 3ppt but unfortunately neither the capital gap nor current account balance (CAB) variables are statistically significant. We also analysed whether the CAB is significant for small open economies and found that the t-statistic of the indicator increased to 1.4 (which is still however not significant at $p<0.10$) and that it leads to a negative coefficient.

Table 3 : Results of the Panel OLS estimations (2001-2013, 27 EU countries)

Variable name:	Model1	Model2	Model3
Constant	-	-	-
Use of Kalman Filtered TFP	-	-	-
Use of hours worked	+	+	+
Economic variables currently used in Output Gap Estimation Process			
Positive growth in GDP	+	+	+
Negative growth in GDP	n.s.	n.s.	n.s.
CUBS indicator	+	+	+
Change in Real Unit Labour Costs (RULC)	n.s.	-	-
Additional financial variables			
Growth in total credit		+	+
Long term interest rates		**-**	**-**
Additional economic variables			
Capital Gap (kgap)			n.s.
Current Acount Balance (CAB/GDP)			n.s.
R^2	0.23	0.27	0.30
adj. R^2	0.21	0.25	0.28

Q2 Would the method have performed better in the pre-crisis period (on a year-to-year basis) with an augmented basket of indicators? To answer this second question, we use the Panel OLS model for the period 2001-2005 to forecast the revisions for 2006-2008 and see whether the revisions would have been smaller. This analysis shows that the improvement in the output gap revisions is small and for some countries there is no improvement at all (Czech Republic, Estonia, Luxemburg, Poland, Slovenia and Slovakia). The most important point from this analysis is that no matter which of the three models one uses, the improvement in terms of the real time reliability of the resultant output gaps is very limited. For example, at the level of the EU as a whole, if any of the 3 models were retrospectively applied to the 2006-2008 period, the improvement in the reliability of the, year-to year, output gap estimates would only be of the order of 10 to 20 percent, & consequently 80 to 90 percent of the pre-crisis output gap estimation errors would still have been generated by the official methodology.

3.3. What are the lessons to be drawn from the pre-crisis problems with the EU's output gap methodology – can these problems be resolved by better indicators / methodological changes or do they reflect the atypical characteristics of the crisis & its asymmetric effects in the boom & bust phases of the cycle ? Section 3.2 showed that the better exploitation of the PF method's cyclical indicators & the inclusion of an augmented basket of indicators would have only marginally helped to improve the PF methods pre-crisis performance (i.e. a gain in year-to-year reliability of only 10% to 20%). In addition, in trying to explain what went wrong with the method in the pre-crisis period, the analysis in section 3.1 raised questions as to why neither the labour cost nor CU indicators rang alarm bells as to the unsustainable nature of pre-crisis trends. It also showed that even if we had been using a trend capital stock concept (rather than the actual capital stock), the additional impact of the "capital gap" on the Euro Area's overall output gap in 2008 would have been very small. Given these results, the current section asks to what extent changes to indicators / methodologies could realistically help in the event of future, finance-induced, crises or, alternatively (as strongly implied by sections 3.1 & 3.2) whether the search for new indicators / methodologies is largely an academic issue given the asymmetric effects of financial shocks during the pre- & post-crisis periods.

a) Perspective 1 : Changes to indicators / methodologies might improve the year-to-year reliability of the method in future "boom" periods : Regarding indicators, given the amount of work done recently on the relative performance of different labour cost indicators & the fact that the insights from this research are now reflected in the methodology, the only area where future indicator changes could realistically help is with respect to capacity utilisation (CU). Better measurement of CU developments in a number of specific service industries & in construction could add to the accuracy of the TFP trend / cycle decomposition exercise.

Regarding methodologies, whilst the more structural model of investment described earlier will help anchor long run investment trends, we are still left with what to do over the short to medium term time horizons (i.e. up to T+5), since there is still no mechanism in place which can deal with the type of asset booms experienced in the pre-crisis period. In this context, new research carried out by the BIS suggests that using information about the financial cycle (real interest rates, credit growth and property prices) to produce "finance-neutral" output gaps could help in alerting policy makers to the role which asset prices are playing in the dynamics of business cycles. The BIS work stresses the importance of incorporating information about the financial cycle / financial instability into the potential output calculations. This is a very

different approach to estimating potential output compared with the EU's current approach which focusses on the inflationary consequences of changes in the level of actual output. If there are no signs of inflationary pressures in an economy, the EU's method assumes that actual & potential output are more or less equal. The BIS approach stresses the Minsky idea that "stability is destabilising", namely that an economy may be on an unsustainable path even under conditions of low or stable inflation (e.g. in the pre-crisis period) & only measures of potential output / "finance-neutral" output gaps which incorporate the impact of financial factors are capable of signalling when an economy is on such an unstable path.

Whilst the BIS approach has clearly a number of appealing features (including the fact that it produces large positive output gaps in the pre-crisis period, with only a small level of revisions), including it directly in the PF method does not seem sensible at the current time since, on the basis of the BIS results, the approach could lead to too many "false alarms" / "inconsistent signals" (e.g. at what point should policy makers get concerned about the growing share of credit as a percentage of GDP or of rising house prices). In addition, the BIS method produces implausibly small, and rapidly closing, output gaps in the post-crisis period. In terms of the selection of variables used by the BIS and their treatment as cyclical versus structural indicators, there is no distinction made in the BIS analysis between the impact which these indicators are having on both the cycle and on potential – everything is attributed to the cycle. This is a questionable approach since the financial market variables used are undoubtedly affecting both the cycle and the trend. Finally, whilst in general not having revisions is a good thing, the observation of having only small revisions for finance-neutral output gap estimates may not necessarily suggest a good real time performance since it could be indicative of having missed important signals.

Despite the many failings of the BIS approach listed above, DG ECFIN still believes that a variant of the BIS approach, which is currently being developed, could serve as a useful complement to the current approach. More specifically, financial market variables could be effectively used as a type of "early warning" signal of bubbles in capital formation (i.e. large cyclical deviations relative to the long run trend level for the capital stock) or alternatively as a signalling mechanism for abnormally low levels of capital accumulation in the bust phase of the cycle. In combination with the other cyclical indicators used in the PF method, such as capacity utilisation and different labour cost indicators, the inclusion of financial market variables could boost the overall information set available to policy makers on the cyclical drivers of growth (although the results from section 3.2 would caution against excessive optimism as to the additional gains to be made).

b) Alternative Perspective : Changes to indicators / methodologies will not help since the last "boom and bust" cycle was not typical : The insensitivity of the labour cost and CU indicators to what, in hindsight, was an obvious, finance-induced, boom in the EU provides one more piece of evidence in favour of the belief that the last boom-and-bust cycle had a somewhat atypical nature, as it did not emerge in the traditional sectors of the economy but had much more to do with developments in sectors which are especially sensitive to credit, such as the construction sector, with knock-on implications for the financial sector / asset markets. Whilst the importance of the financial sector in the economy has been significantly increasing, its share of total output remains relatively small in most EU countries. For this reason, even large (and possibly persistent) deviations of capacity utilization from its natural level in this specific sector may not be picked up in the more aggregated measures of capacity utilization which are used in the methodology. Whilst it is in the nature of fluctuations in the financial sector to have a non-proportional impact on the

rest of the economy and to amplify shocks coming from elsewhere, they are unlikely, to directly influence the TFP gap and hence should not be expected to be picked up by any currently available, or indeed future, TFP trend extracting econometric methods.

Consequently, this alternative view suggests that since our current potential output methodology focusses on capacity constraints, in the absence of evidence pointing to such constraints, it is unrealistic to expect such an approach to be able to decipher when an economy is on a, finance-induced, asset-price driven, unsustainable path. This difficulty is not only due to the stylised fact that financial / investment booms tend to coincide with positive supply side shocks but also that the demand expansions associated with such booms tend to weaken supply constraints. This weakening in supply constraints is particularly apparent with respect to labour supply developments, with strong periods of growth boosting domestic participation rates & the inflow of foreign workers. In addition, the unsustainable nature of many financial booms has more to do with the sectoral misallocation of financial resources (towards construction & other, credit-sensitive, sectors) rather than to the generation of economy-wide capacity constraints.

Concluding Remarks

This paper has provided evidence to support the EPC / ECOFIN Council decision in 2002 to adopt the production function (PF) methodology for estimating output gaps as the "commonly agreed" reference method to be used in the EU's fiscal (&, by default, also structural policy) surveillance procedures. Whilst uncertainty will always be a feature of output gap calculations, this paper has shown that in relative terms the real time reliability of the EU's calculation method has been superior to the methodology used in the EU up to 2002, namely the HP filter, as well as outperforming the OECD's methodology in every single year of the 2004-2014 period & that of the IMF in all years except one.

Not only was the EU's methodology "top of its class" over the period as a whole (in terms of relative reliability), it also outperformed the HP, OECD & IMF methods around the crucial turning point of the crisis. In Spring 2009, there was a very wide range of estimates for the Euro Area's output gap in the year 2009, ranging from an economically implausible zero estimate from the HP filter to -4.3% / -5.5% respectively from the IMF / OECD, with the EU's PF method in the middle of these estimates (-2.8%). Five and a half years later in Autumn 2014, the ex post output gap estimates for 2009 for the Euro Area now produced by the IMF & OECD methods are almost identical to the real time Spring 2009 estimate produced by the EU's methodology i.e. -2.9% versus -2.8%.

In addition, these relatively large OECD & IMF output gap revisions for the specific year 2009, compared with those of the EU's method, suggest that the respective IMF / OECD methodologies made a fundamentally different assessment of the impact of the financial crisis. The extremely large negative output gaps predicted by the IMF & the OECD in Spring 2009 for the year 2009 were consistent with a view at that time that the effects of the financial crisis on potential output would be relatively limited & temporary in nature. The much smaller negative gaps produced by the EU's methodology suggested a less benign interpretation of the crisis, namely that the impact on potential would be much more significant & more prolonged (a view which has been subsequently confirmed by economic developments in the post crisis period). This interpretation is supported by a recent Bundesbank analysis which stressed that the drop in actual output in the post-2009 crisis period was initially interpreted by the IMF & the OECD as a cyclical phenomenon. It was not

until the economic recovery proved weak that it became clear that the preceding upward movement in potential was in fact unsustainable.

Whilst the relative performance of the EU's methodology suggests that it is "state of the art", unfortunately its absolute performance, especially in the pre-crisis period from 2006-2008, leaves a lot to be desired. Whilst the EPC / ECOFIN Council can correctly claim that the EU's method easily outperformed the HP, OECD & IMF methods in terms of reliability over the pre-crisis period, however this provides little comfort from a fiscal surveillance perspective since it is now clear that big output gap (& consequently structural budget balance) errors were made over this period. To put the level of these errors into perspective, revisions to the PF's output gap estimates in the pre-crisis period were roughly five times greater than those of the post-crisis, 2009-2014, period. This is a particularly humbling statistic given that one of the EU's primary motivations in 2002 for moving away from the HP filter to the PF approach was the expectation of reduced levels of procyclicality (especially in the upswing stage of cycles) which were expected to result from alleviating the risk of end-point biases & from the prospect of agreeing a new multivariate estimation methodology to reduce the cyclicality of the trend TFP estimates.

Section 3 of the present paper was devoted to examining the poor pre-crisis performance of the EU's methodology. It showed that economic activity appeared deceptively robust in the run-up to the crisis, with the "great moderation" clearly masking the underlying financial vulnerabilities in the system. With such abnormal, asset price driven, business cycles, deciphering cyclical from structural developments was especially treacherous. In hindsight, the PF method was clearly lulled into a false sense of security by the prevailing conditions of low / stable inflation and failed to pick up the unsustainable nature of the finance induced boom which occurred. The amount of spare capacity in the pre-crisis period was large, with supply conditions being relatively elastic & with increases in demand easily accommodated without showing up in inflation (a view in keeping with the "secular stagnation" ideas of Summers & others). Of course, as we now know, the pre-crisis orthodoxy that stable inflation was a sufficient condition for economic stability was an illusion. The apparently benign real economy developments masked the growing debt levels in the system which eventually led to the "Minsky moment" in Autumn 2008 when the financial system imploded.

Whilst a clearcut answer to what happened in the pre-crisis period is not yet possible, nevertheless section 3 offers some hope that the extent of the mistakes made by the EU's PF methodology in the years running up to the crisis in 2008 will not be repeated in the next upswing phase :

- Firstly, the improvements made to the EU's PF methodology in the post-crisis period (i.e. the introduction of the KF TFP method in 2010 & the new NAWRU method in 2014) will undoubtedly help in reducing the risk of procyclical output gap estimates being produced in the next upswing phase. In this context, it is encouraging to note that the NAWRU, & especially the TFP, changes to the method appear to be paying off in the post-crisis period where the widely repeated criticism that the EU's methodology is too procyclical is simply not credible since in general the method has revised down, not up, potential after observing the new facts on actual growth developments in the post crisis period (suggesting that it could be the asymmetric role played by finance in the upswing phases of cycles which bears primary responsibility for the past procyclicality errors in the method). In addition, the strong overall performance of the EU's method in the post-crisis period, with average revisions of

less than ½% of GDP, the lowest since the method was introduced in 2002, suggests that the method is making progress, justifying the decision of the EPC in 2002 to set up a working group (i.e. the Output Gap Working Group) explicitly dedicated to improving the output gap estimates used in EU policy surveillance procedures.

- Secondly, section 3 discussed the issue of integrating financial market variables into the method & whilst the gains from doing so may not be as great as some commentators suggest, it is accepted that, in future upswings, the production of finance-neutral output gaps could serve as a useful complementary warning system for policy makers. In this regard, an important lesson to be learnt is that more work is needed on understanding the link between finance & sustainable growth. More specifically, we have to find ways of distinguishing in the method between "normal" times (i.e. when the activities of the financial sector are consistent with economic stability) and "boom" times (when high leverage & high risk taking leads inevitably to asset price bubbles & to fundamental changes in the macro context). We are still a long way from being able to make such distinctions but, using the BIS work as a starting point, we hope to gradually complement the official methodology, over the coming months & years, with a series of indicators which can alert policy makers to the type of financial sector instabilities which were arrogantly ignored in the pre-crisis era.

- Finally, & linked with the preceding points, some comfort can be taken from the atypical nature of the last boom-&-bust cycle, with the financial market reforms introduced in the post crisis period lessening the likelihood of an equivalent atypical cycle in the near future (as long, of course, as policy makers remain vigilant). Since the shock was centered on the systemically important finance sector, its effects in the pre- & post-crisis periods were asymmetric in nature, with the gains from the construction boom in the pre-crisis years being more than offset by the output losses experienced in the post-crisis period. The flexible supply side response to the boost in demand in the pre-crisis period led the method to conclude that the increase in GDP was sustainable. There was a clear asymmetry in the demand response in the downturn phase, with demand being constrained due to the implosion of the banking sector & the need for public and private sector deleveraging.

To conclude, the excessive optimism of the pre-crisis period with respect to underlying growth trends in the EU, underlines, yet again, that handling the upswing stage of cycles is the ongoing "achilles heel" of mainstream output gap estimation methods, & why policy makers are right to continuously warn that so called "good times" is where the most significant fiscal & structural policy errors are made (note : the structural policy errors are mainly ones of omission rather than commission, using the cover of "good times" for policy inaction). Consequently, whilst one can legitimately argue that the pre-crisis period was a once in a generation financial shock & that the real time reliability performance of the EU's commonly agreed method has been exceptionally good around the turning point of the crisis & in the post-crisis periods, nevertheless it is hoped that the EPC will recognise the importance of its OGWG continuing to improve the EU's commonly agreed methodology, with a particular focus on attenuating the procyclicality risks in the upswing phase of cycles.

References

Borio, C., Disyatat, P., & Juselius, M. (2013), "Rethinking potential output: Embedding information about the financial cycle". *Monetary and Economic Department. Bank for International Settlements.*

Bundesbank (2014), "On the reliability of international organisations' estimates of the output gap", *Deutsche Bundesbank Monthly Report (April).*

Cayan, J-P, & van Norden, S (2005), "The reliability of Canadian output gap estimates", *North American Journal of Economics & Finance 16, 373-393.*

Cottarelli, C., Giammusso, F., & Porello, C (2014), "Fiscal policy & potential output – the fall of actual & potential GDP after the crisis", *LAVOCE.*

Cottarelli, C (2015), "Potential growth rates & the working of the SGP fiscal rules", *VOX (March).*

Darvas, Z (2013), "Mind the gap & the way structural budget balances are calculated", *Bruegel Blog (October).*

Darvas, Z (2015), "Mind the gap (& its revision)", *Bruegel Blog (May).*

Economic Policy Committee (2001), "Report on potential output & the output gap" *(EPC/670/01/EN).*

Financial Times (20 November 2014), "Italy accuses Brussels of shaky accounting"**.**

Hers, J & Suyker, W (2014), "Structural budget balance : a love at first sight turned sour" *CPB policy brief.*

Nelson, E & Nikolov, K (2003), "UK inflation in the 1970's & 1980's : the role of output gap mismeasurement", *Journal of Economics & Business 55, 353-370.*

Orphanides, A & van Norden, S (2002), "The unreliability of output gap estimates in real time", *Review of Economics & Statistics 84, 569-583.*

Planas, C, Roeger, W & Rossi A (2010), "Does Capacity Utilisation help estimating the TFP cycle ?", *European Economy, Economic Papers 410**.*

Rabanal, P & Sanjani, M (2015), "Financial Factors : Implications for Output Gaps"; *IMF Working Paper 15/153.*

Annex 1 : Euro Area - Absolute Average "Forecast-to-Forecast" Revisions – HP filter 100 versus EU's Production Function (PF)

Absolute Average "Forecast-to-Forecast" Revisions
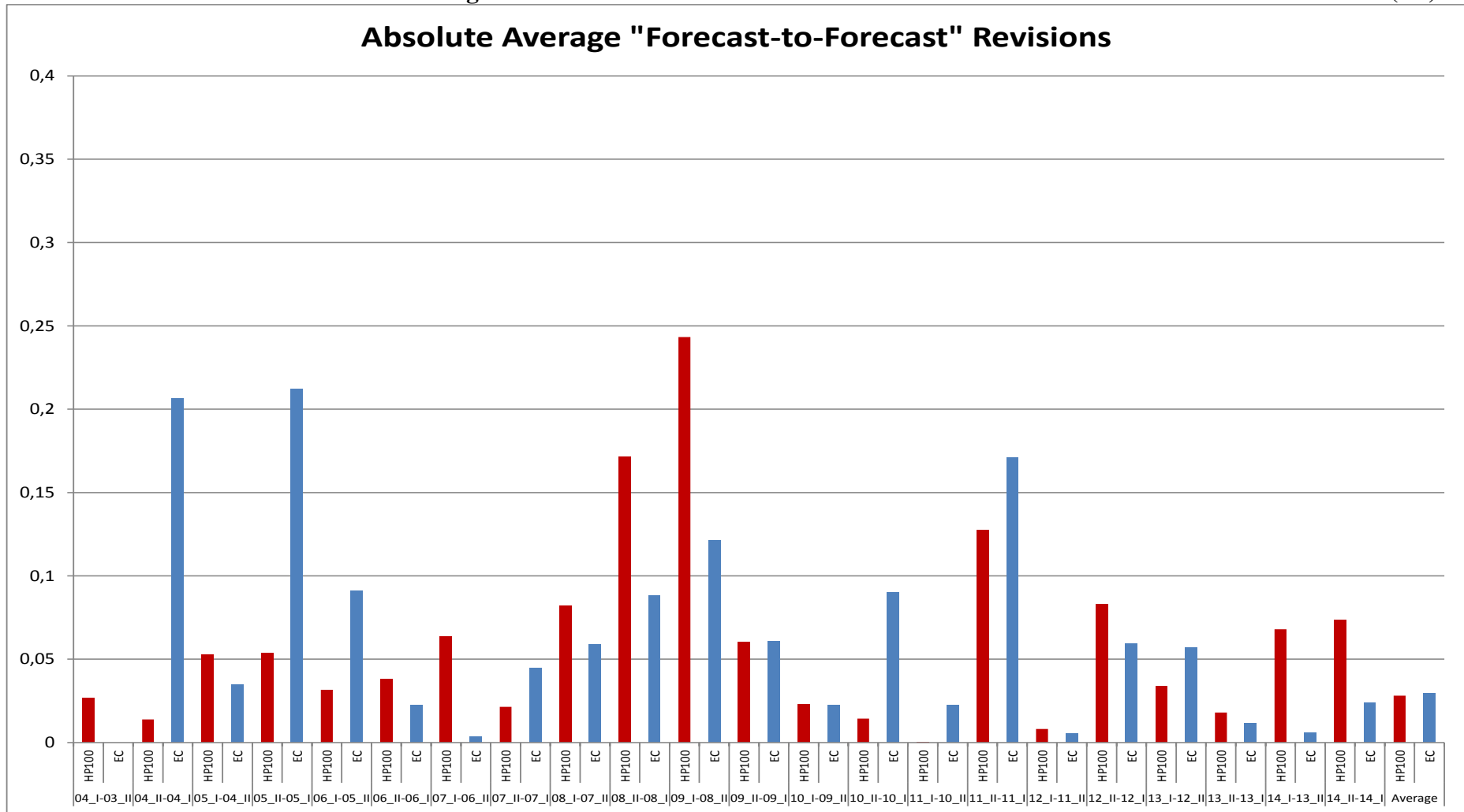
Euro Area - Absolute Average "Forecast-to-Forecast" Revisions – EU's Production Function (PF) versus OECD & IMF Methods

Absolute Average "Forecast-to-Forecast" Revisions