An assessment of the relative quality of the Output Gap estimates produced by the EU's Production Function Methodology

K. Mc Morrow, W. Roeger, V. Vandermeulen and K. Havik

EUROPEAN ECONOMY

Economic and Financial Affairs

European Economy Discussion Papers are written by the staff of the European Commission's Directorate-General for Economic and Financial Affairs, or by experts working in association with them, to inform discussion on economic policy and to stimulate debate.

The views expressed in this document are solely those of the author(s) and do not necessarily represent the official views of the European Commission.

Authorised for publication by Anne Bucher, Director for Growth, Competitiveness and Structural Reforms.

Europe Direct is a service to help you find answers
to your questions about the European Union.

Freephone number (*)**:**
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on http://europa.eu.

An assessment of the relative quality of the Output Gap estimates produced by the EU's Production Function Methodology

K. Mc Morrow, W. Roeger, V. Vandermeulen and K. Havik

Abstract

This paper assesses the performance of the EU's production function (PF) methodology for estimating output gaps since its introduction in the EU's policy surveillance procedures in 2002. It looks at how the methodology has performed relative to the method used up until 2002 (i.e. the Hodrick Prescott filter), with respect to its ability to track the euro area's business cycle. It also compares the PF method with the equivalent OECD and IMF methodologies in terms of its stability, real-time reliability and financial crisis performance. The analysis shows that the EU's PF method has performed better than the HP filter and the equivalent OECD & IMF methods. The results consequently strongly support the 2002 ECOFIN Council decision to adopt the PF method as the EU's 'commonly agreed' reference method. Nevertheless, whilst the PF method has clearly done well in relative terms since it was first introduced, the analysis also recognises the absolute size of the output gap errors made by all of the methods in the pre-crisis period. These errors underline 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.

Contacts: K. Mc Morrow, W. Roeger, V. Vandermeulen and K. Havik, European Commission, Directorate General for Economic and Financial Affairs, Fiscal Policy Directorate, B-1049 Brussels, Belgium; Kieran.Mcmorrow@ec.europa.eu, Werner.Roeger@ec.europa.eu, Valerie.Vandermeulen@ec.europa.eu, Karel.Havik@ec.europa.eu.

CONTENTS

INTRODUCTORY REMARKS[1]

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 (see Box 1). 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 for all of the EU's 28 Member States (due to their usage in the calculation of structural budget balances) 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 stability & 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

---

[1] Whilst the analysis in the main section of the paper focusses only on the euro area, the evidence provided in annexes 1 & 2 show that the euro area conclusions also apply to the majority of the EU's 28 Member States.

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 method 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 (see Annex 7).

- 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 stability & 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. 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 business 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 the 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 two 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 less stable & more unreliable than those of the OECD & the IMF ? (Section 2)

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***)

1. COMPARISON BETWEEN THE PF METHODOLOGY AND THE HP FILTER APPROACH

As mentioned earlier, 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 evaluates the **revision** properties of both the PF and HP methods since 2002 on the basis of a number of criteria, such as :

- Firstly, looking at the **short-term stability of the estimates**, i.e. how much are the output gap estimates revised from one forecast to the next. In terms of the EU's fiscal policy surveillance exercises conducted under the Stability & Growth Pact, this "forecast to forecast" revisions criterion is possibly the best one to use if one is solely interested in evaluating the methods with respect to their implications for the change in the structural fiscal balance.

- Secondly, by comparing the **long-term, real time, reliability / accuracy of the methods** – this criterion is the one to focus on if the aim is to reduce policy errors in real time (e.g. to avoid the type of errors made in the aftermath of the 1970's oil price shocks);

- Thirdly, by assessing the **performance of the methods during the financial crisis** (the 2009 crisis was arguably the biggest test of output gap estimation methods since the 1929 crash); &

- Finally, by examining the **economic plausibility of the output gap estimates (i.e. the "smell test")**, in particular whether the estimates are consistent with other cyclical indicators & 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 or not.

1.1 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.

Consequently, forecast-to-forecast stability of the output gap estimates is an important criterion when evaluating a method's performance. Graph 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 of the Commission's forecast vintages from 2002-2014, of only 0.02 / 0.03 pp respectively). Whilst the relative stability performance of both methods is similar for the 2002-2005 & the 2006-2008 sub-periods, the HP filter method does better than the PF method over the post-crisis years (2009-2014). Regarding these post-crisis differences, the additional graphs for the Euro Area shown in Annex 1 indicate that the HP filter's better performance is somewhat distorted by the large offsetting revisions around the crisis period (with, for example, a positive revision of 0.3 for the Spring 2009 forecast vintage, followed

by a negative revision of 0.3 in the Autumn 2009 vintage[2]). In overall terms, given that the average HP & PF revisions, over all of the 2002-2014 vintages, are so small & so similar (0.02 versus 0.03), it is not justifiable to use the short run stability criterion to discriminate between both of these methods[3].

Graph 1 : Short Term (Forecast-to-Forecast) Stability of the PF & HP filter Output Gap Estimates (Euro Area : 2002-2014) (Period Average Revisions to Estimates)
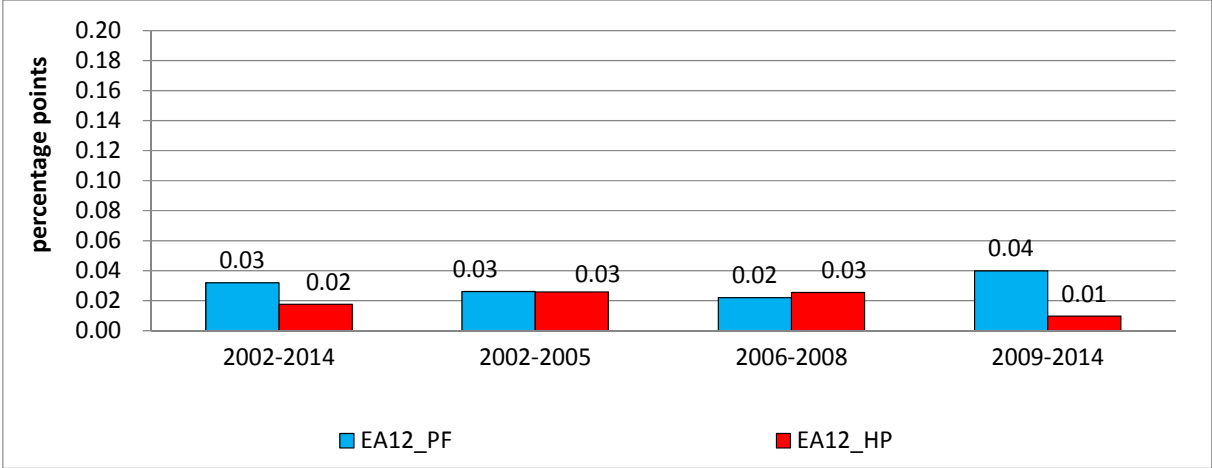

1.2 Long-term real-time reliability of PF and HP estimates : Whilst relatively stable short-term output gap estimates are important, stability should not come at the expense of long-term real-time reliability. In this sub-section we compare the real-time and *ex post* PF and HP filter estimates for the euro area, produced by the PF & HP filter methodologies for the period 2002-2014, using the Commission's autumn forecast vintages. The autumn 2014 vintage is used as the *ex post* reference, with the size of the differences between the real-time and *ex post* estimates used as an indication of the methods' relative reliability[4]. The results are shown in Graph 2. The main conclusions are:

- Firstly, for the 2002-2014 period as a whole, the overall degree of revisions differences between both methods was similar, with average absolute revisions of around 1% point for both methods.

- Secondly, whilst the average 2002-2014 revisions properties of both methods may be broadly equivalent, this average performance hides a much more interesting and nuanced dynamic picture when one looks at the evolution of revisions over time. Graph 2 suggests that the relative reliability of the PF method has gradually improved over time. Whilst 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), 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), the PF method again

---

[2] Please note that these offsetting effects are not a feature of the long run reliability graphs since average **absolute** differences are used for these graphs.

[3] The small forecast-to-forecast revisions for both the HP & PF methods confirm the view of the 2001 EPC report that whilst the **level of the output gap** is highly uncertain, **changes in the output gap** estimates between successive forecast vintages (i.e. forecast-to-forecast revisions) are much more stable & consequently so are changes in structural fiscal balances.

[4] Note : The real-time versus ex-post test of output gaps is much tougher than the forecast-to-forecast test.

outperformed the HP filter (due *inter alia* to the introduction of the new TFP method in Autumn 2010)[5].

- Finally, whilst a method's relative revisions performance is important, policy-makers should not lose sight of the absolute size of those revisions. Graph 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[6]. 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 by both estimation approaches. 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 output gap revisions in those years that were multiples of those made in the 2002-2005 and the 2009-2014 periods.

Graph 2 : Long Term, Real Time, Reliability of the PF & HP filter Output Gap Estimates (Euro Area : 2002-2014)
(Period Average Absolute Revisions) (Real Time versus Ex Post Estimates)
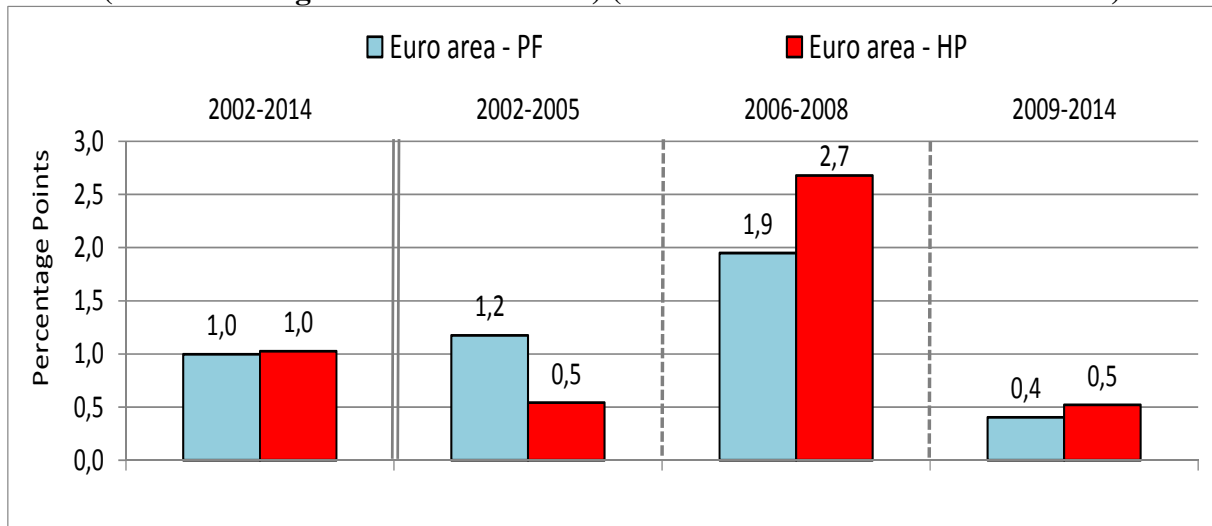
1.3 Performance of methods around the turning point of the financial crisis : Graph 3 shows enormous differences between the output gap estimates produced by the two methods around the turning point of the crisis, with the HP filter pointing to a zero output gap for 2009 and 2010 in the spring 2009 forecast vintage, compared with an average of roughly -3 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 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 1/4 % 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 the PF's performance around this crucial cyclical "turning point".

---

[5] In addition, 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.

[6] The large absolute revisions highlight the optimistic bias in both methods in the pre-crisis period.

**Graph 3 : Turning Point of Financial Crisis 2008-2010 (Spring 2009 Forecast)
PF versus HP filter Output Gap Estimates (Euro Area)**
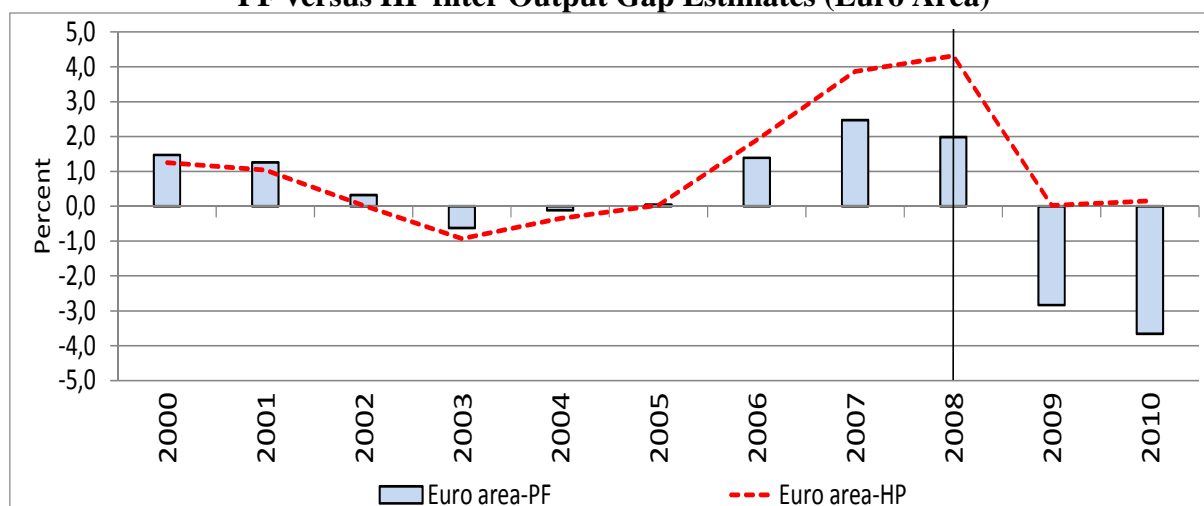
1.4 Economic plausibility of PF and HP estimates : In addition to the introduction of hours worked in 2005, EU policy-makers included a new TFP estimation method in the PF framework in 2010[7]. This helped further enhance the credibility of the overall methodology. For example, a comparison of the real-time 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[8].

The impact from the introduction of the KF TFP method is particularly visible in terms of the **level of the overall output gap & consequently its economic plausibility**. Graph 4 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 until autumn 2010; and

---

[7] The old HP filtered TFP approach was replaced by a Kalman Filter (KF) approach in the official methodology in 2010. 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 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. 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.

[8] *What has been the impact of the introduction of the new TFP methodology in terms of Output Gap revisions for the Euro Area as a whole over the period 2010-2014 ? :* If one 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, for the post crisis years 2010-2014, one sees 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. In the first years after the crisis (2010-2011), the introduction of the Kalman filter TFP did lead to significant reliability gains for the euro area 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. 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 (see Mc Morrow et al -2015).

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

Graph 4 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. From 2006, the HP output gap is generally higher than the PF output gap, with the PF output gap using the HP TFP component often in the middle of the overall set of results. Graph 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 post crisis years. The gain 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 4 : Average Size of Euro Area's Total Economy Output Gap (Autumn 2014 Forecast) 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)
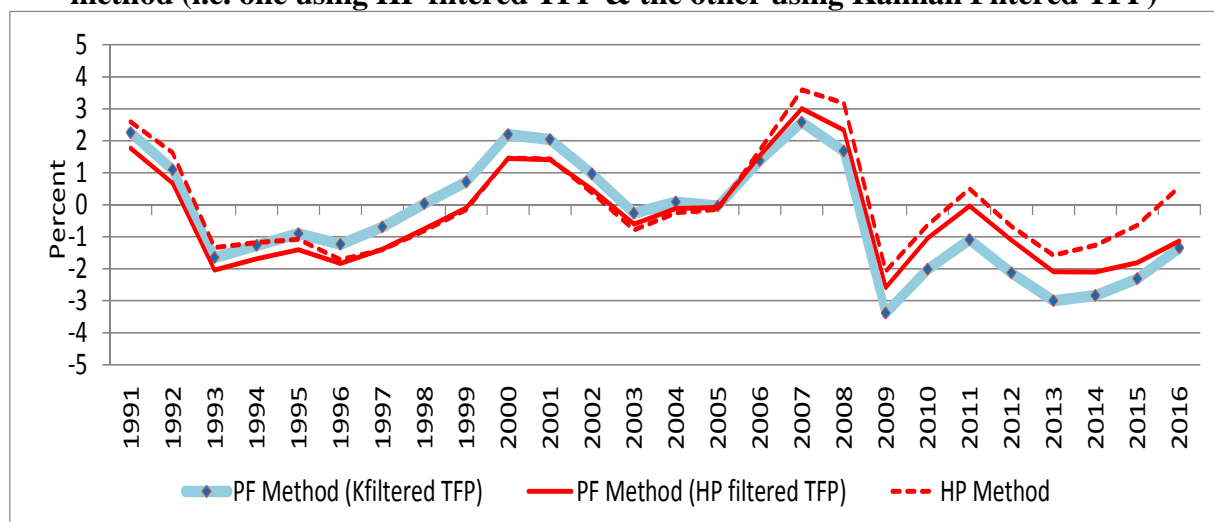
1.5 : Overall evaluation of the relative performance of the PF and HP methodologies over the period 2002-2014 : was the shift to the PF method justified or not ? This section has provided a significant amount of evidence in support of the EPC / ECOFIN Council decision to shift to the PF method in 2002. 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.

2. COMPARISON OF THE EU'S PRODUCTION FUNCTION METHODOLOGY WITH THE EQUIVALENT OECD AND IMF METHODOLOGIES

The previous section stressed the better performance, based on a number of criteria, of the EU's PF method over the 2002-2014 period, 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 superior, a more pertinent current issue for policy makers is how the PF method has been performing compared with the equivalent OECD and IMF methods (since Member State policy makers are much more interested in this comparison rather than the one with the HP filter). This section now addresses this question by assessing the respective performances of the EU, OECD and IMF methods over the period as a whole (2004-2014), as well as in the pre-crisis (2006-2008) and post-crisis (2009-2014) sub-periods. In addition, in order to provide a longer-term perspective, the section goes on to examine the results from an equivalent comparative revisions exercise published by the German *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 and 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 same approach to the EU output gap estimates for the equivalent 1980-2010 period, with the objective of comparing the relative degree of uncertainty surrounding the estimates from the EU, OECD and IMF methods respectively.

2.1 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 5 summarises the forecast-to-forecast revisions for the three institutions for 2004-2014. Unlike the PF/HP comparison, where stability differences were relatively small, Graph 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 is also a feature of the sub-periods shown, with the outperformance particularly striking for the 2009-2014 sub-period, where EU revisions are roughly half those of the IMF and a quarter of those of the OECD. This could have non-negligible implications for policy-relevant fiscal indicators such as the change in the structural fiscal balance. Whilst

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 5 : Short Term (Forecast-to-Forecast) Stability of Output Gap Estimates : EU-PF versus IMF & OECD Methods (Euro Area : 2004-2014) (Period Average Revisions to Estimates)
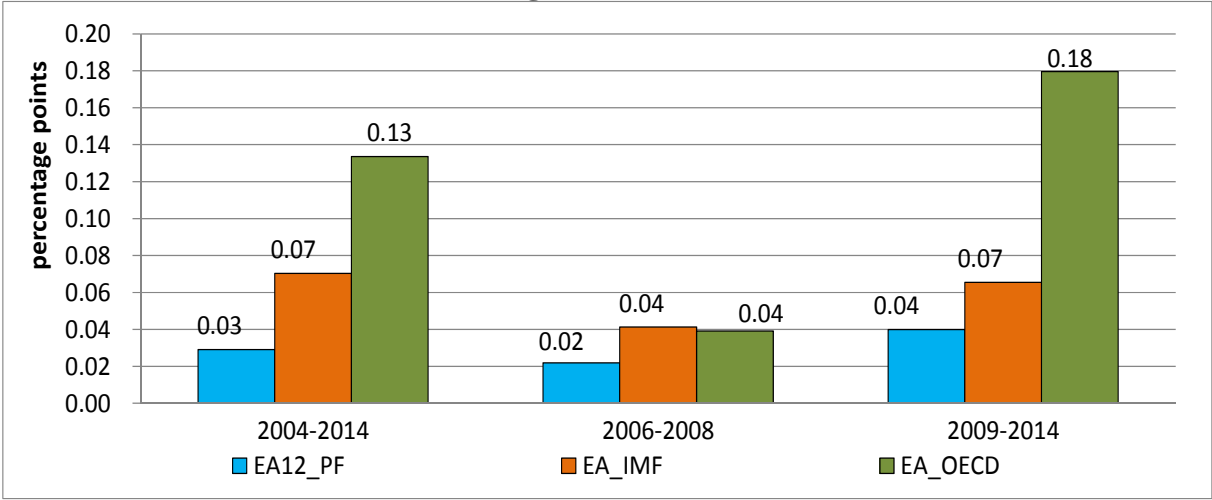
2.2 : Long-term real-time reliability of EU-PF vs OECD and IMF estimates : Graph 6 compares the real-time and *ex post* (i.e. autumn 2014) output gap estimates for the euro area as a whole for the years 2004-2014 (autumn vintages), as produced by the EU, IMF and OECD methods. It shows the average absolute revisions for the three sets of estimates for the period as a whole (2004-2014) and for the pre-crisis (2006-2008) and post-crisis (2009-2014) sub-periods respectively. The autumn 2014 vintage is used as the *ex post* reference in all three cases. The size 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 6 are as follows:
- Firstly, regarding the **period as a whole**, the real-time reliability/accuracy of the output gap estimates produced by the EU's method was significantly better than in the case of the IMF and OECD methods. The average absolute revisions for the EU's method were less than half those of the OECD method (0.9 vs 2.0) and 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, and especially relative to the OECD;

- Finally, for the post-crisis period **2009-2014**, whilst the real-time reliability performance of the EU and IMF methods converged substantially, the OECD's performance remained very much an outlier, with revisions roughly three times greater than those of the EU & IMF methods. Regarding the EU and IMF methods, it is quite striking how similar the real-time and *ex post* output gap estimates have been for the Euro Area in each year of the post-crisis period (see annex 2).

Graph 6 : Long Term, Real Time, Reliability of Output Gap Estimates : EU-PF versus IMF & OECD Methods (Euro Area : 2004-2014)
(Period Average Absolute Revisions to Estimates) (Real Time versus Ex Post Estimates)
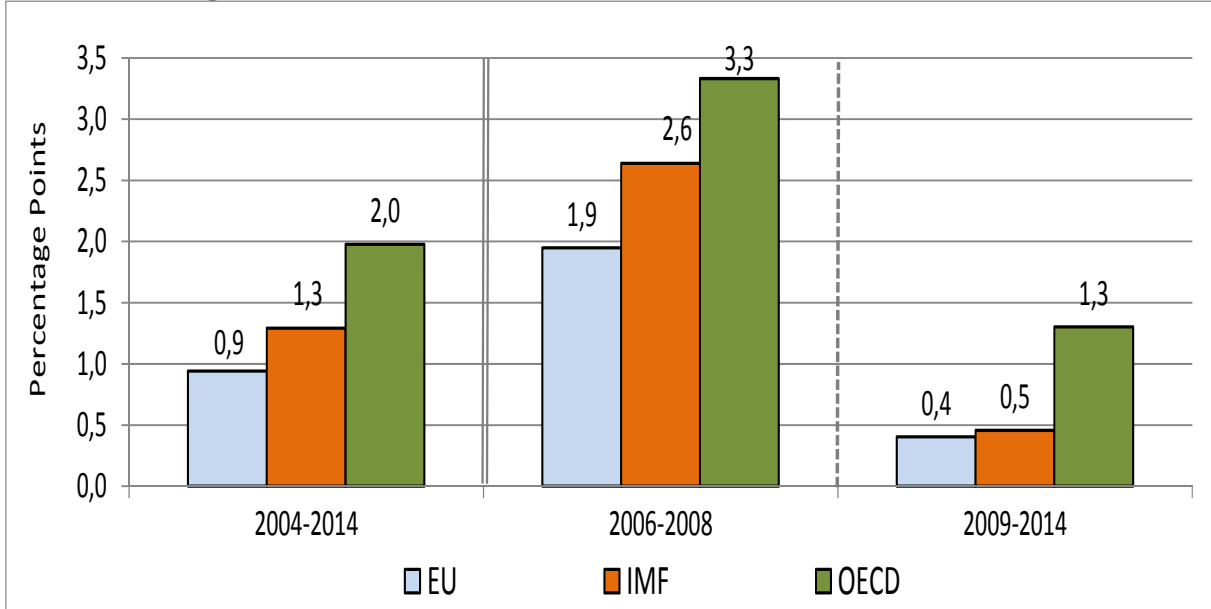
2.3 Performance of EU vs IMF and OECD methods around the turning point of the financial crisis : Regarding the relative performances of the different institutions around the time of the crisis, Graph 7 shows the real-time (spring 2009) and *ex post* (autumn 2014) output gap estimates for 2009 from the EU (for both the PF and HP filter), IMF and OECD methods, with the difference between the real time & ex post estimates indicative of the relative reliability of the four methods.

Graph 7 : 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
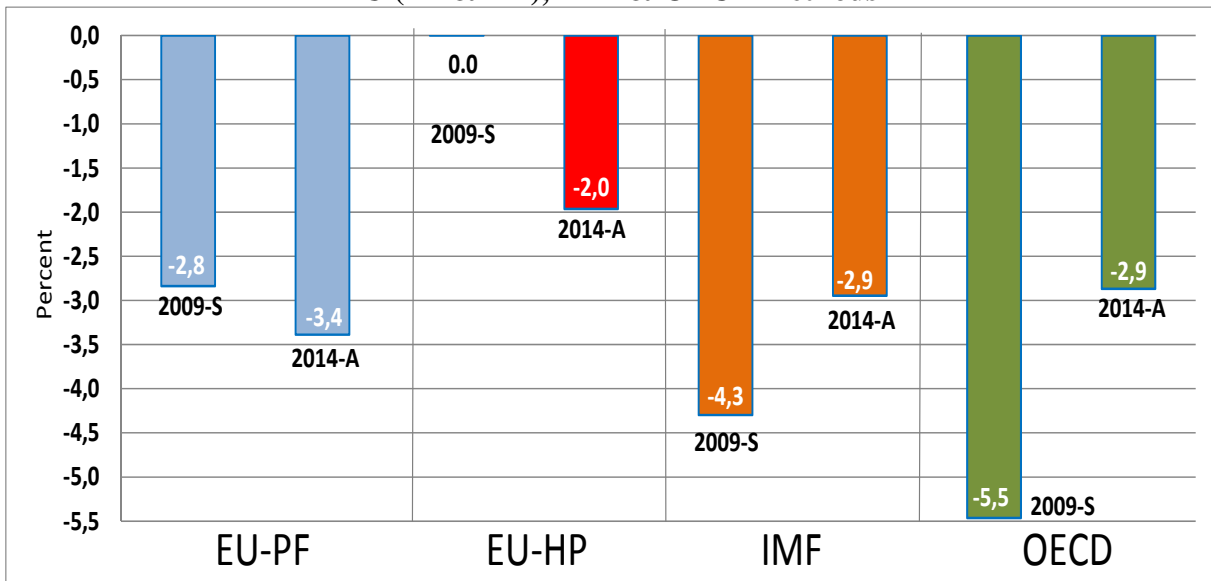
The graph shows that, in spring 2009, there was a very wide range of forecasts for the Euro Area's output gap in the year 2009, ranging from a zero forecast from the HP filter (implying that most of the effects of the crisis were structural) to -4.3 % and -5.5 % from the IMF and the OECD respectively (implying the opposite conclusion, namely that most of the effects were cyclical), with the EU PF method in the middle of these estimates (-2.8 % - implying that the crisis would have both cyclical & structural implications). Five and a half years later (in autumn 2014), the revisions for the year 2009 suggest that not only did the PF method do significantly better than the HP filter in terms of revisions (i.e. a revision of only ½ a % point, indicating strong real time reliability around this crucial turning point), 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 and OECD methods (-2.9 %) are almost identical to the EU's initial (i.e. real-time) Spring 2009 estimate (-2.8 %).

2.4 : Assessing the 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 (see Annex 7 for details). 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, 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 time 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 section of table 1, the average output gap spread for the two time periods is reported, with the second section giving the number of times that an output gap estimate has changed its sign, at least once, for a certain year in each period[9]. Table 1 shows that if one applies the Bundesbank's approach to the EU's methodology, one would draw the following two main conclusions :

---

[9] See Annex 4 to examine the detailed results from replicating the Bundesbank's analysis for France over the period 1998-2010.

- 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 – see annex 4 for details

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 / revisions will always be a feature of output gap calculations[10], the paper has shown that, in relative terms, the PF calculation method has outperformed the HP filter, OECD & IMF methodologies over the period 2004-2014, on the basis of a significant number of objective criteria. In addition, the performance of the EU's method has steadily improved over time, with extremely low absolute revisions in the post-crisis years.

Not only was the EU's methodology better than the OECD & IMF methods in terms of relative stability & 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 (-2.9%) are almost identical to the real time Spring 2009 estimate produced by the EU's methodology (-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.

---

[10] Although the degree of uncertainty surrounding output gap estimates has been reduced, it is also clear that uncertainty has not been, and never will be, completely eliminated. There will never be a method which will remove the need for all revisions, with forecast & data uncertainties remaining a fact of life for policy makers to grapple with. Distinguishing cyclical from structural factors in real time will continue to be prone to error & an element of judgement will always be needed in assessing output gap developments in the context of policy surveillance exercises, most notably in the fiscal arena. Consequently, the Commission accepts that any fiscal surveillance framework cannot exclusively rest on a mechanical interpretation of one single structural indicator of fiscal effort. In fact, the Commission uses a three-pillar approach, namely : a) *the change in the structural balance* : where the emphasis is on giving fiscal policy makers the best, real time, estimate of the cyclical position of the economy; **b)** *the expenditure benchmark* : where the emphasis is on using a much more stable 10 year potential growth rate as a benchmark against which to judge whether the discretionary fiscal actions of Member States were affordable or not based on the medium to long run output potential of their respective economies; & finally, c) *non-structural, "bottom-up" measures of fiscal effort*.

Whilst the relative performance of the EU's methodology suggests that it is well-designed, 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.

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 all mainstream output gap estimation methods, & why policy makers are right to continuously warn that so called "good economic 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 economic 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[11]. In this context, referring back to the point raised earlier by BIS economists (Borio et al. 2014), with their suggested solution being to augment the conventional output gap calculation method by using information about the financial cycle to identify investment booms (& in this way to produce "finance-neutral" output gaps), further work on this issue will be described in a forthcoming ECFIN Economic Paper[12].

---

[11] 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 focuses 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.

[12] Whilst the pre-crisis period showed up serious weaknesses with the PF method, nevertheless there are some hopeful signs that the extent of the pre-crisis mistakes can be reduced in the next upswing. Firstly, the post-crisis methodological improvements made to the method in 2010 (TFP) & 2014 (NAWRU) will help to reduce the risk of procyclical output gap estimates in the next upswing phase. Secondly, it is not inconceivable that there will be a research breakthrough on how the PF method could better handle investment in the boom phase of cycles. Thirdly, a number of non-methodological improvements can also be made, including impressing on desk officers the need to avoid overly optimistic actual GDP forecasts & asking national statistical institutes & Eurostat for better official data series (e.g. on hours worked; labour cost & capacity utilisation indicators; migration; longer sample lengths, especially for some of the "new" Member States etc). Finally, the production of finance neutral output gaps could serve as a useful complementary warning system for policy makers.

ANNEX 1
Short Term (Forecast-to-Forecast) Stability of Output Gap Estimates

The average revisions between forecast rounds are calculated as the average of these revisions over a period of 14 years, including one forecast year.

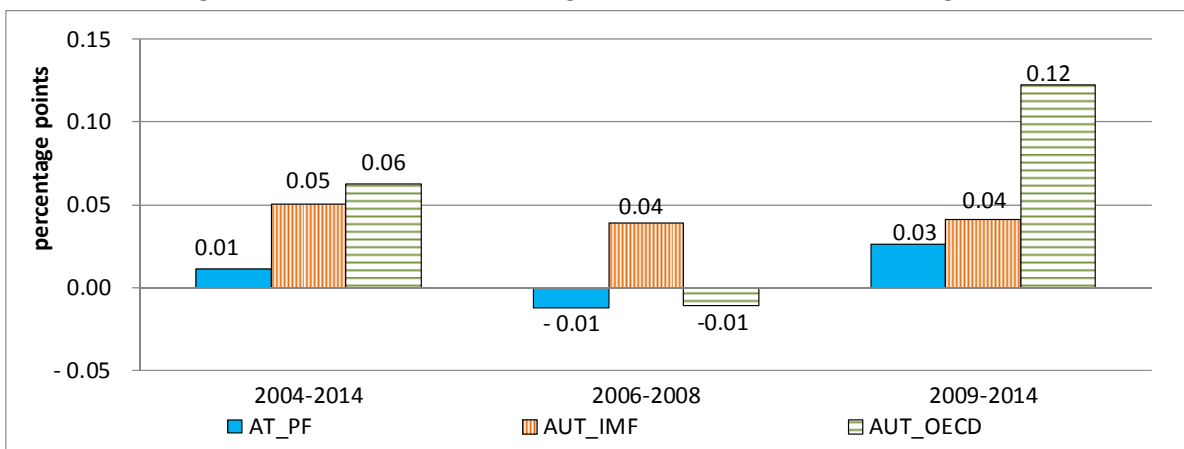(Graphical Analysis for all 28 EU Member States plus Euro Area aggregate)

BE

Short-Term Output Gap Revisions
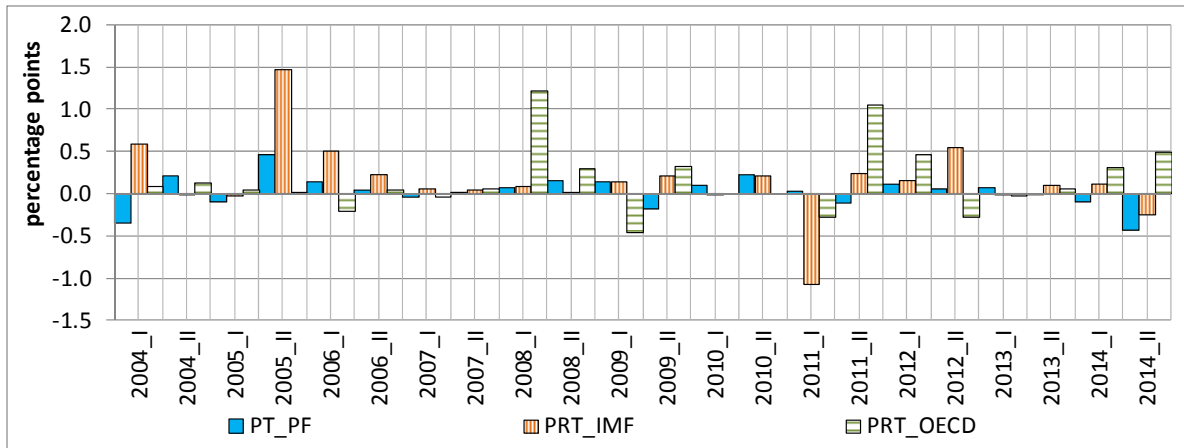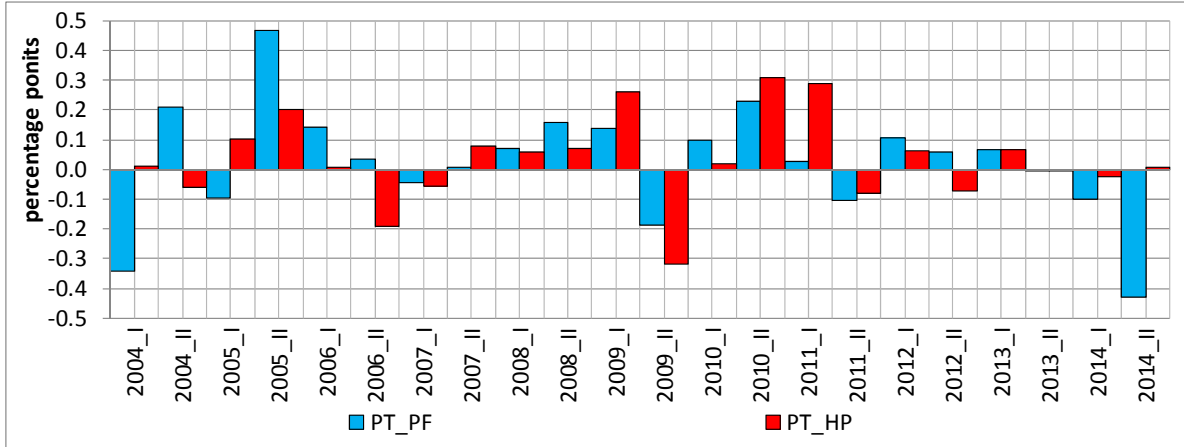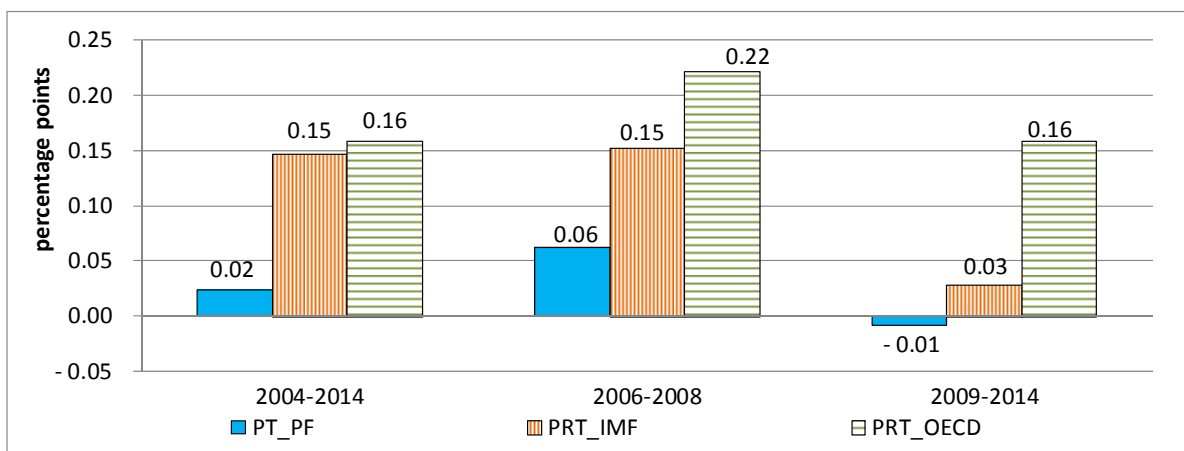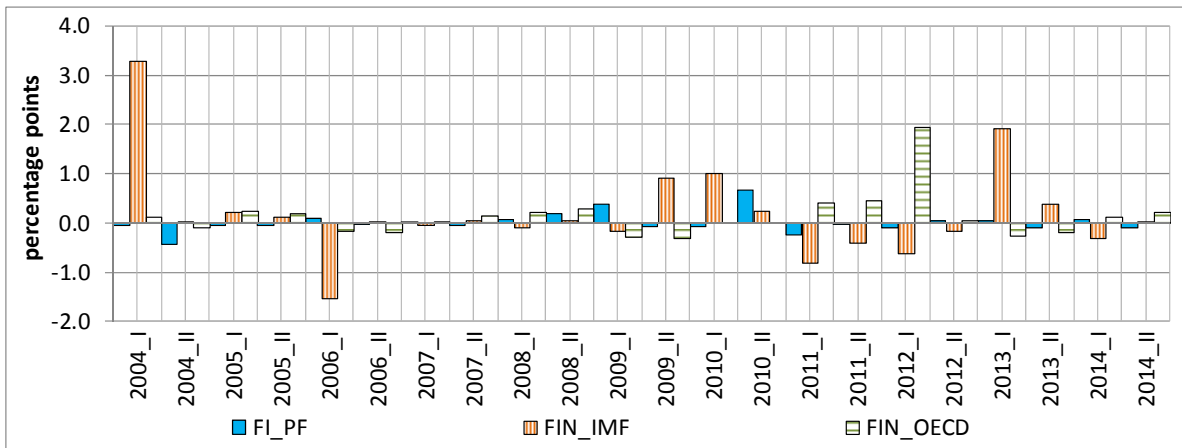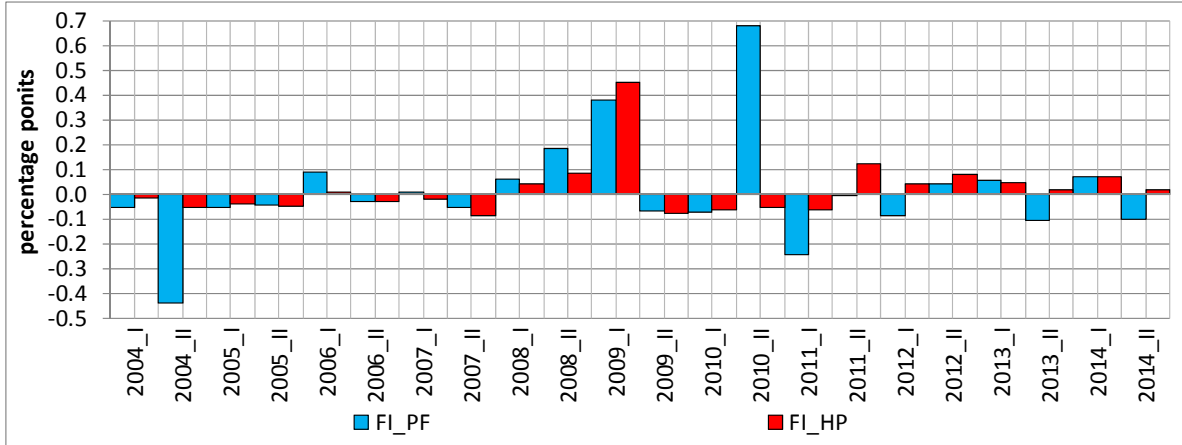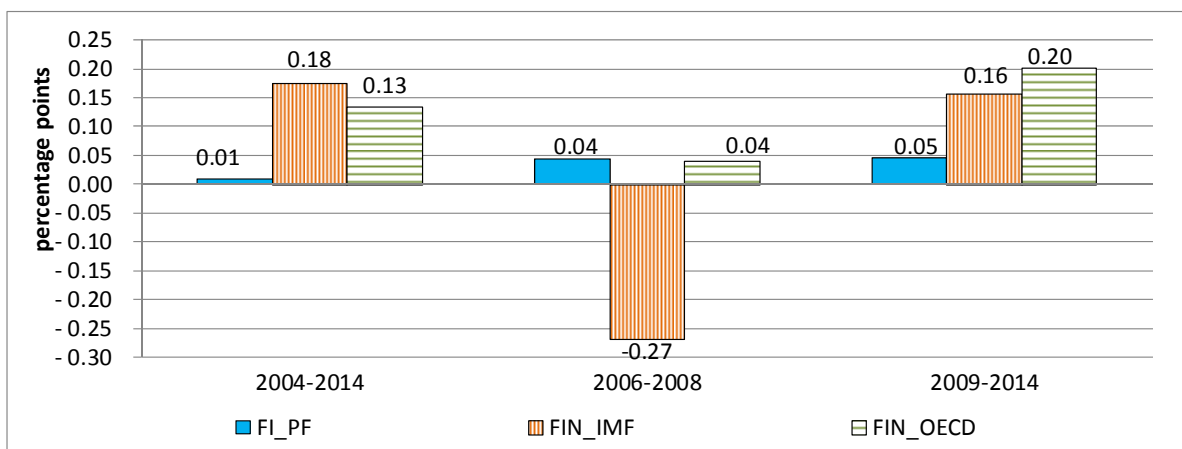### (each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
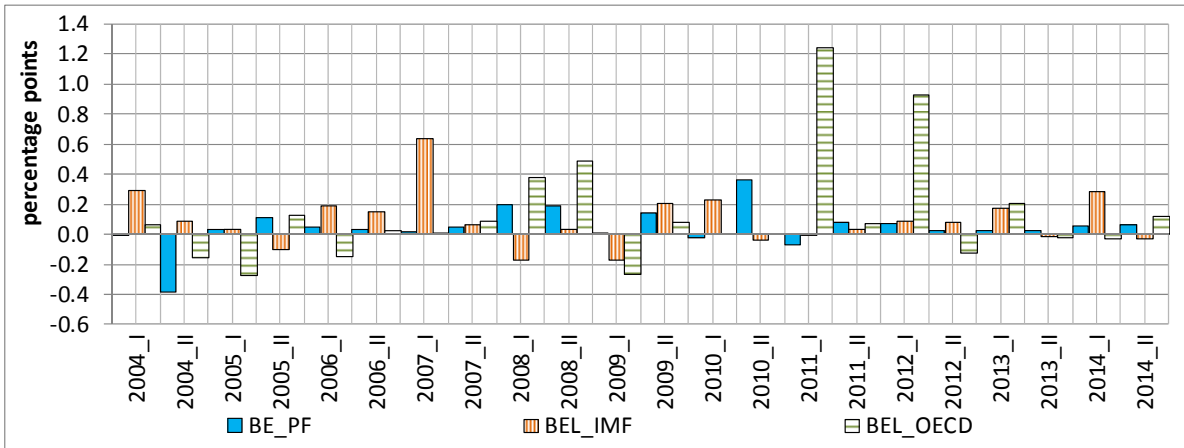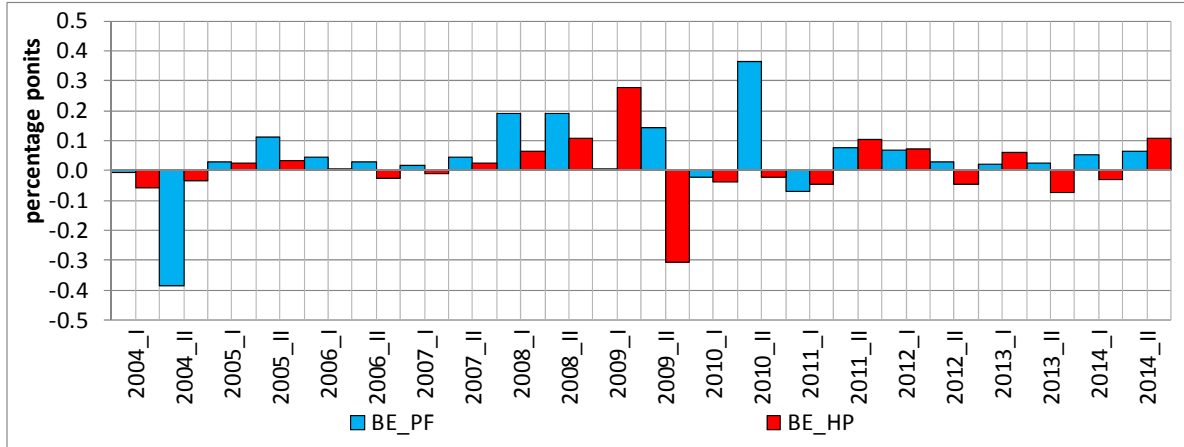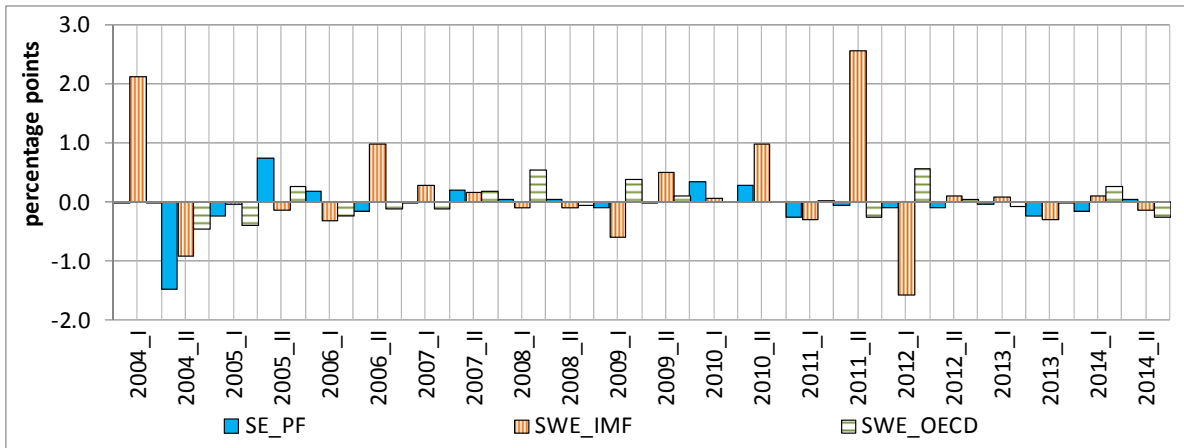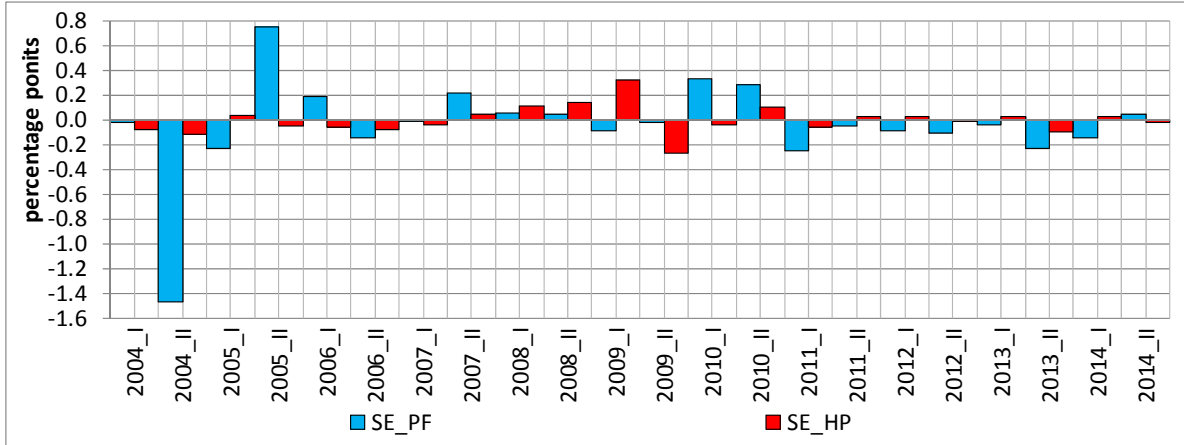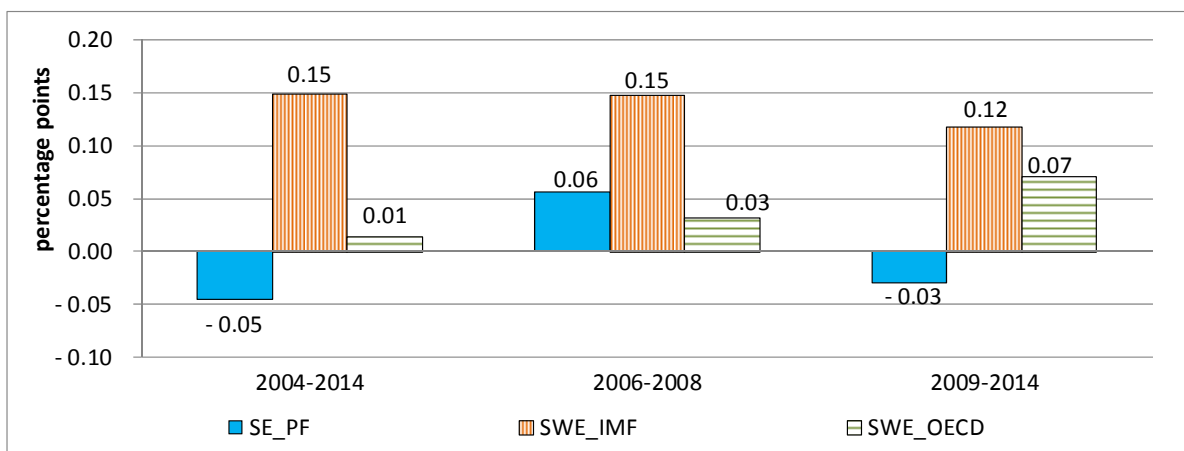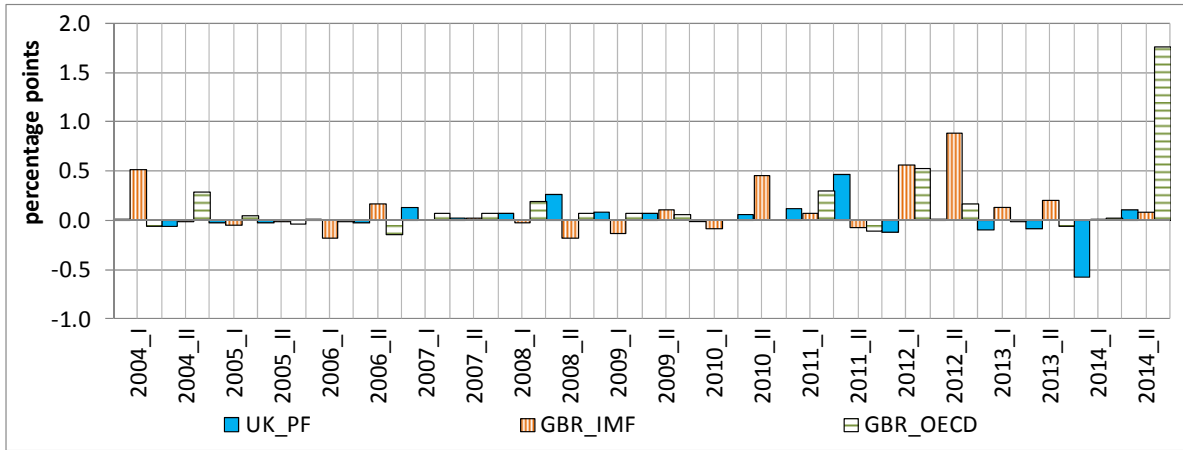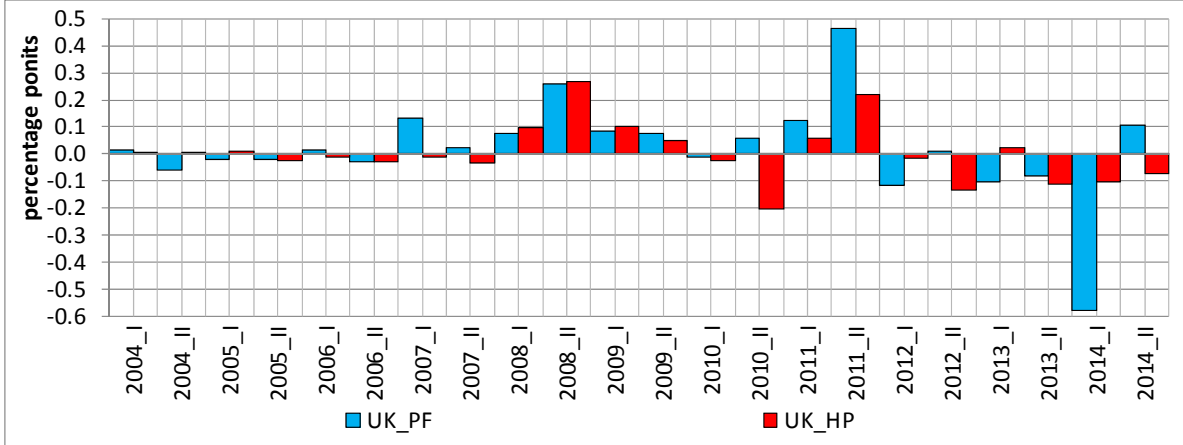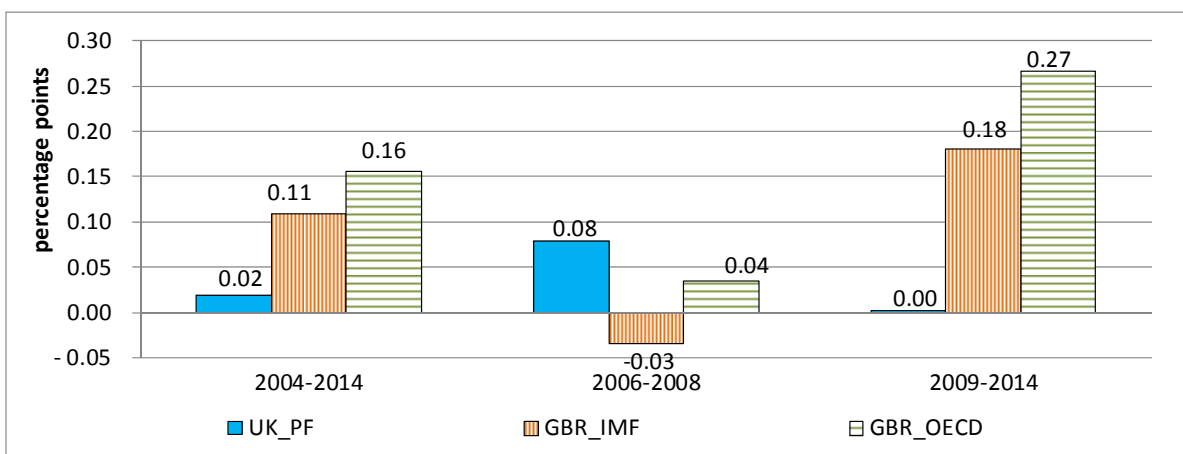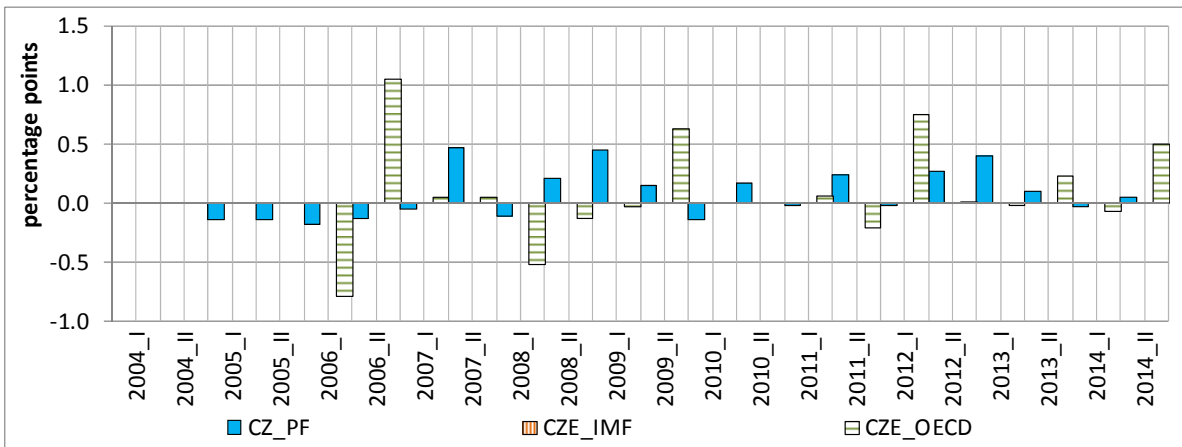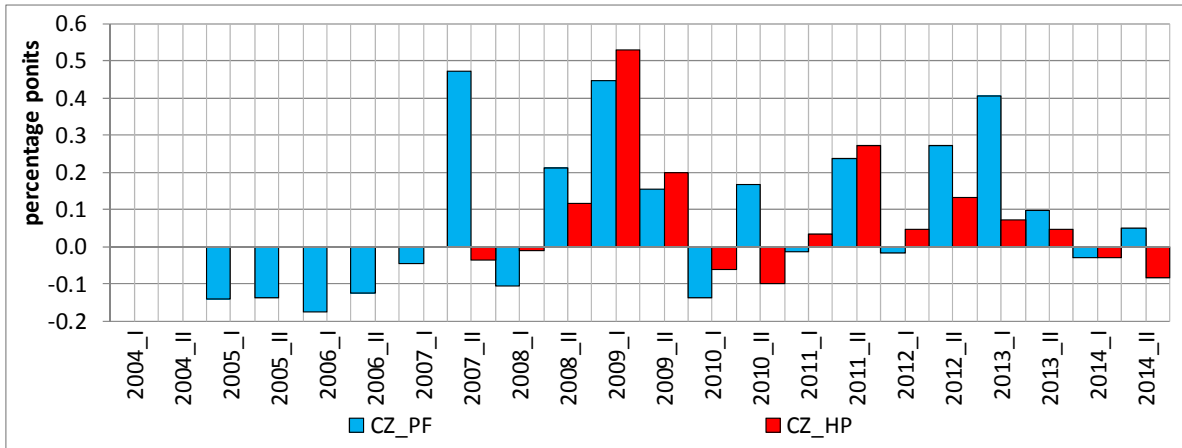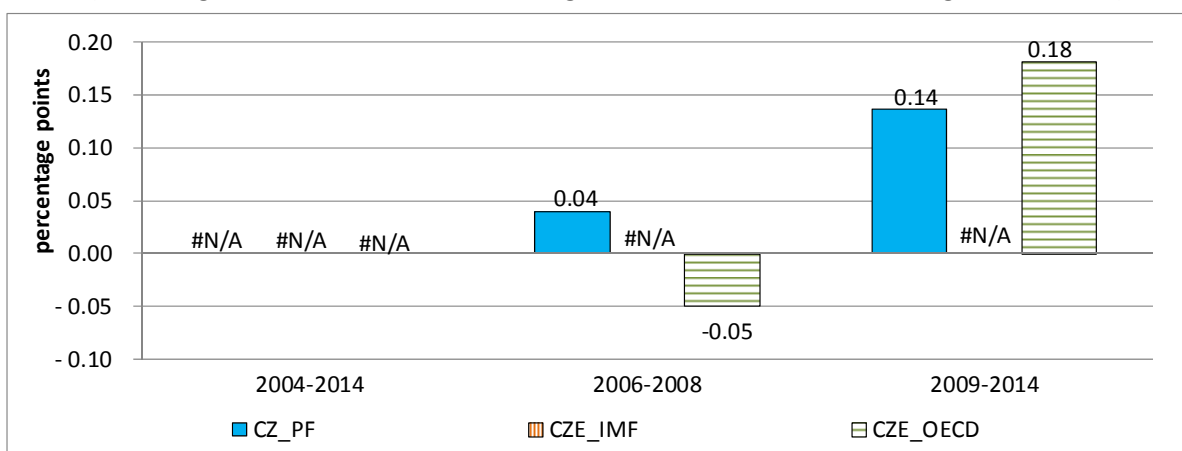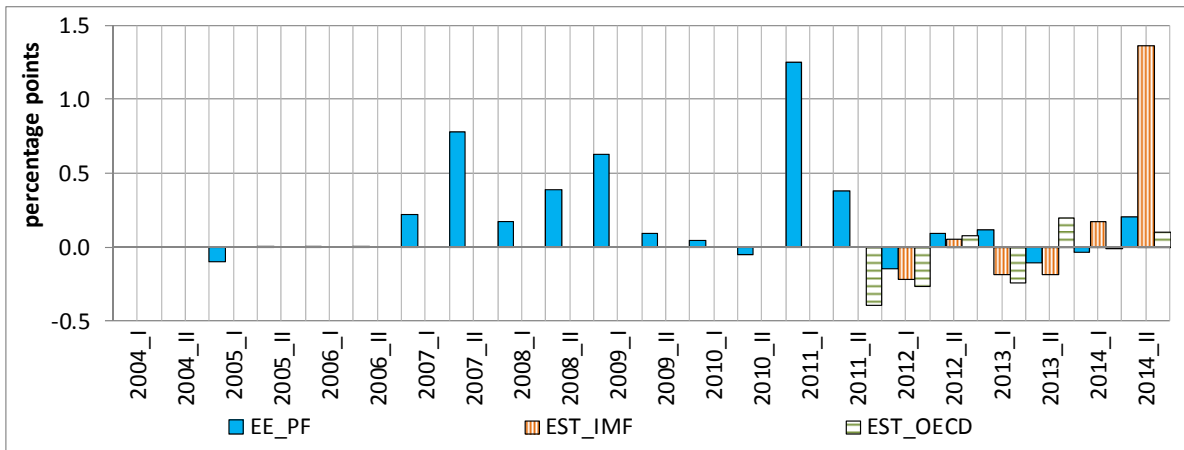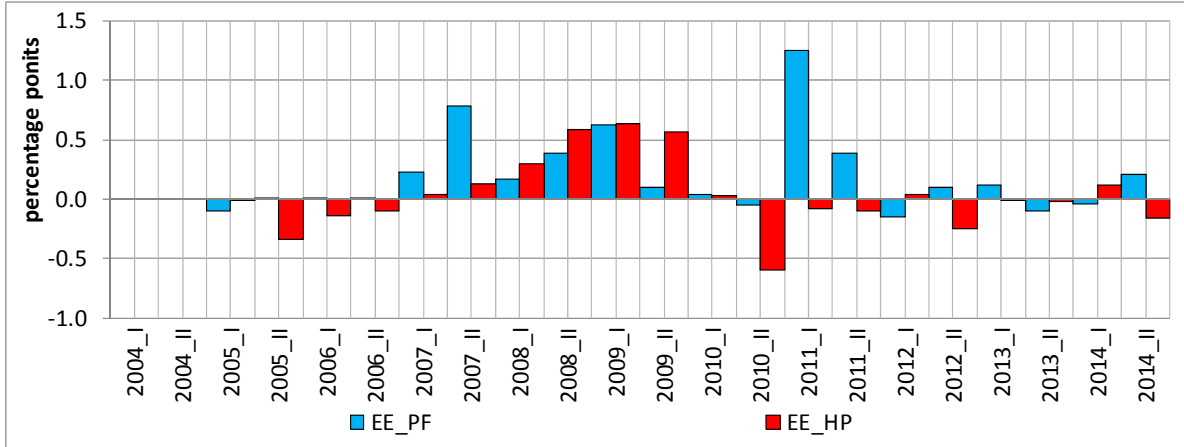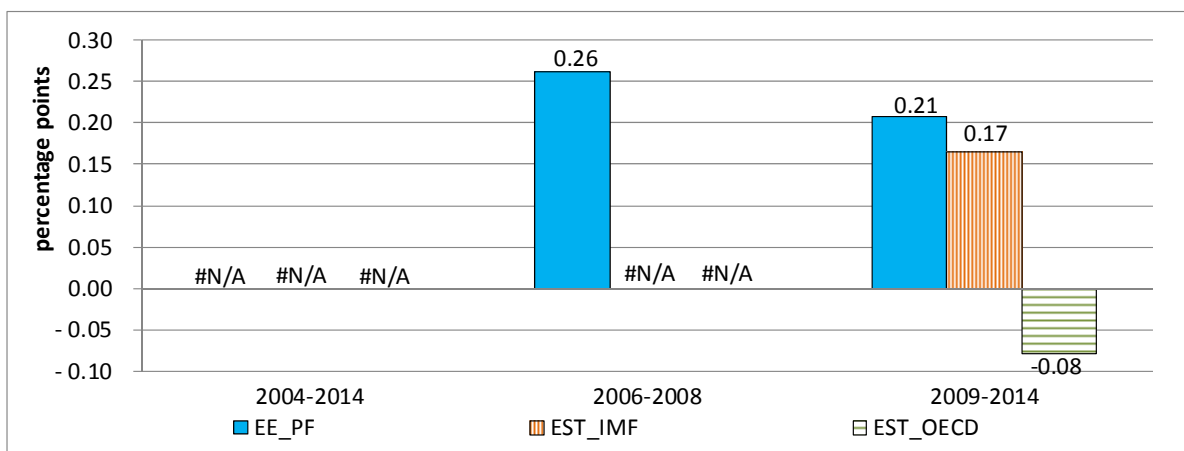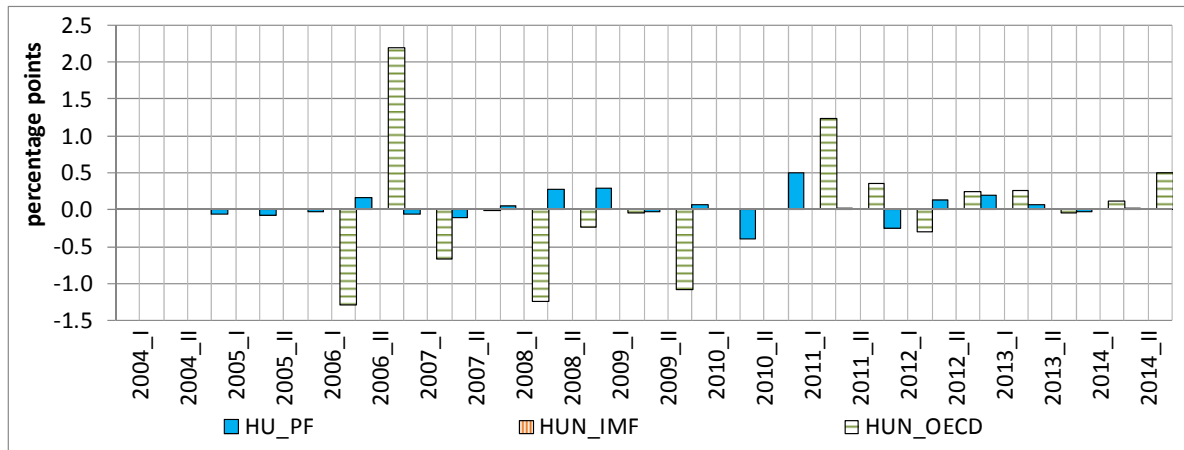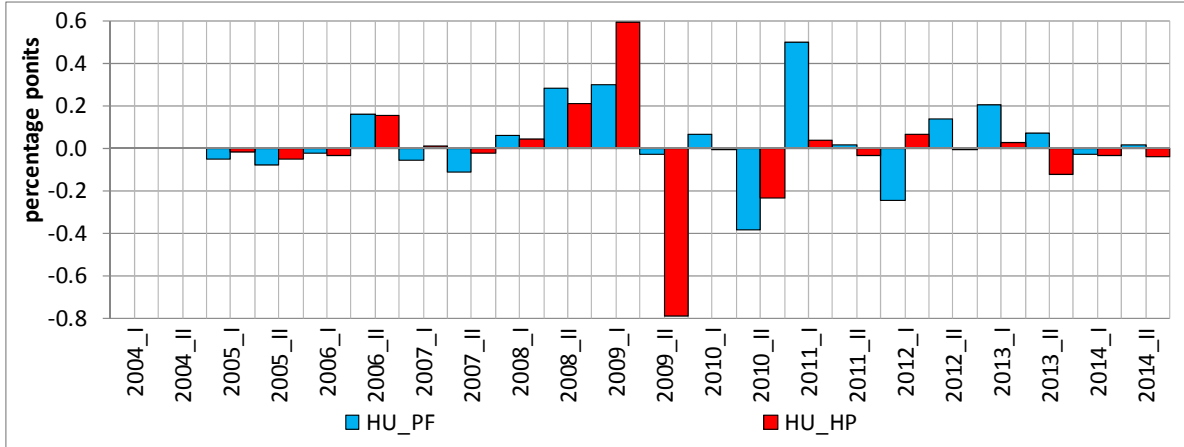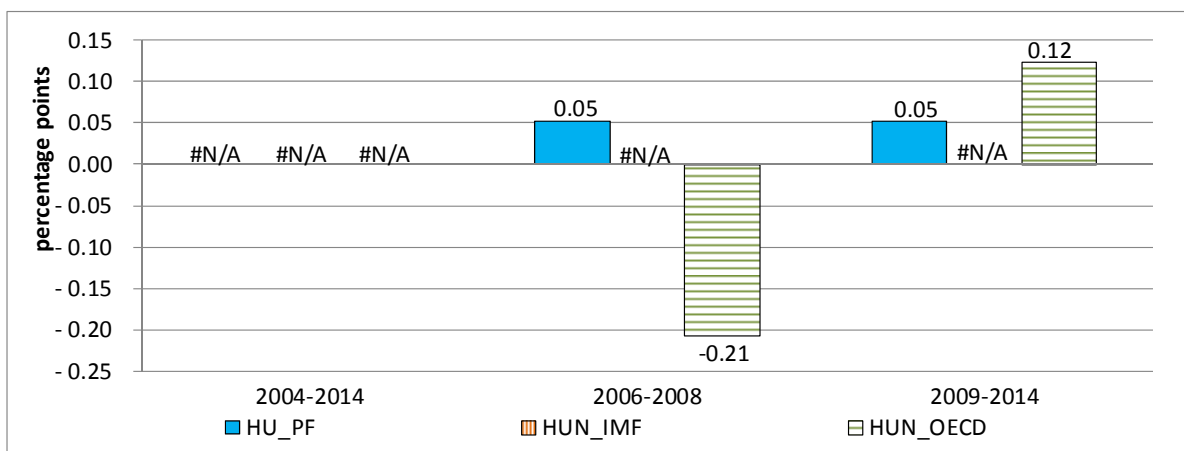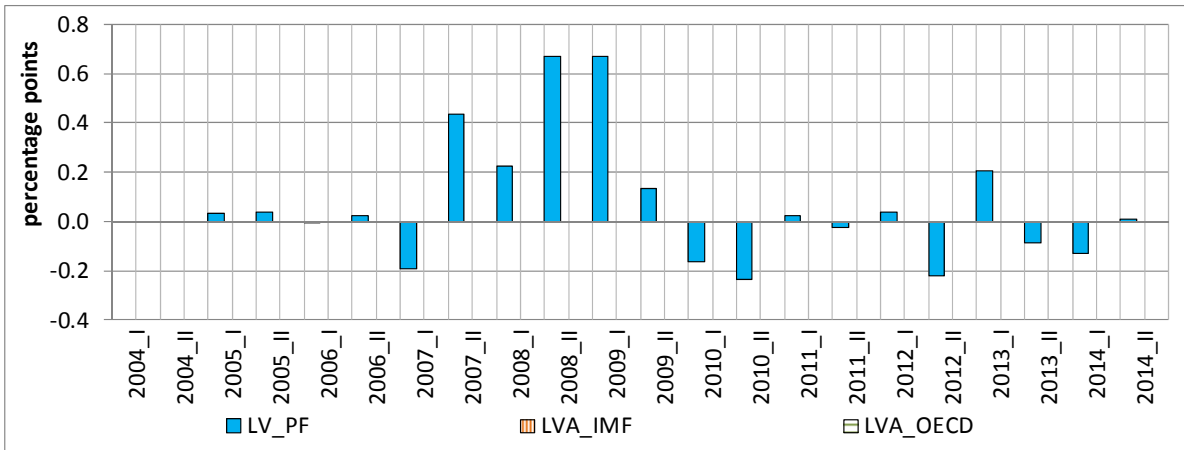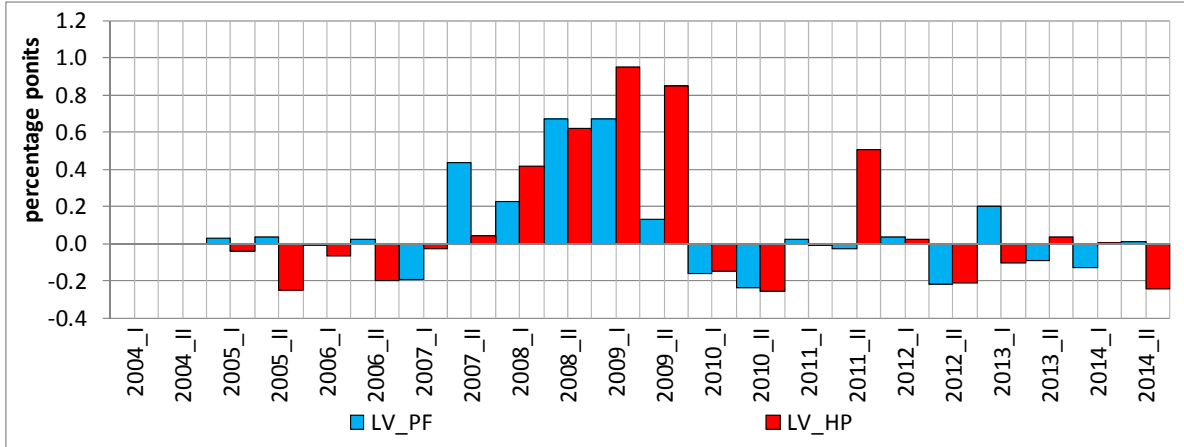

Output Gap Period Average Revisions
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
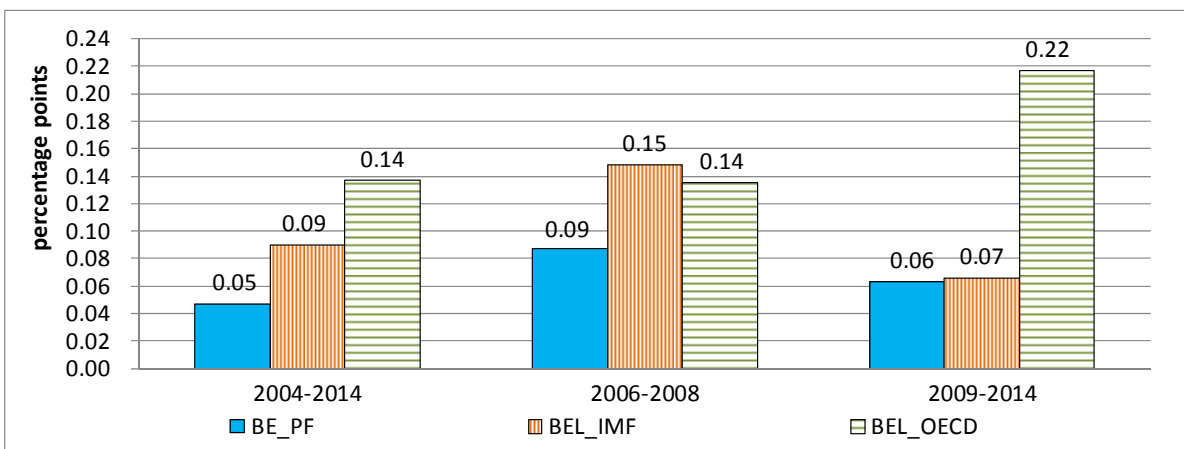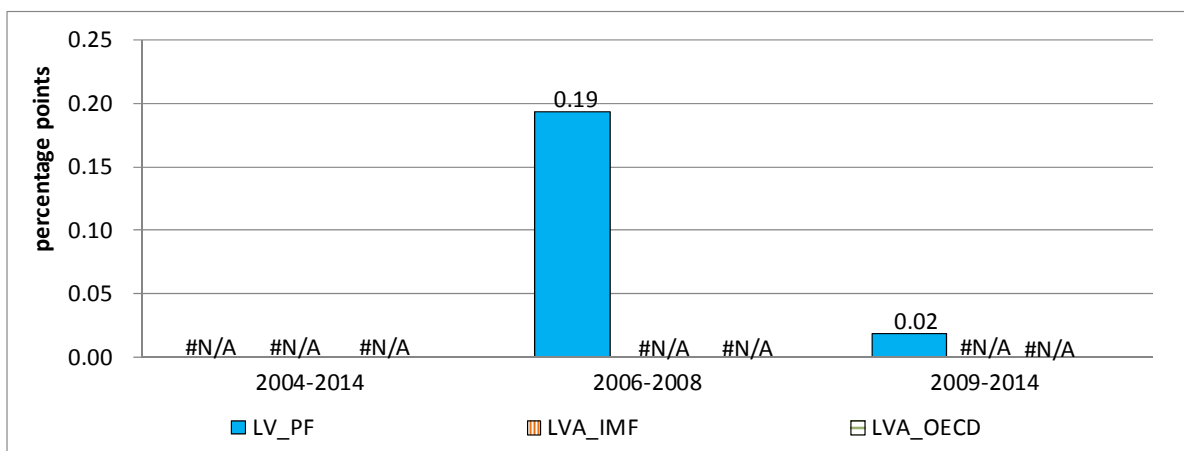
DE

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
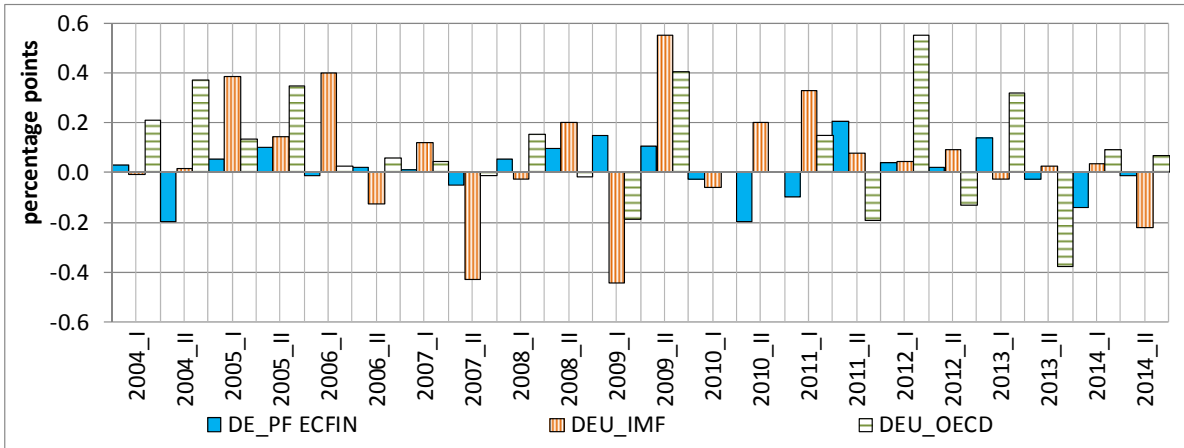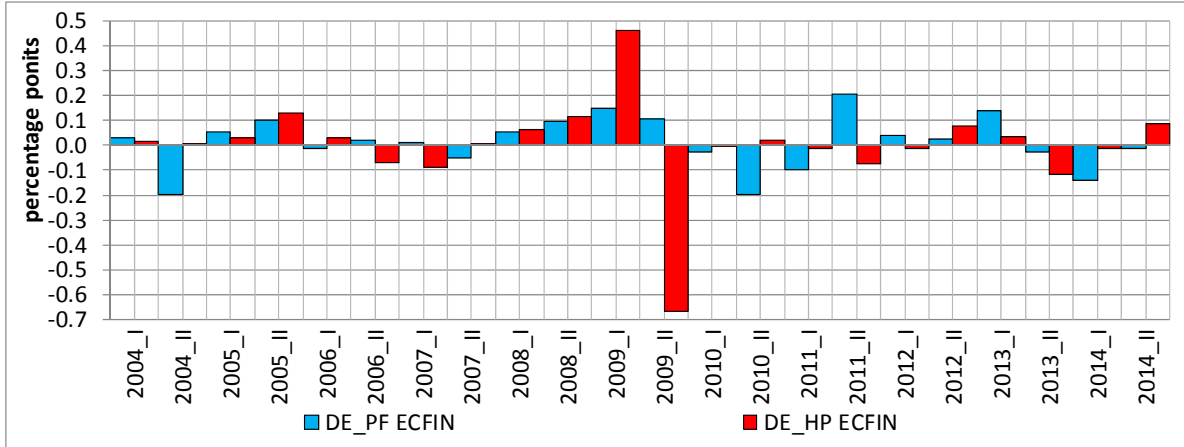
Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
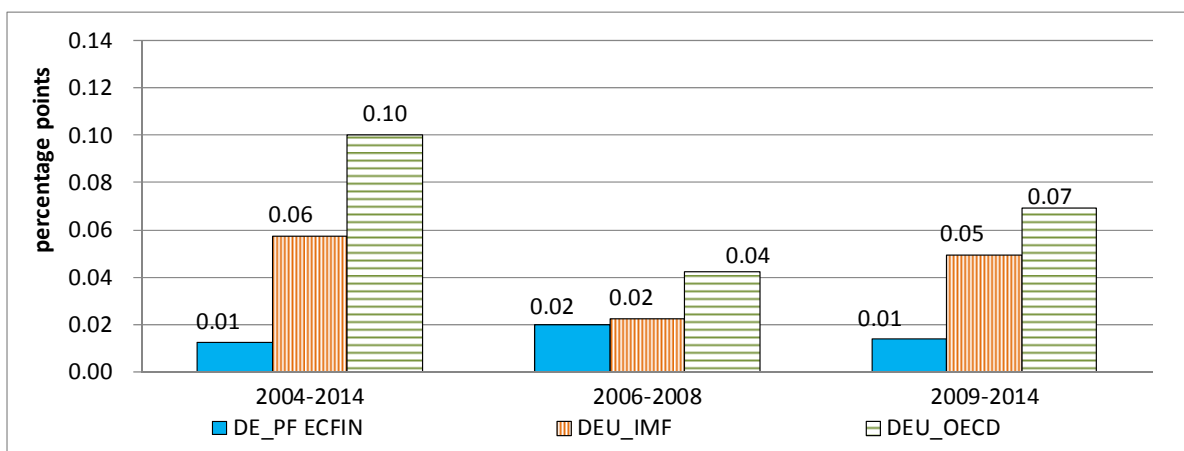
DK

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
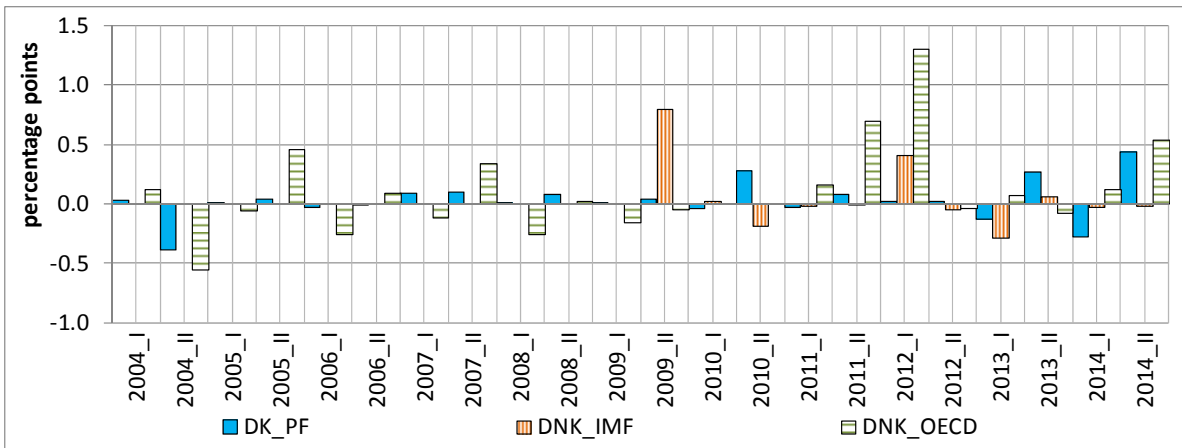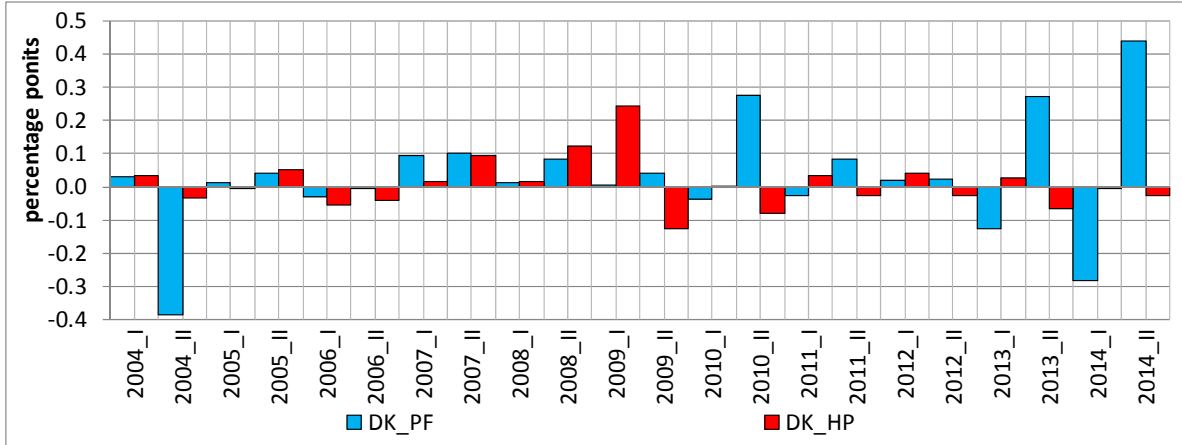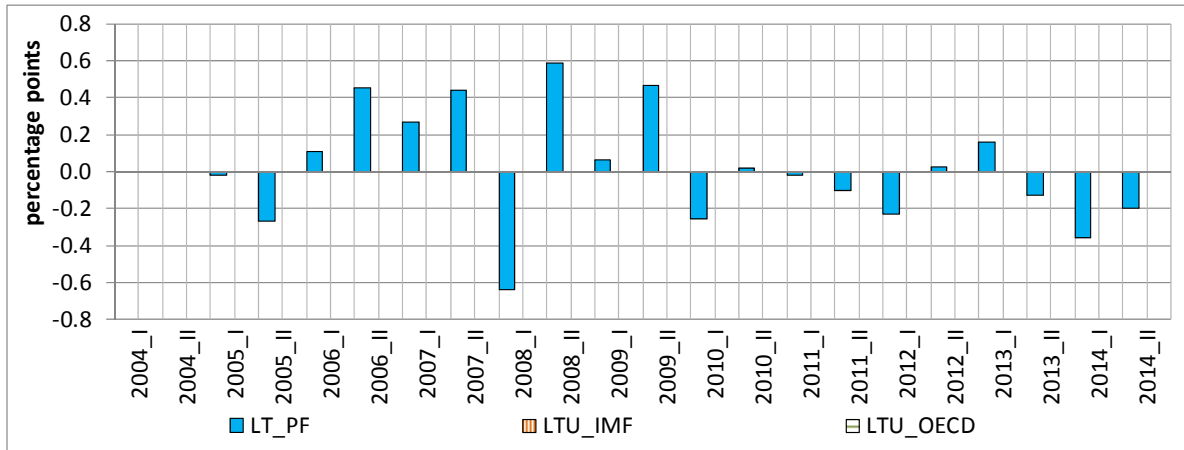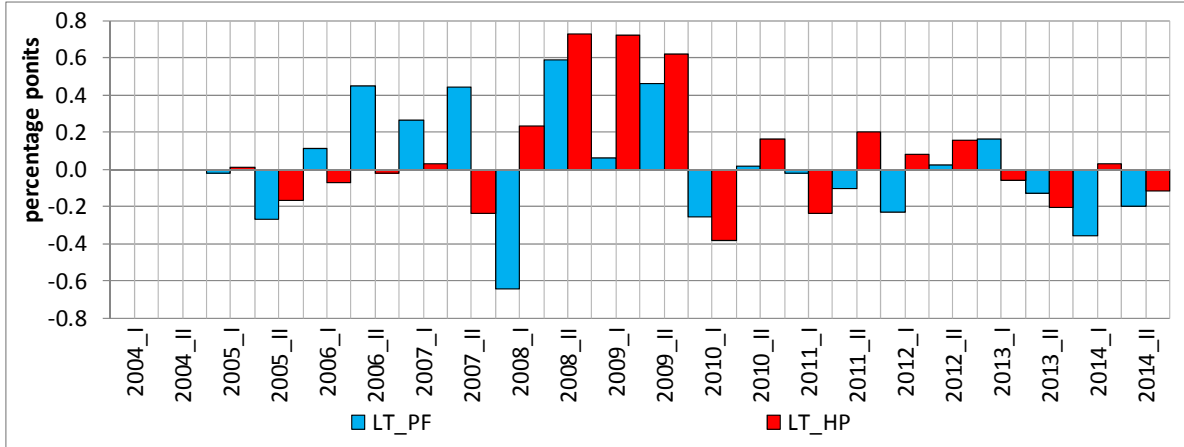

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
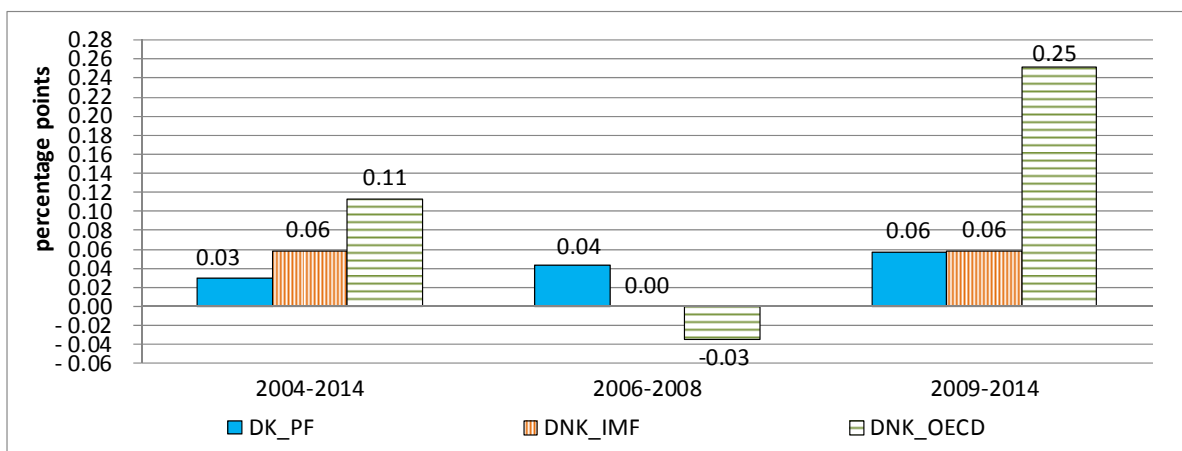
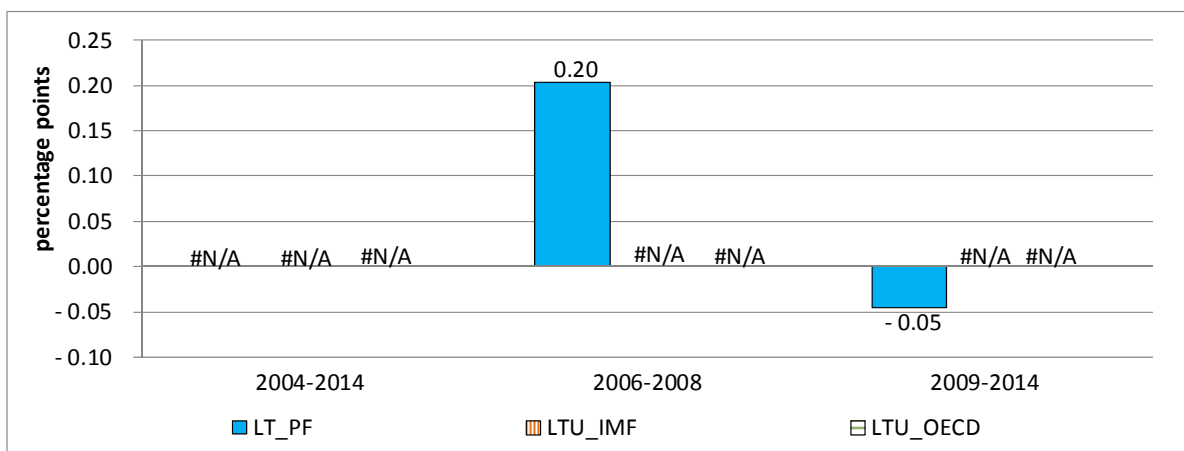note: DK vintages of the IMF start in 2009

EL

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
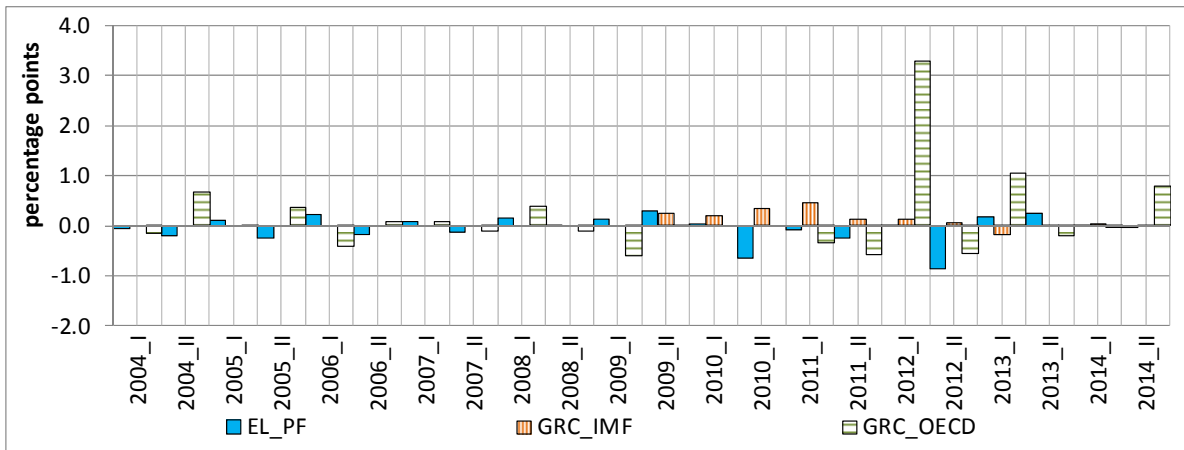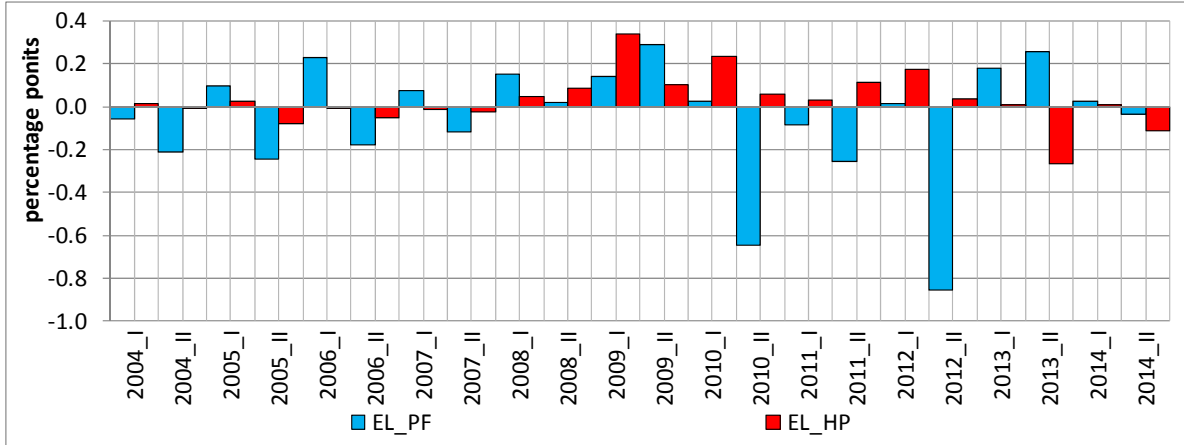
Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
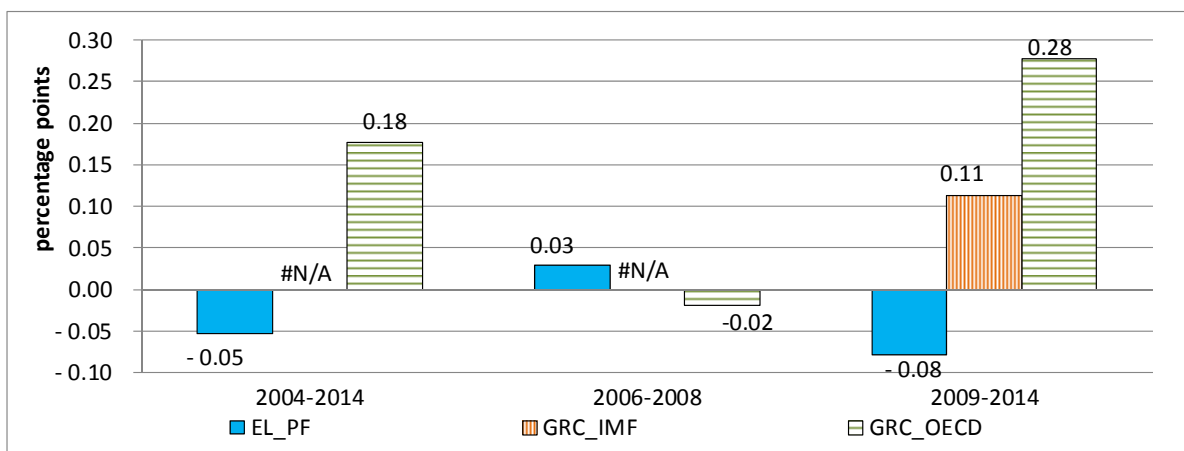
note: EL vintages of the IMF start in 2009

ES

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
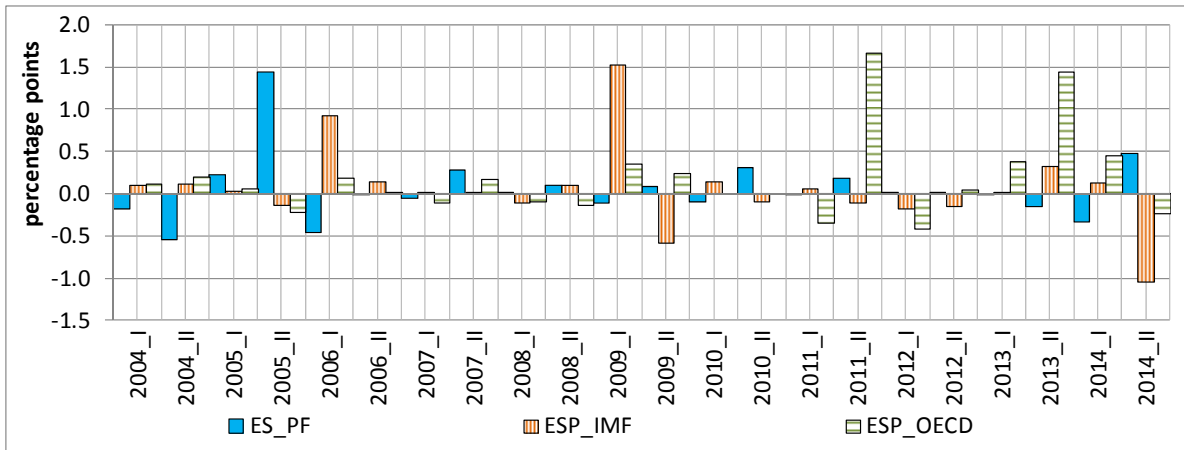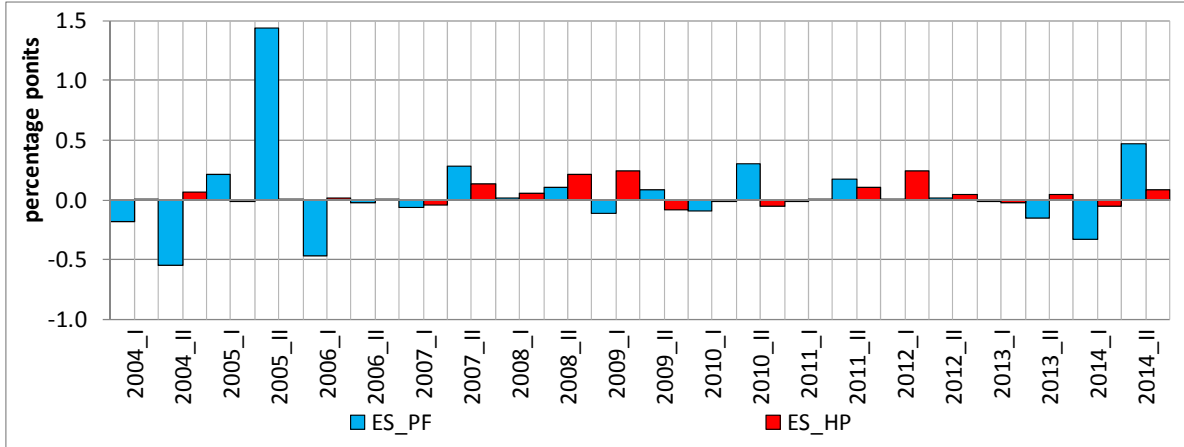
Output Gap Period Average Revisions,

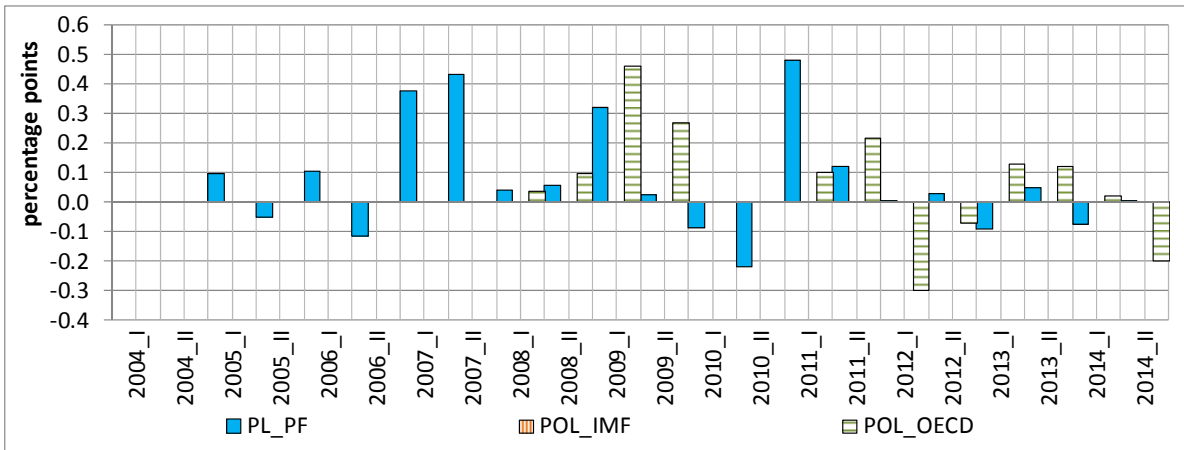(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
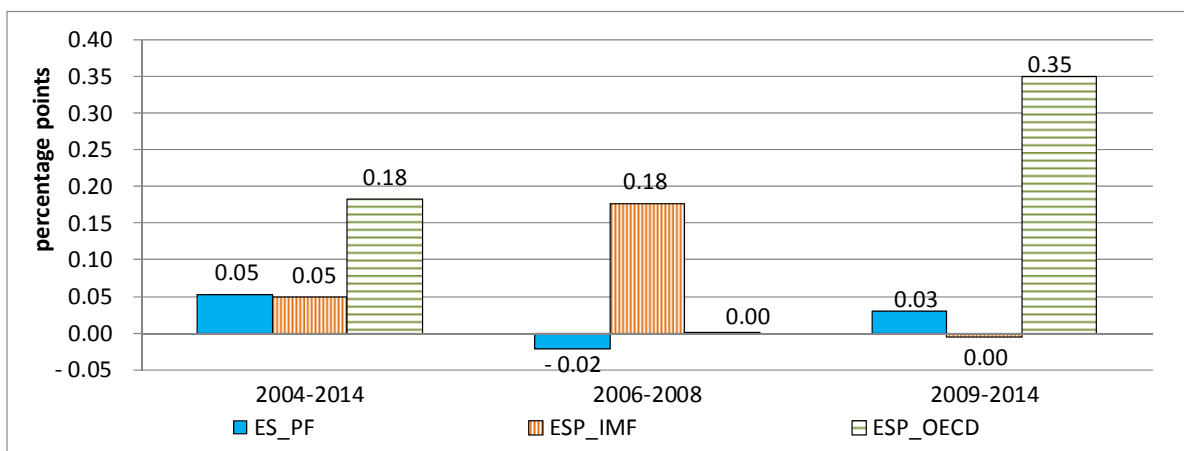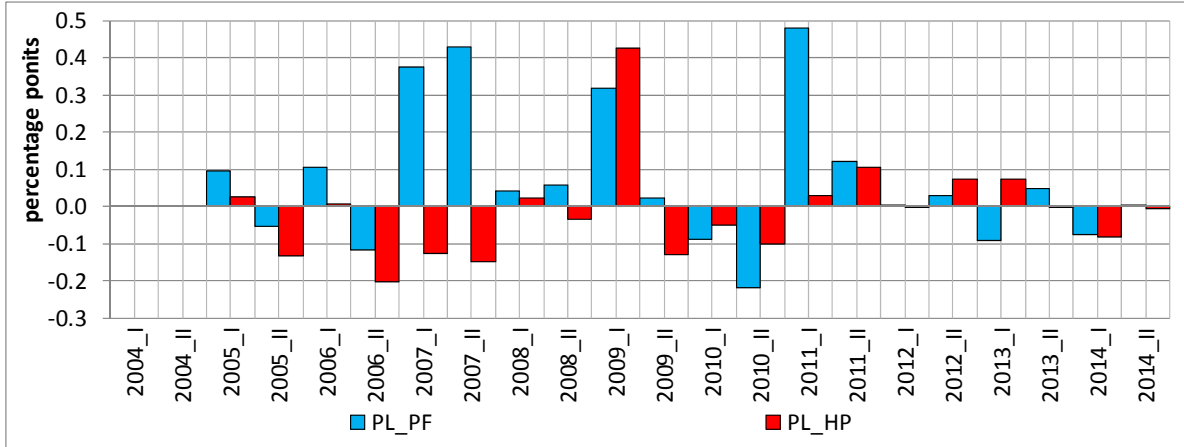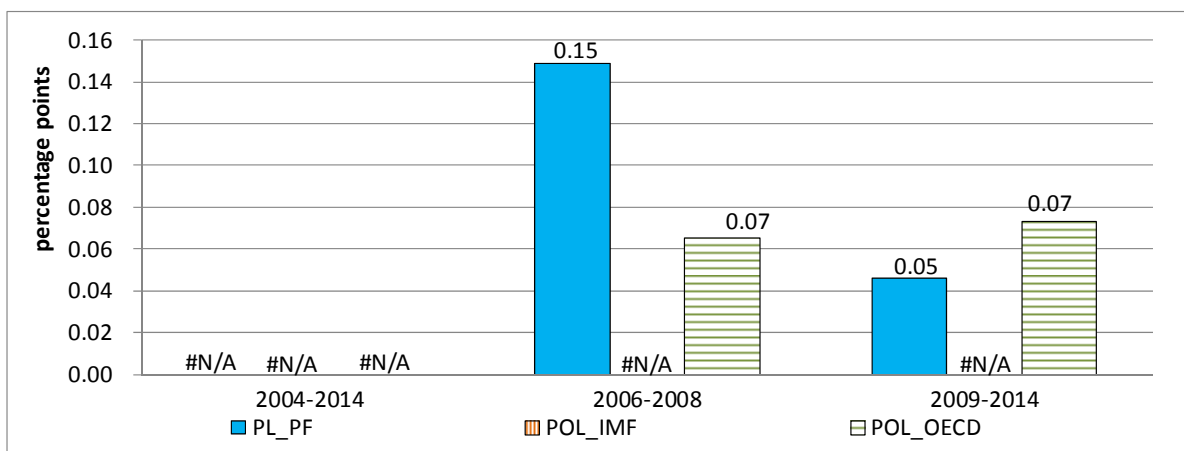
FR

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
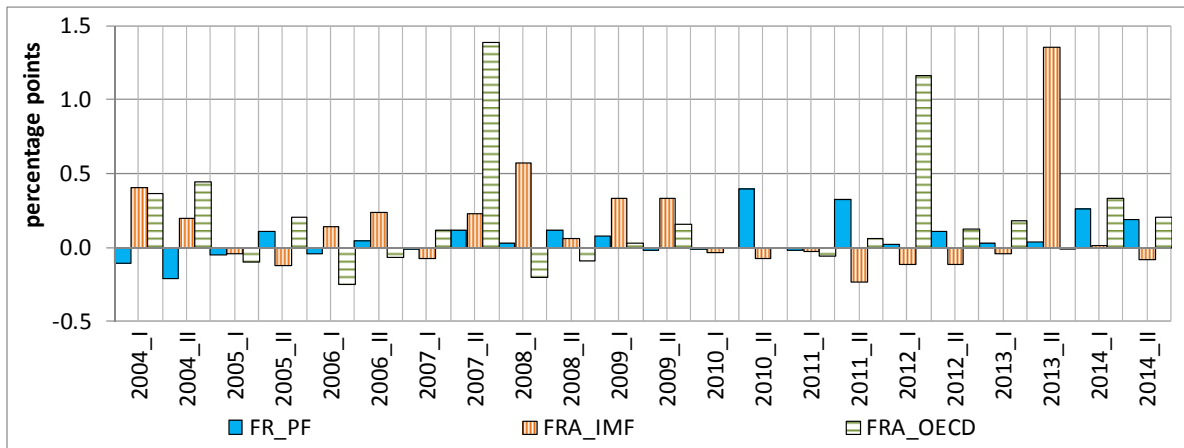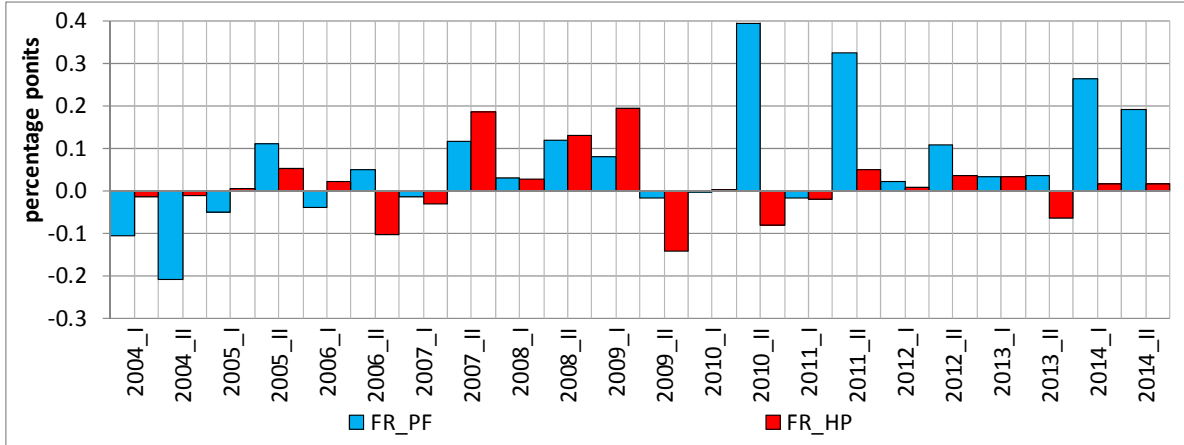
Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

IE

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
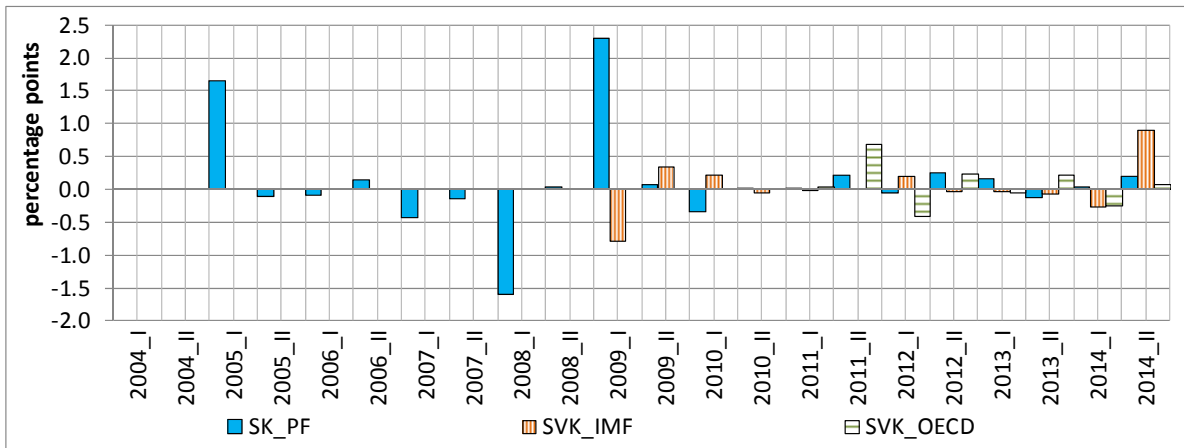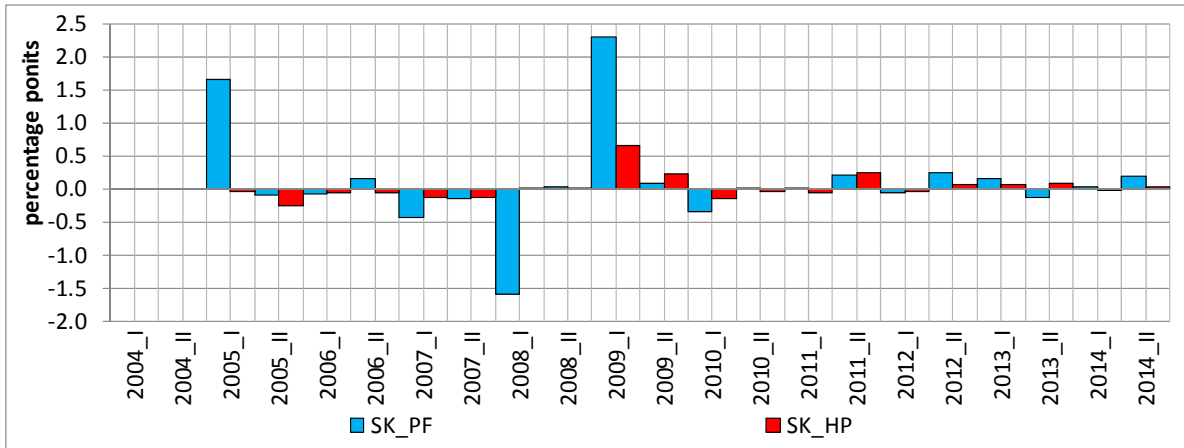

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
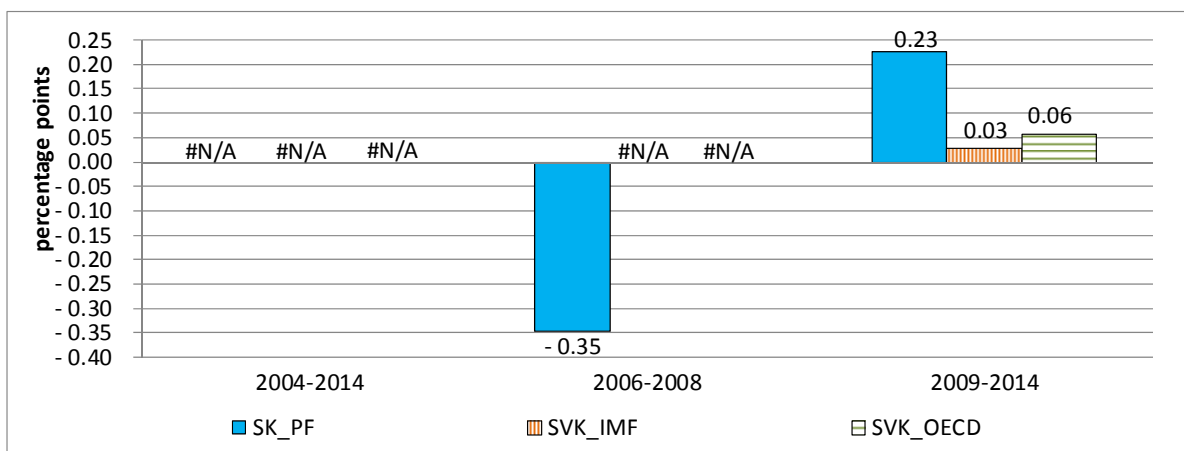
IT

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

LU
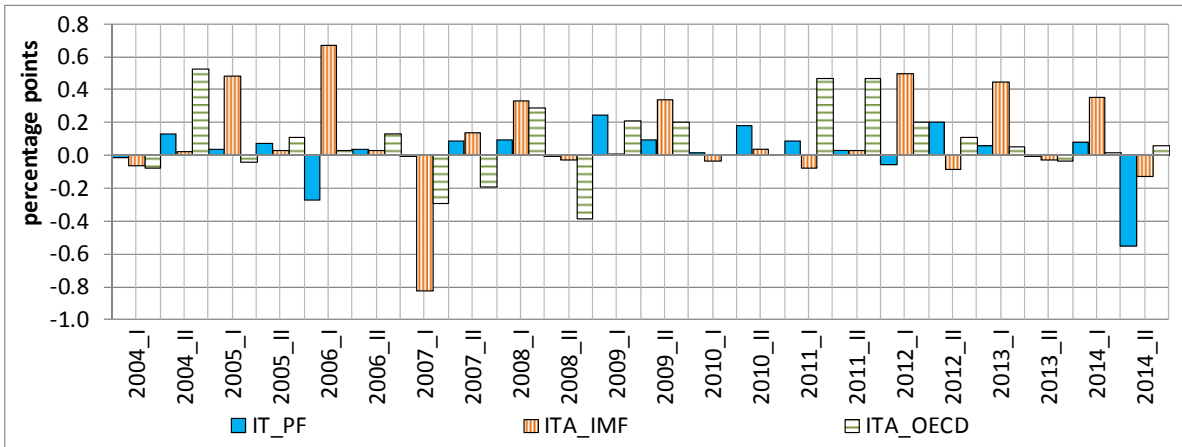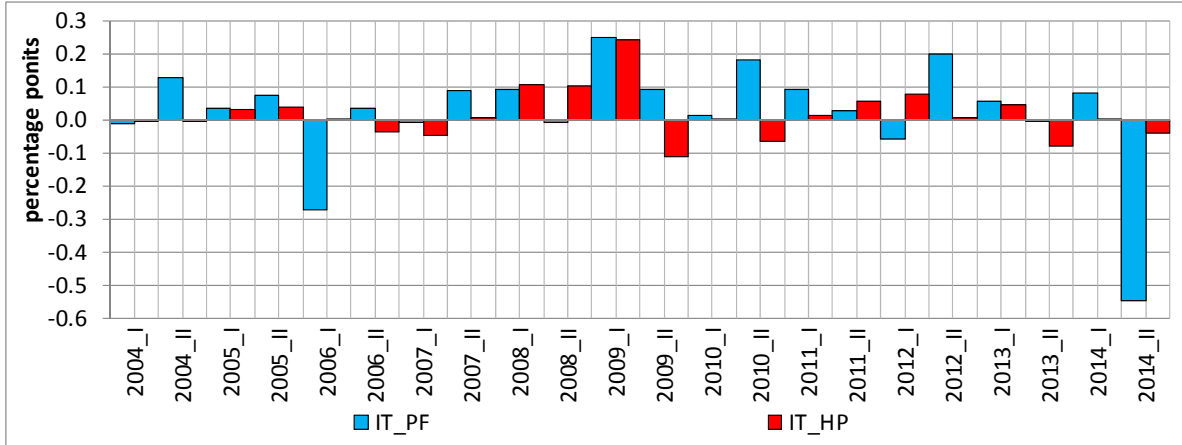
Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,
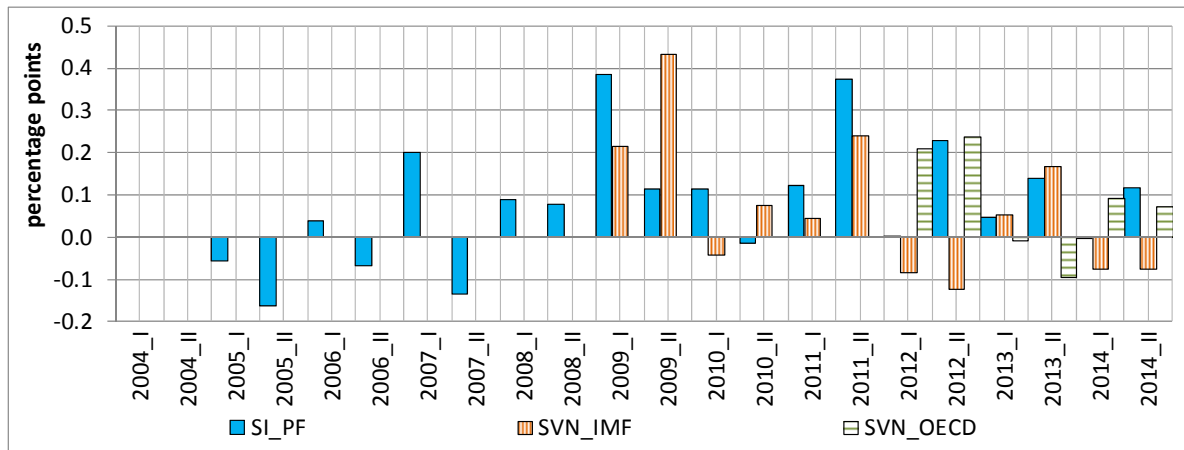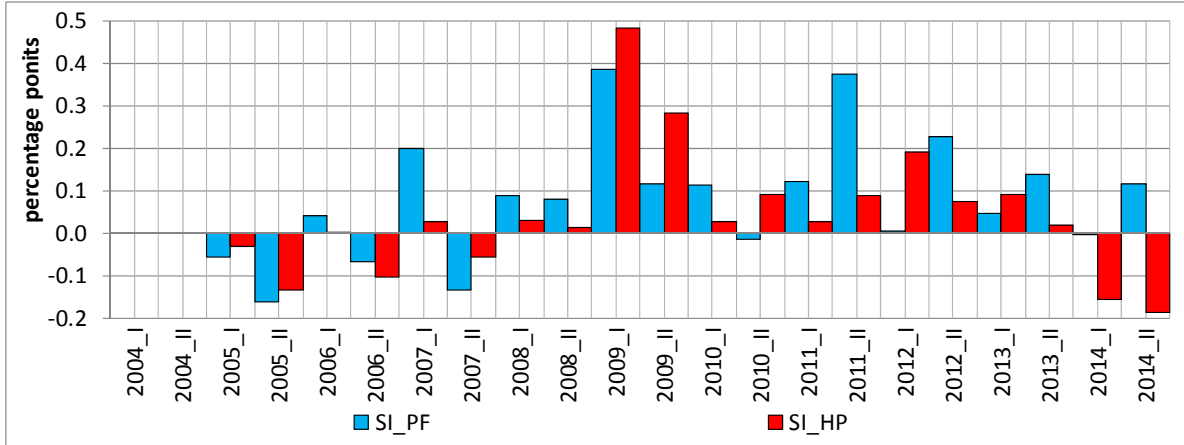### (All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: LU vintages of the IMF start in 2012, the OECD starts in 2006 and 2014 is missing
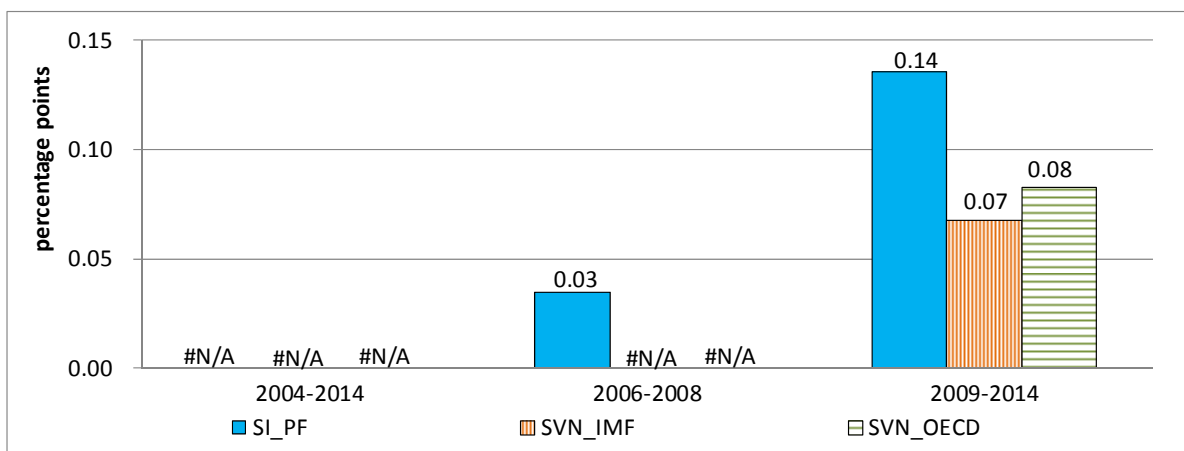
NL

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
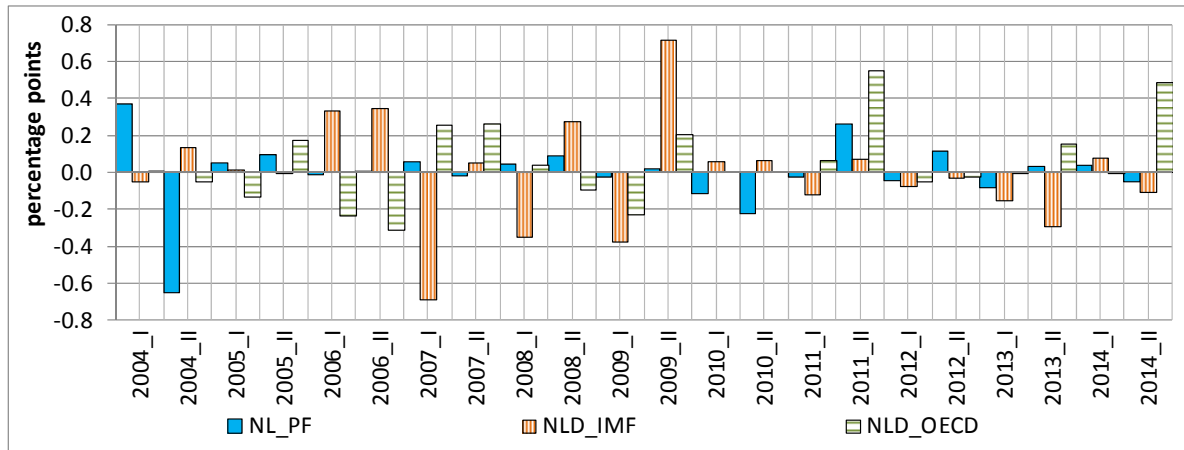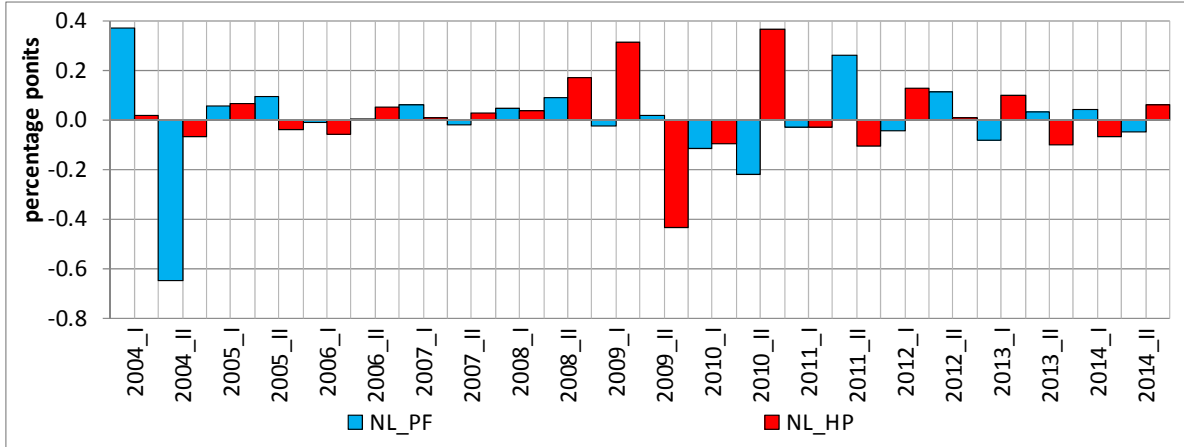
Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

AT

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

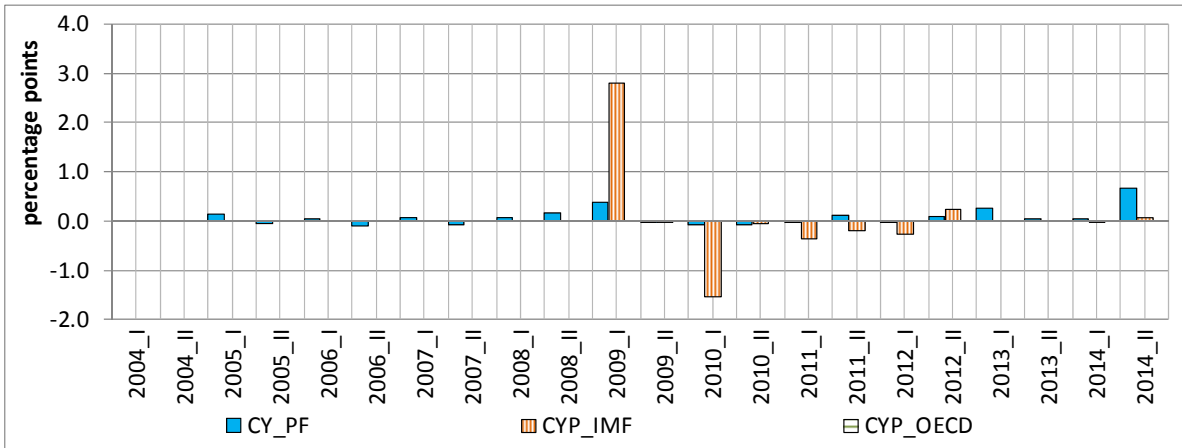
Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

PT

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
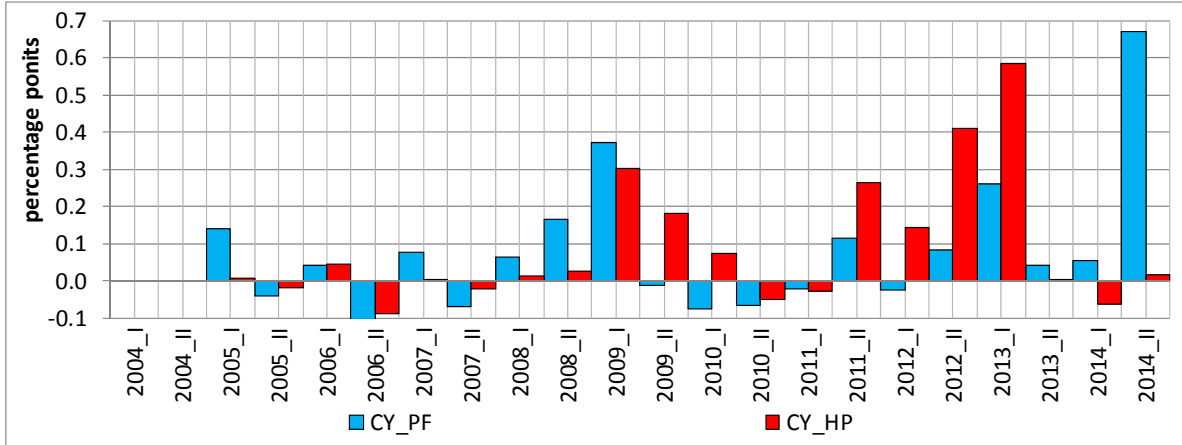

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
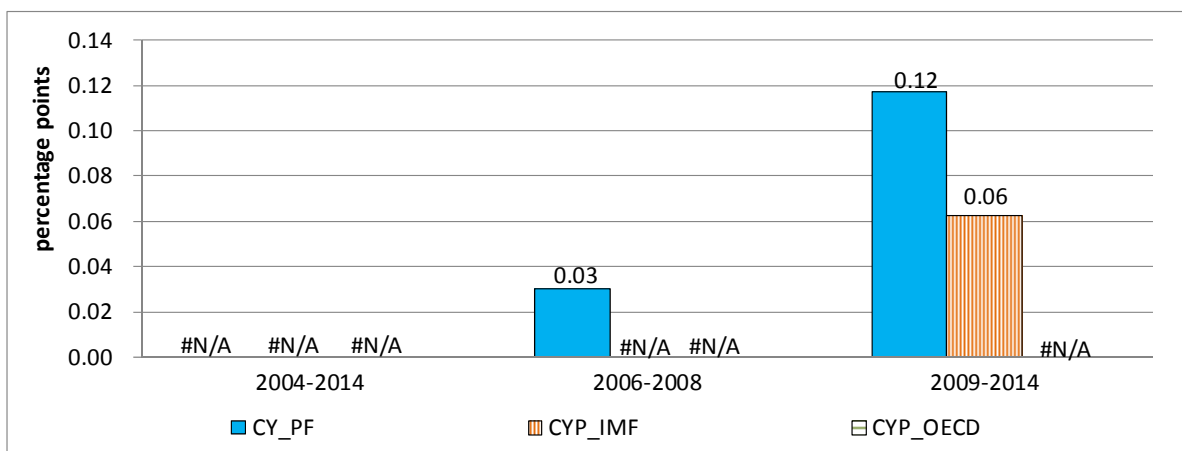
FI

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
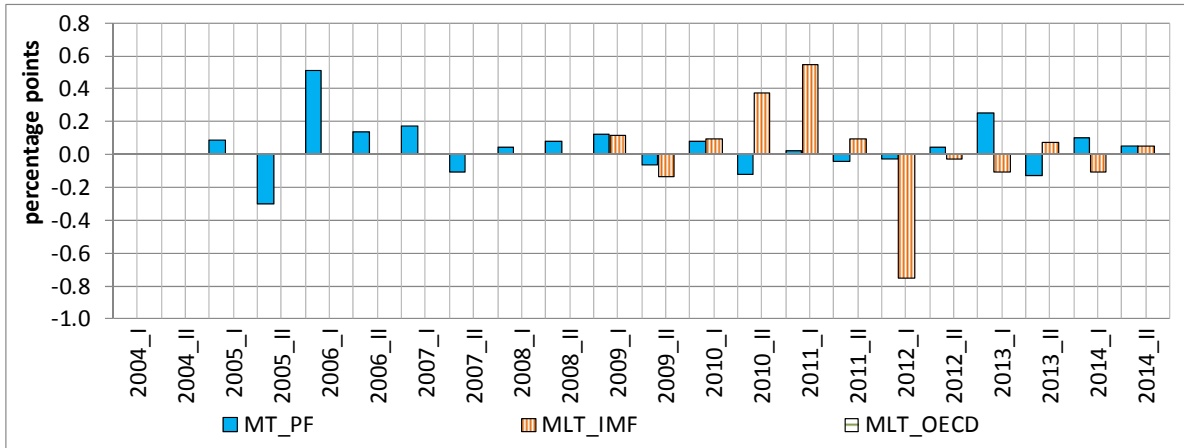

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

SE

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
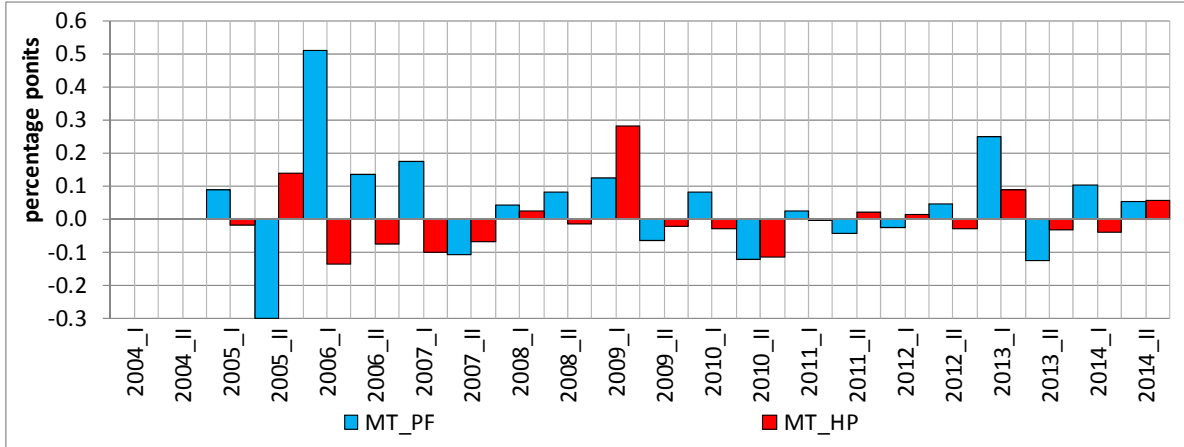

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
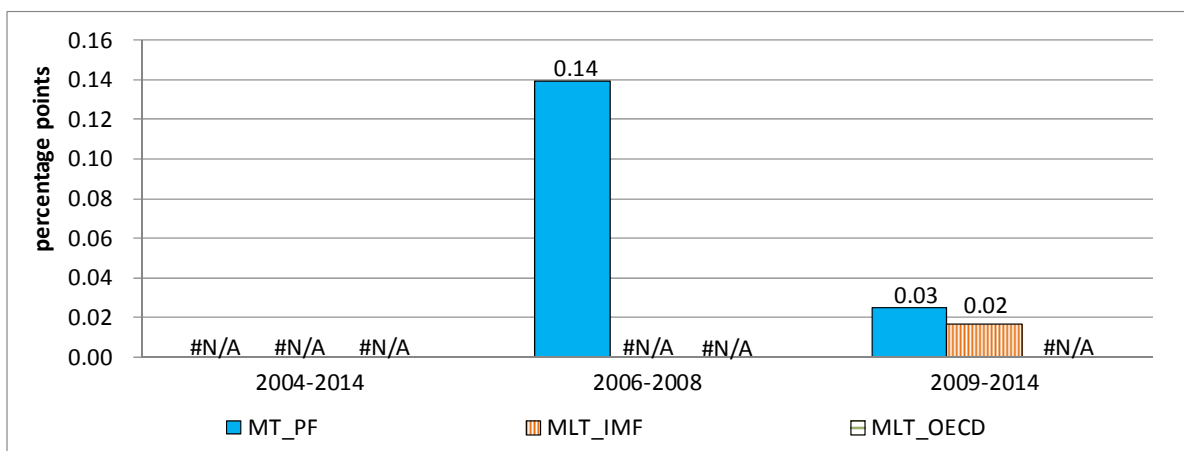
UK

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

CZ

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,
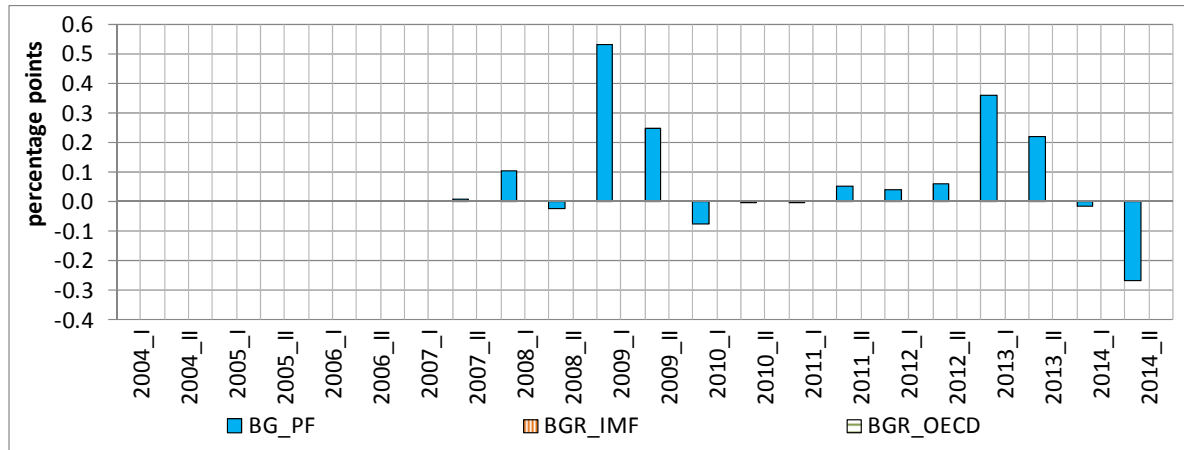### (All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
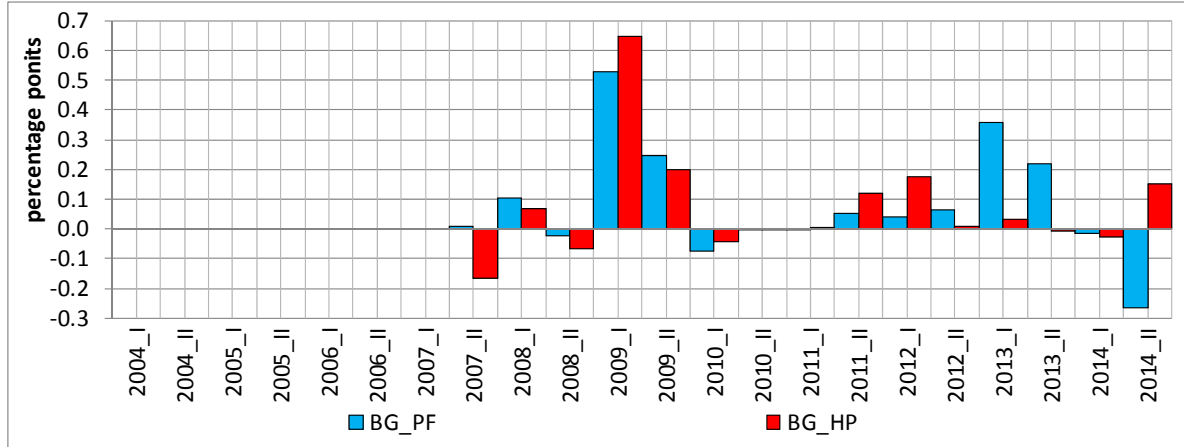

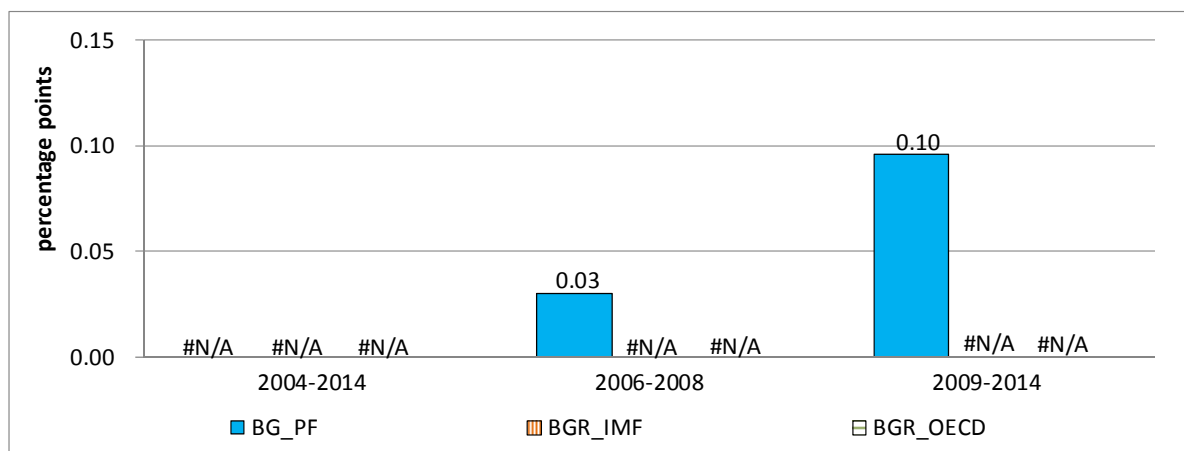note: no IMF vintages for CZ, the OECD starts in 2006

EE

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: EE vintages of the IMF start in 2012, the OECD starts in 2011

HU

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,
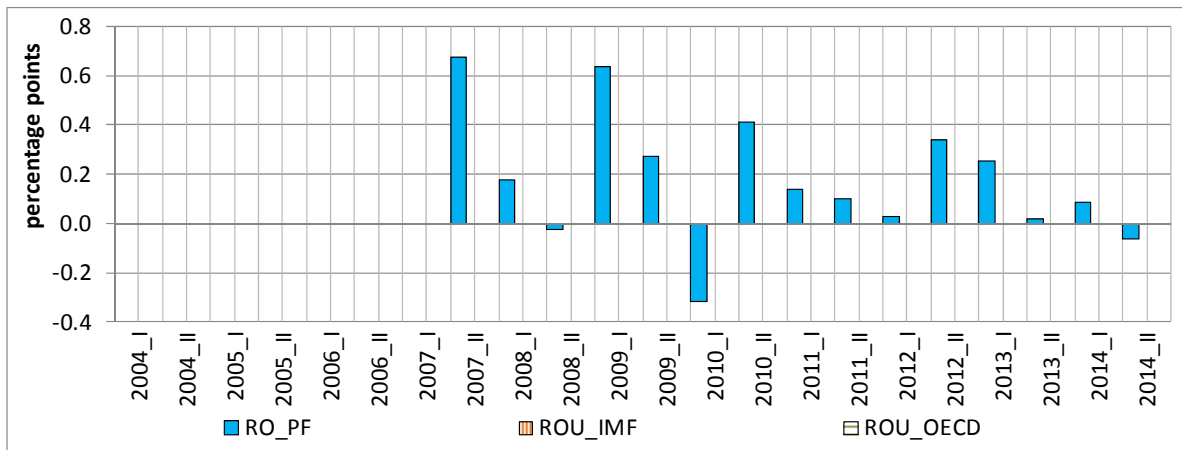(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
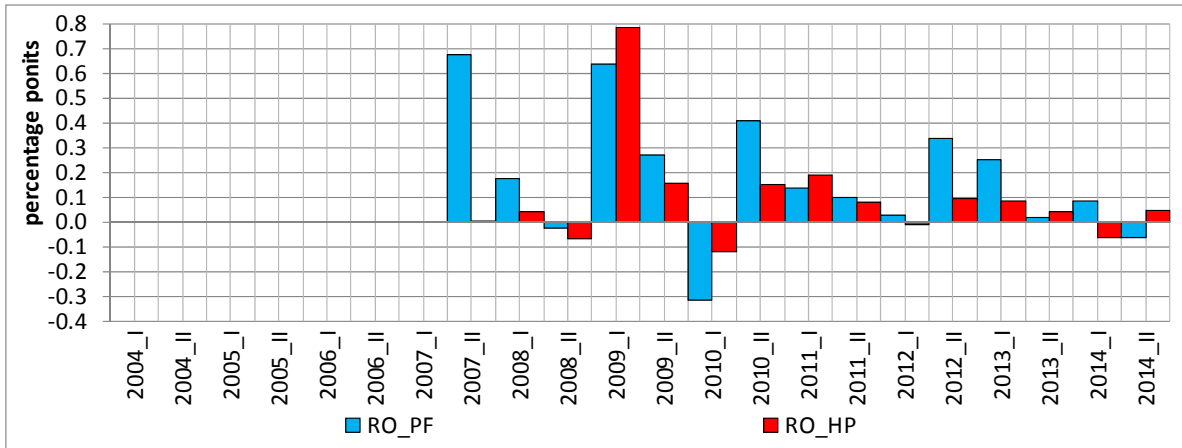

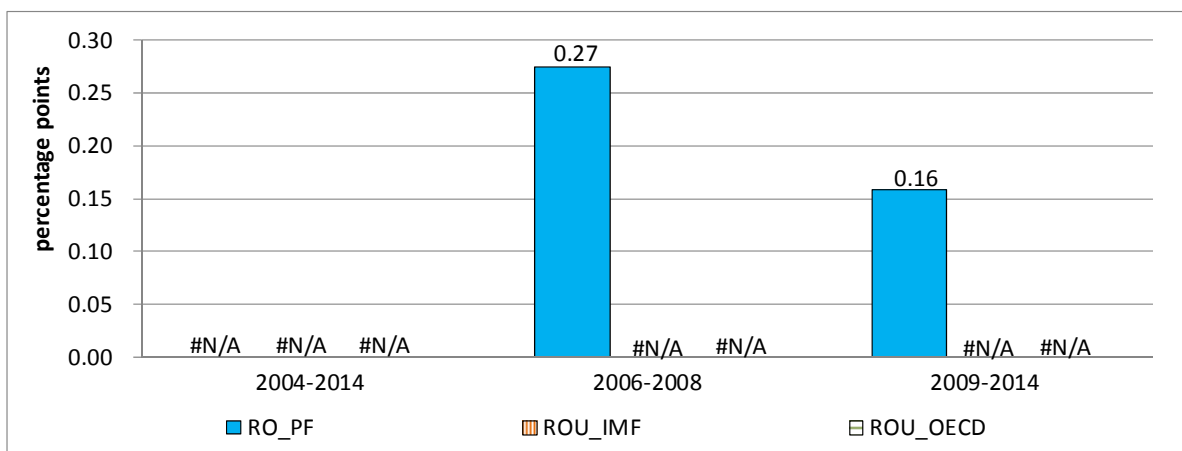note: no IMF HU vintages, the OECD starts in 2006

LV

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: no IMF and OECD vintages for LV

LT

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)
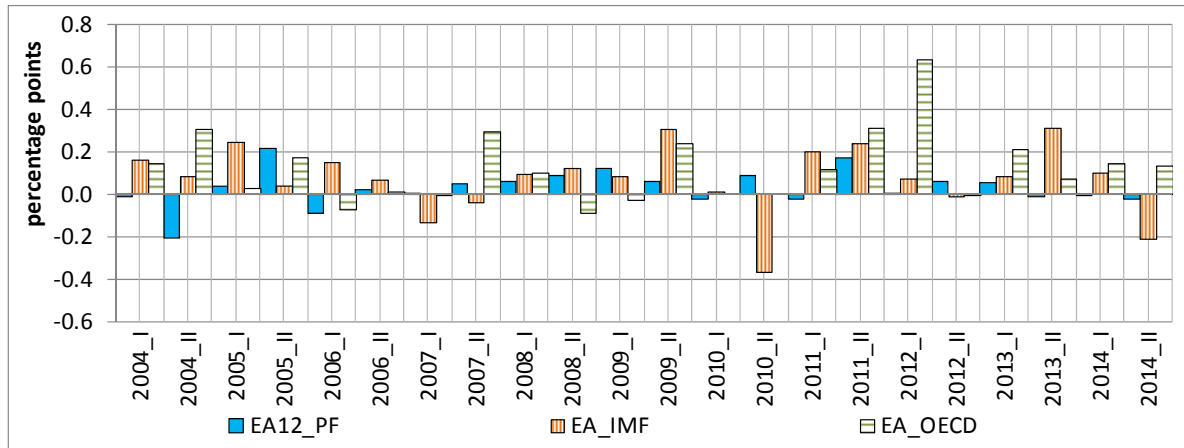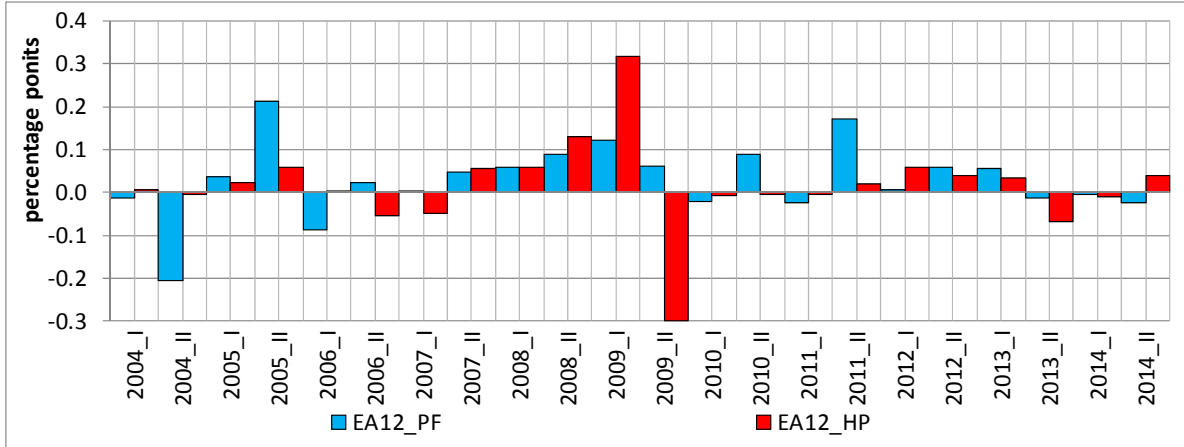

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: no IMF and OECD vintages for LT

41

PL

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: no IMF vintages for PL, the OECD starts in 2008

SK

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: IMF vintages for SK start in 2009, the OECD starts in 2011

SI

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: IMF vintages for SI start in 2009, the OECD starts in 2011

CY

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: IMF vintages for CY start in 2009, 2013 no data, no data for the OECD

MT

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: IMF vintages for MT start in 2009, no data for the OECD

BG

Short-Term Output Gap Revisions
(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,
(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: no data for IMF, no data for the OECD

RO

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)

note: no data for IMF, no data for the OECD

EA

Short-Term Output Gap Revisions

(each Forecast Vintage minus the preceding Forecast Vintage – Average of 14 Years of revisions)

Output Gap Period Average Revisions,

(All Vintages 2004-2014; Pre-crisis Vintages 2006-2008; Post-crisis Vintages 2009-2014)
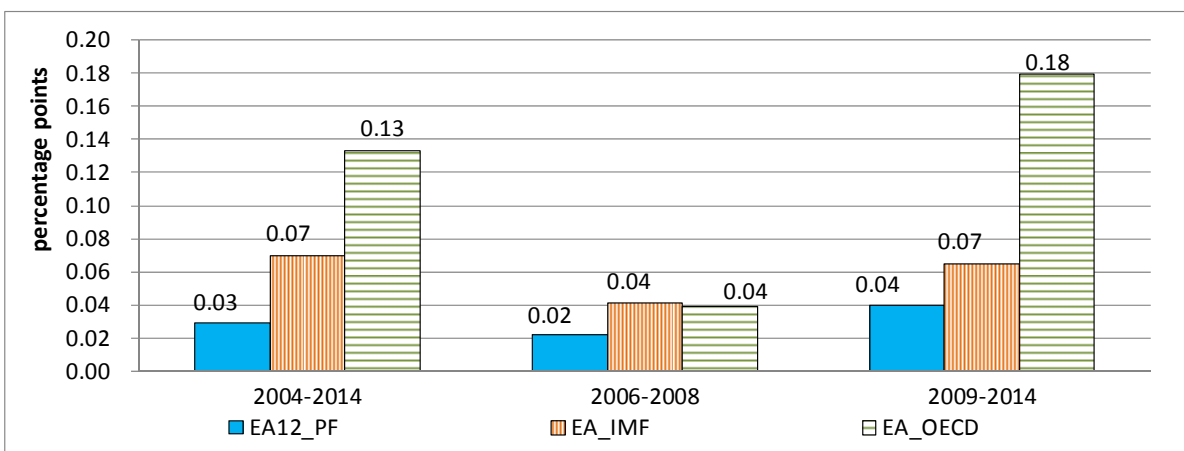
EU-PF Output Gap Estimates more highly correlated with EU-HP Estimates (in the pre- & post-crisis periods) than with IMF or OECD Estimates (Euro Area : 2004-2014)*
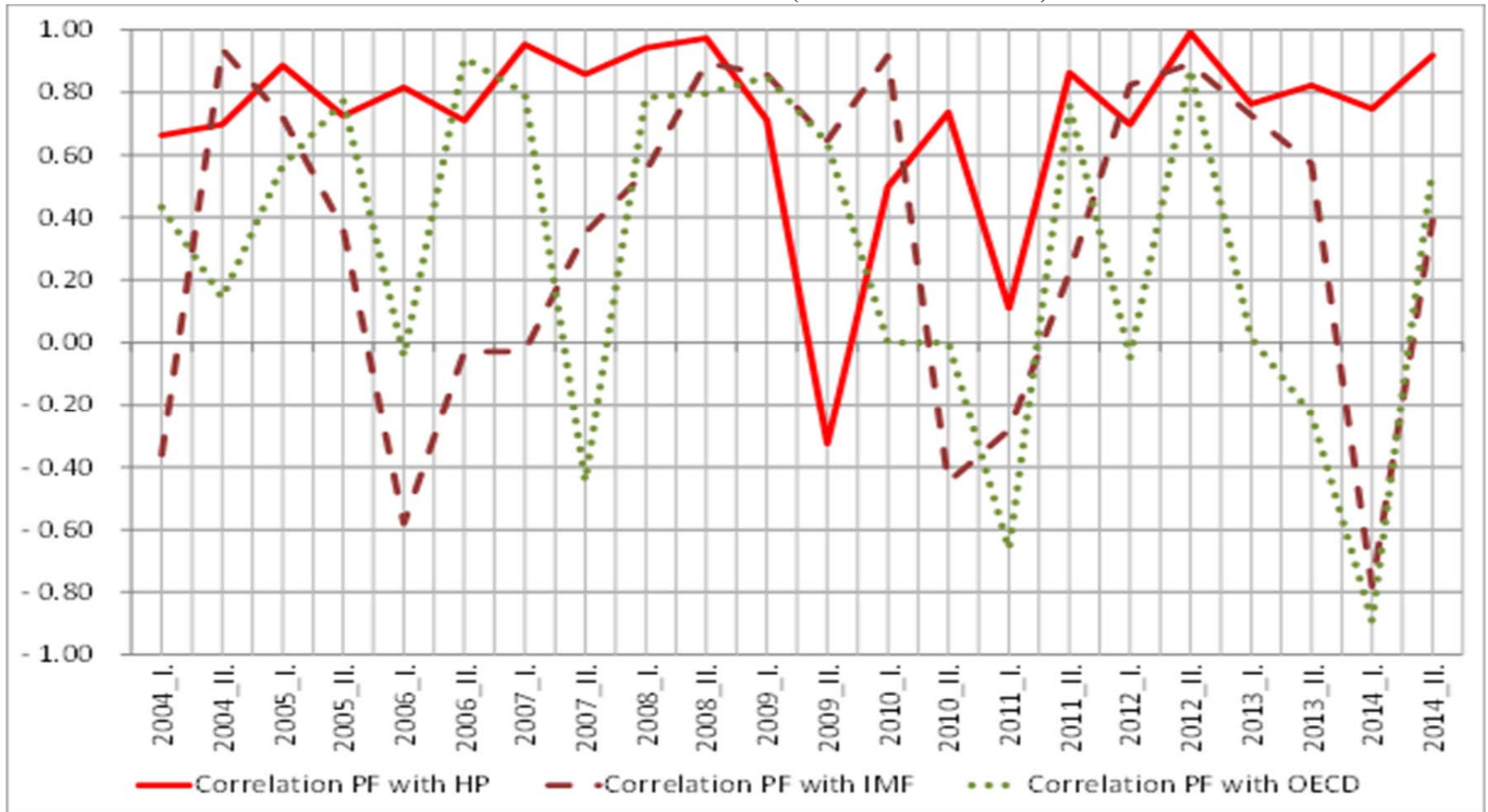
*Correlation between the 14 year average revisions

Long Term, Real Time, Reliability of the Output Gap Estimates produced by the EU PF, EU HP, OECD & IMF methodologies

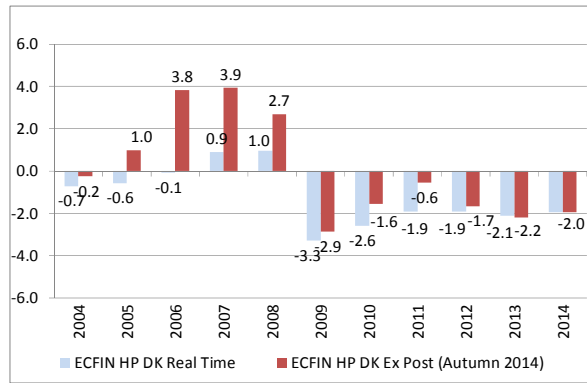**(Graphical Analysis for all 28 EU Member States plus Euro Area aggregate)**

BE

DE

DK

EL

ES

FR

IE

IT

LU

ECFIN Pf LU Real Time / ECFIN Pf LU Ex Post (Autumn 2014)

2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
-0.2	-0.2	1.0	3.9	1.5	-3.9	-1.9	-1.1	-1.8	-2.2	-2.1
-1.9	-1.6	-0.9	0.0	0.7	-5.5	-4.4	-2.2	-3.3	-3.1	

ECFIN HP LU Real Time / ECFIN HP LU Ex Post (Autumn 2014)

OECD LU Real Time / OECD LU Ex Post (Autumn 2014)

IMF LU Real Time / IMF LU Ex Post (Autumn 2014)

ECFIN Pf LU Real Time / ECFIN HP LU Real Time / OECD LU Real Time / IMF LU Real Time
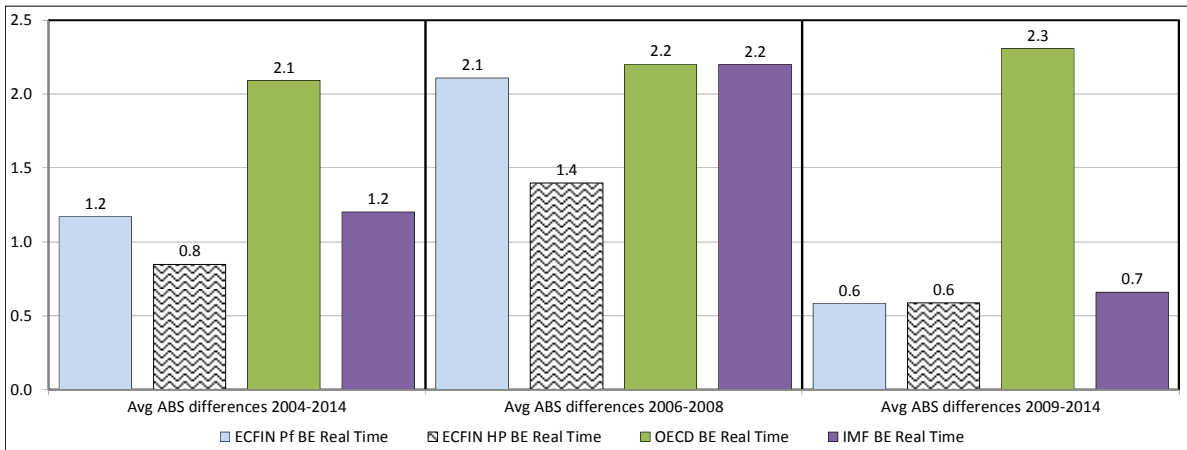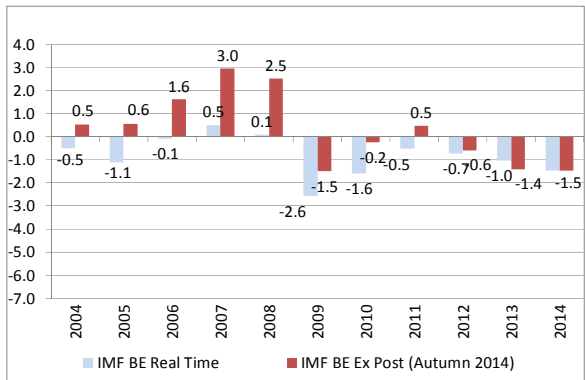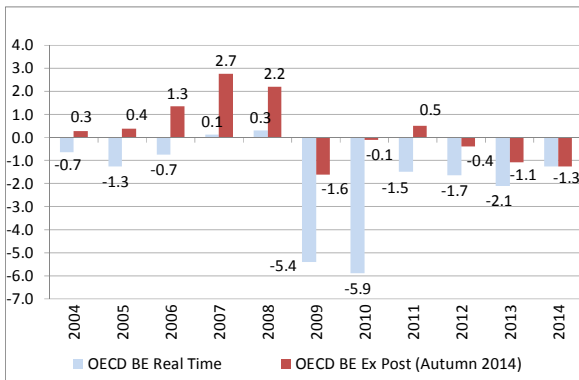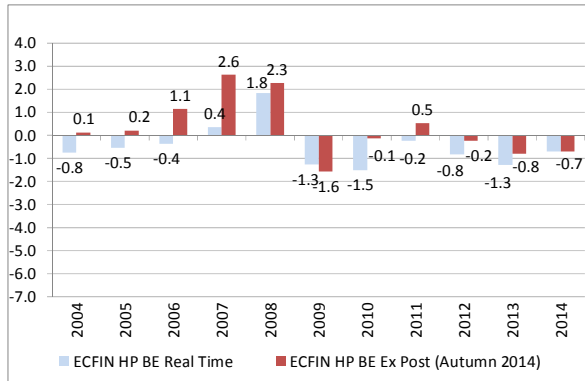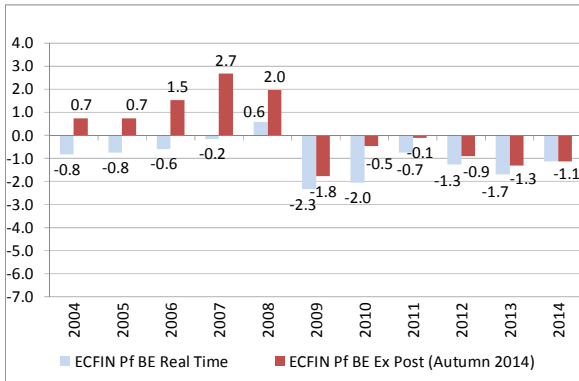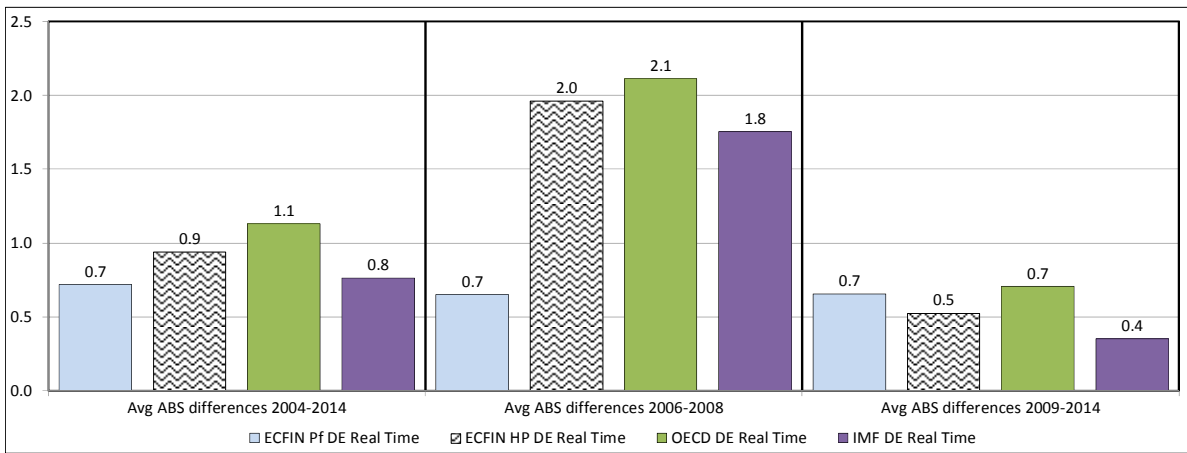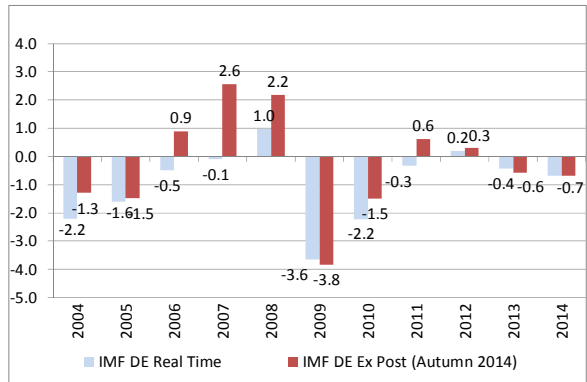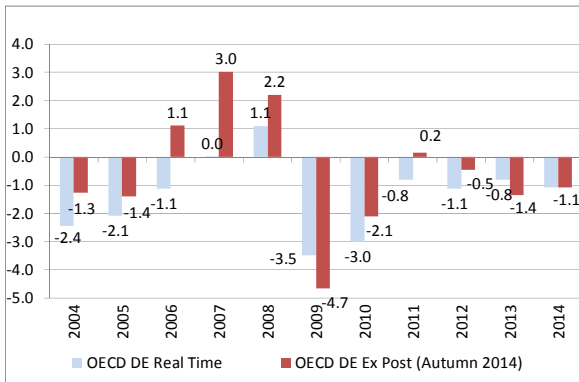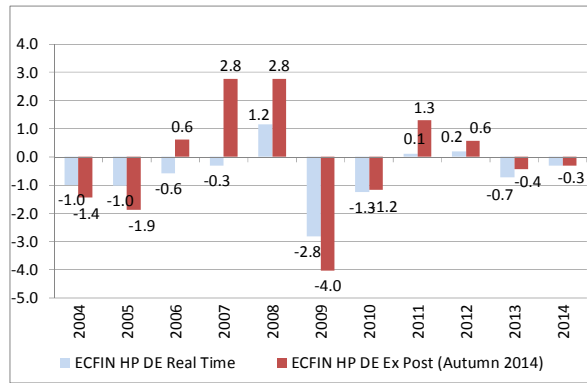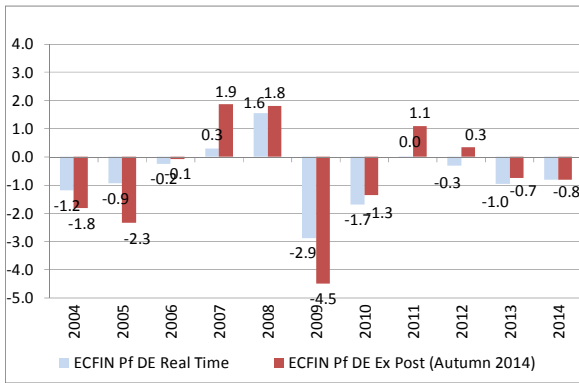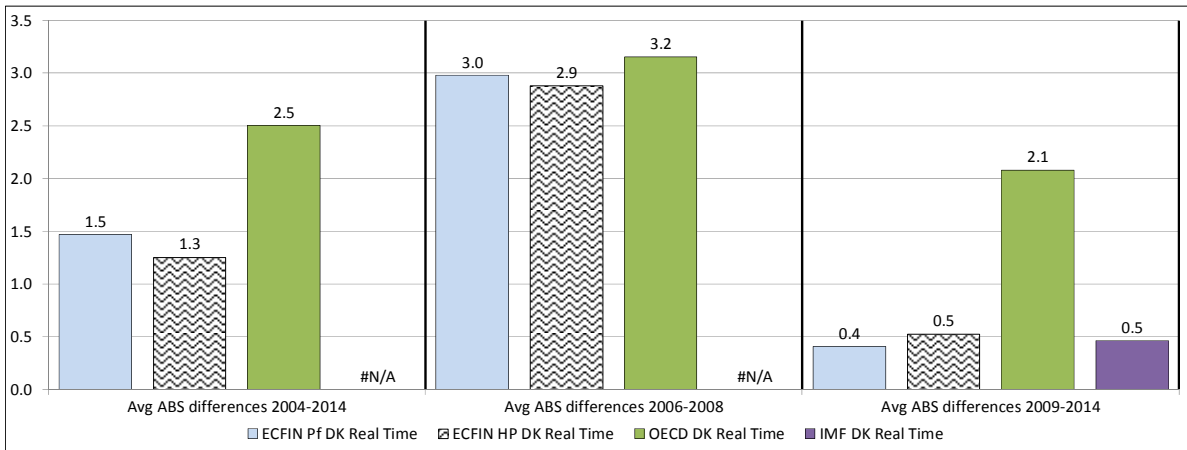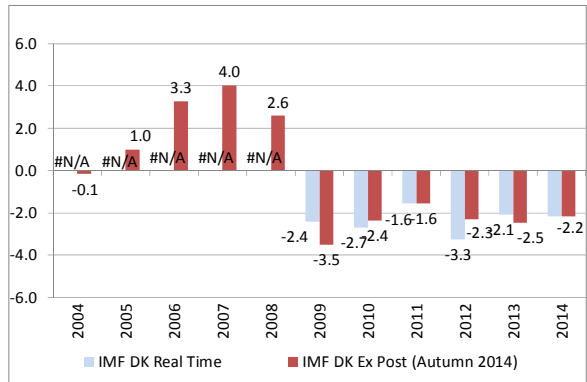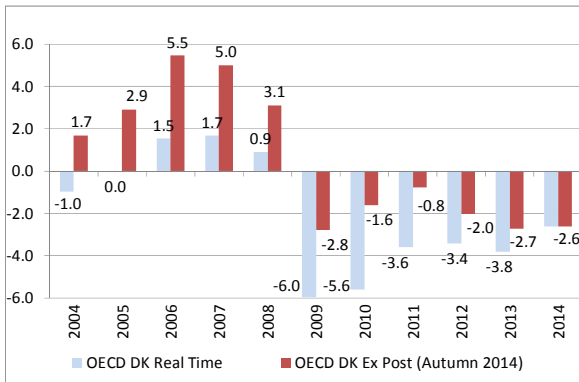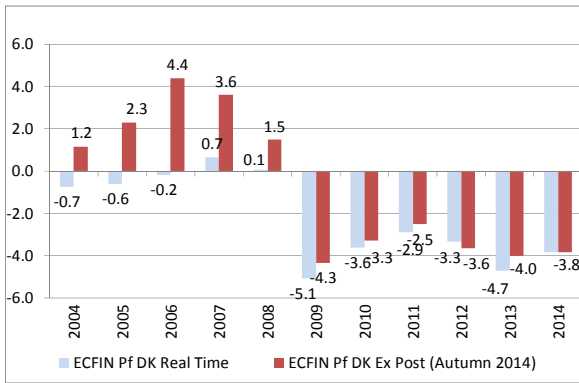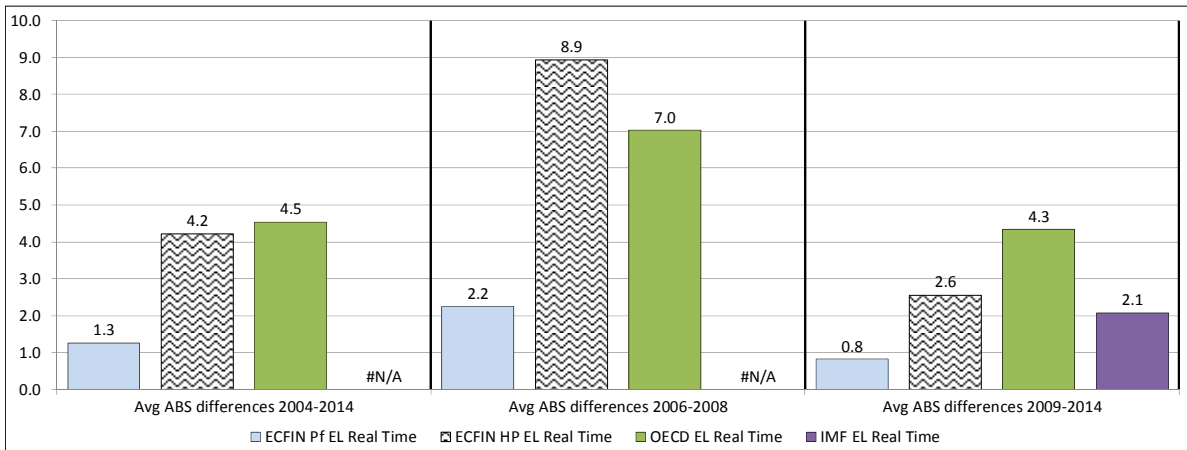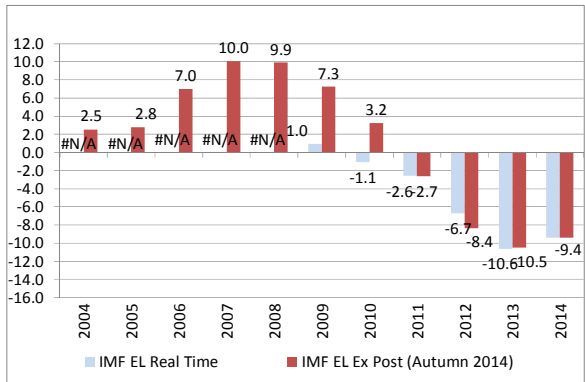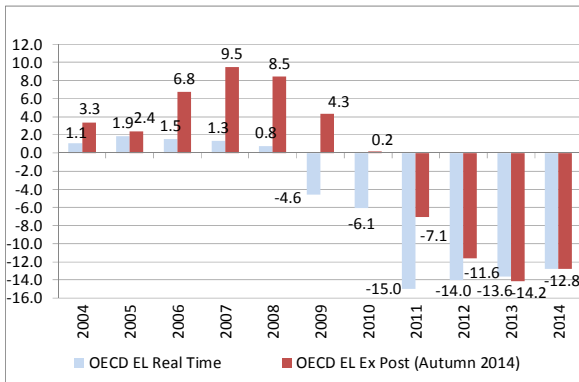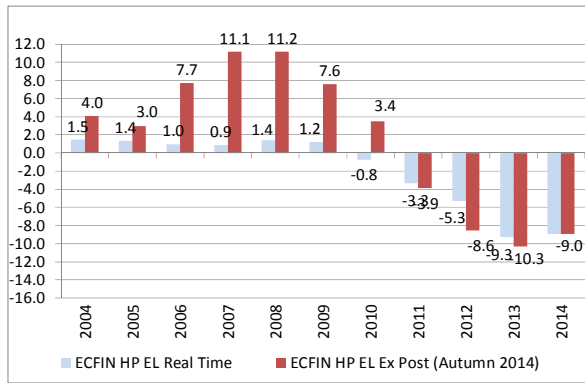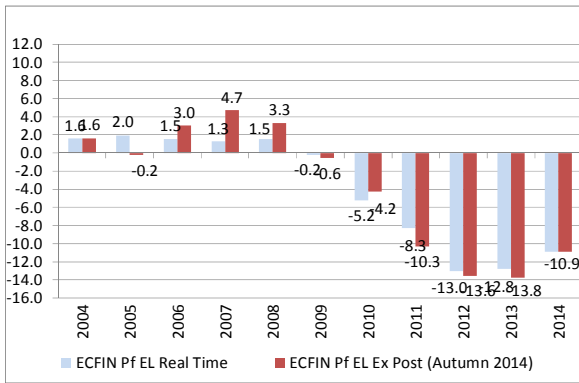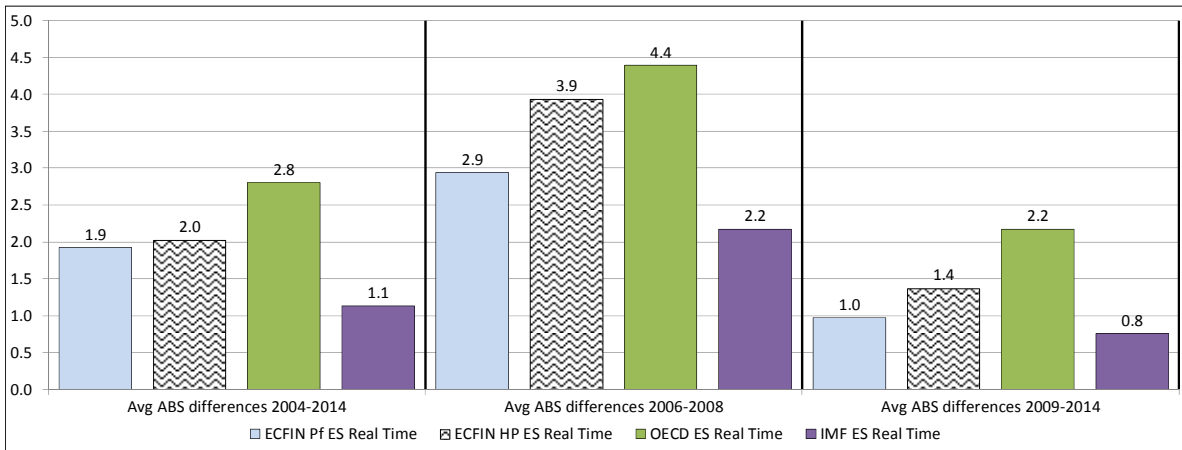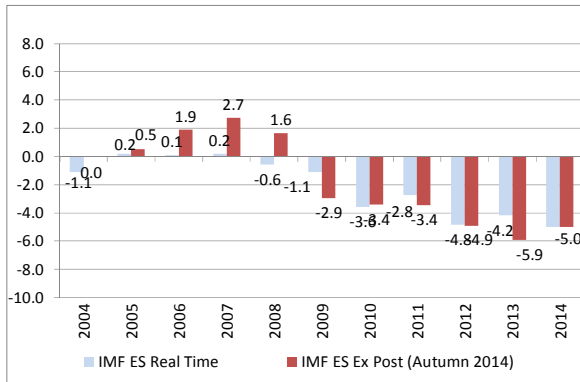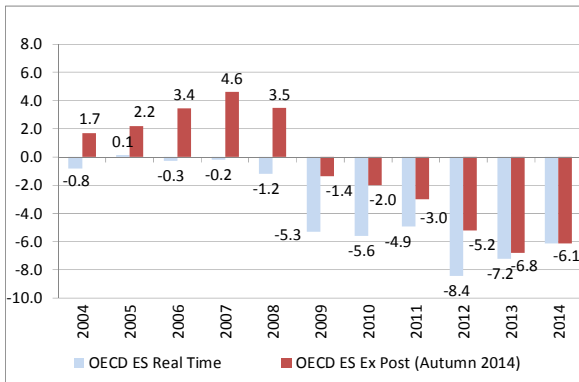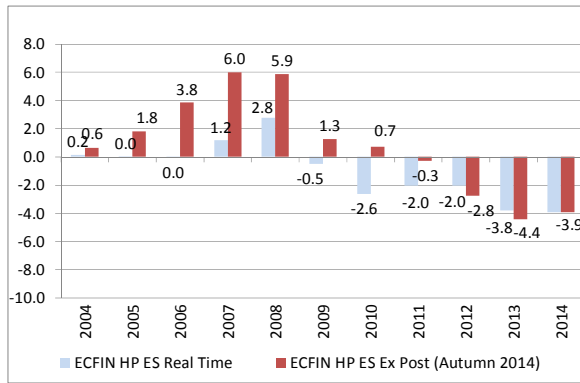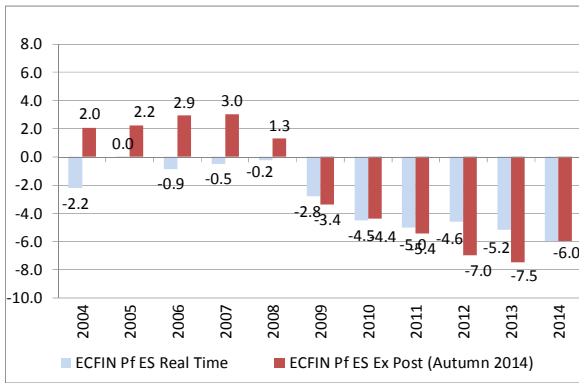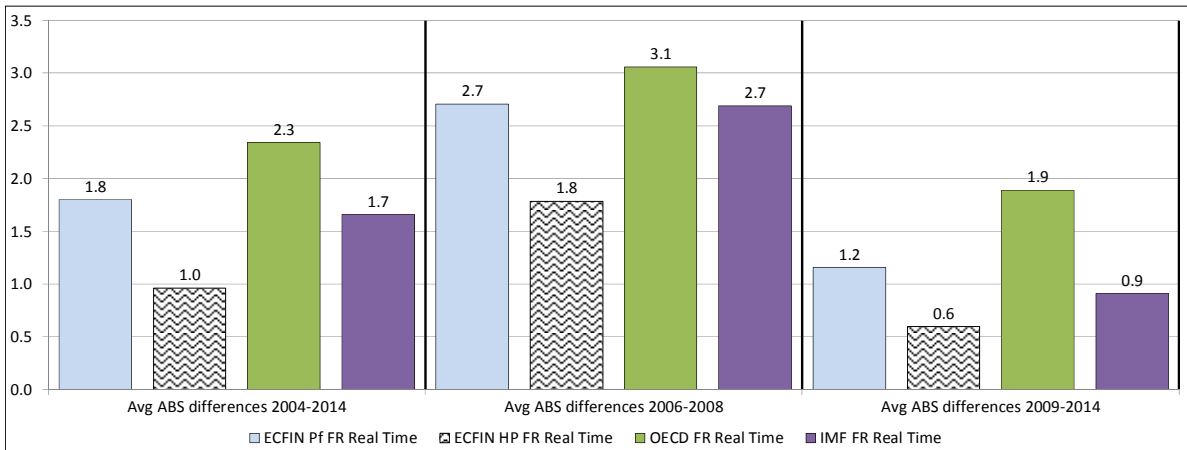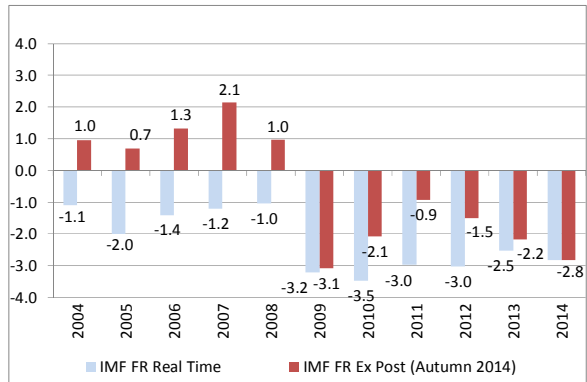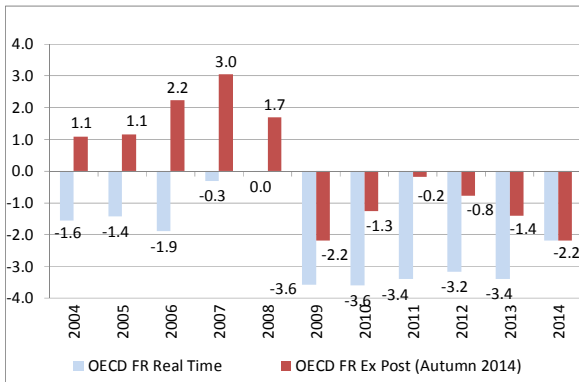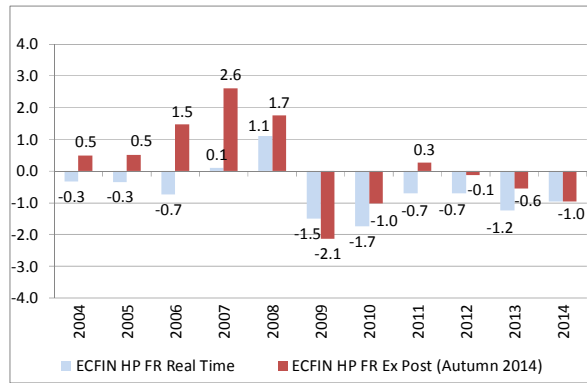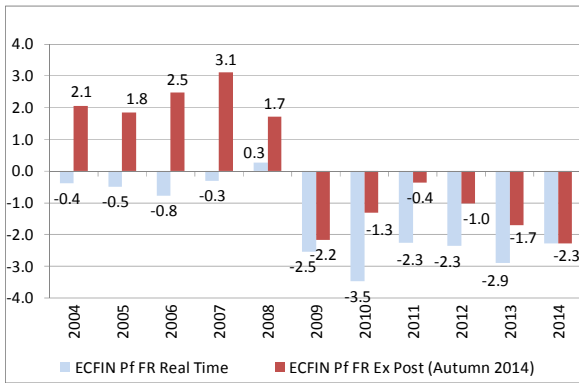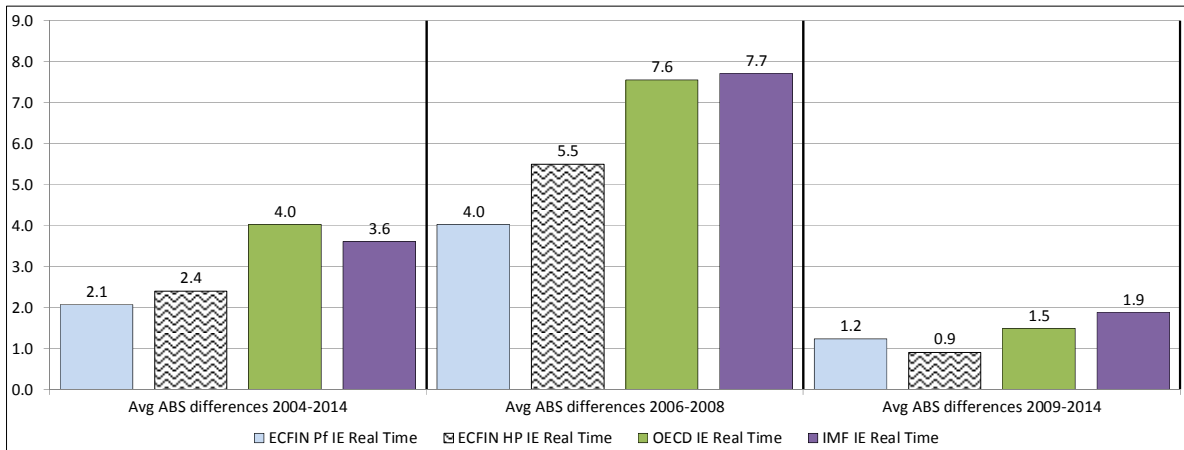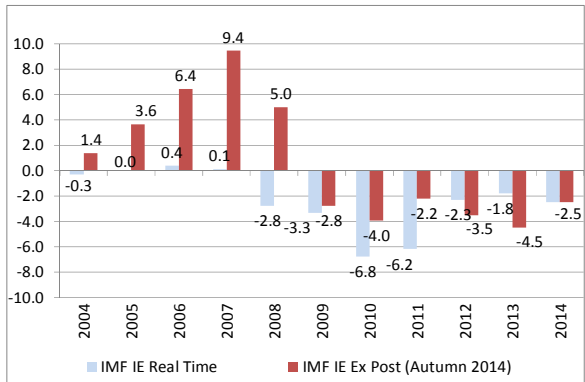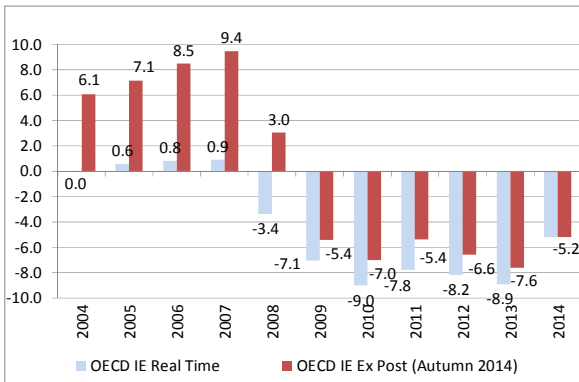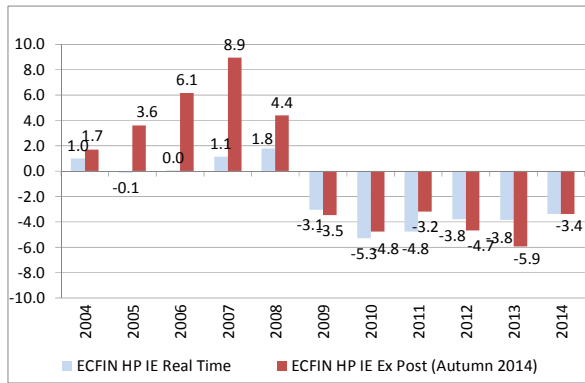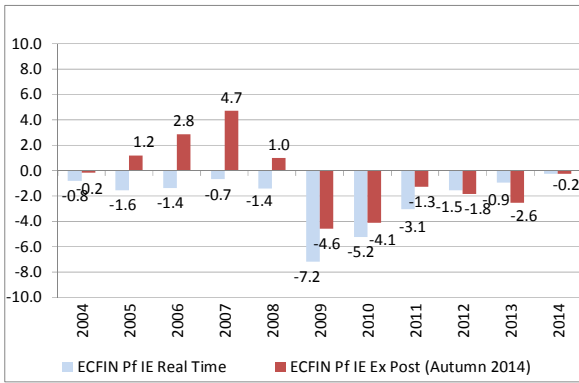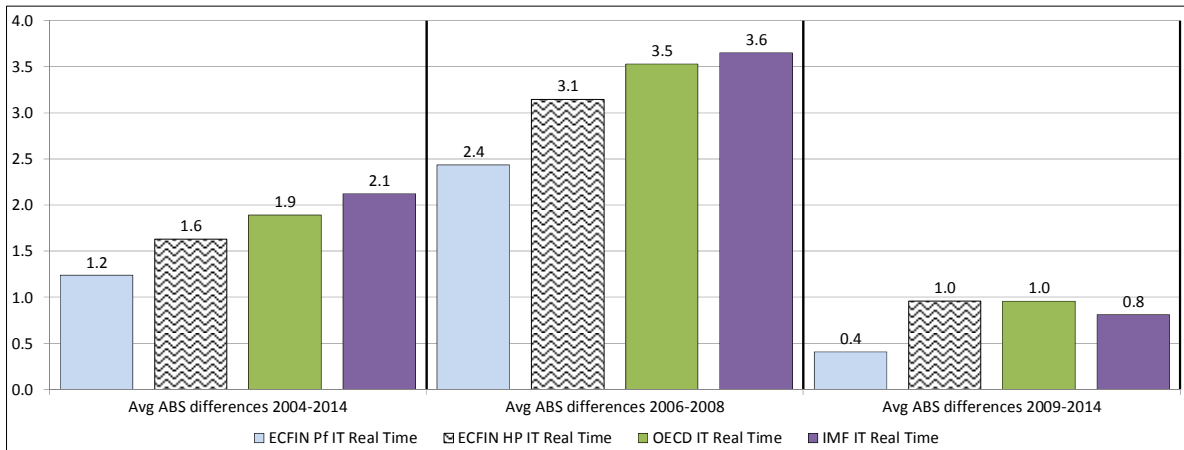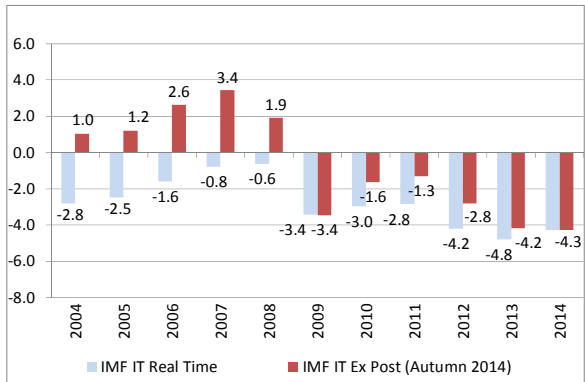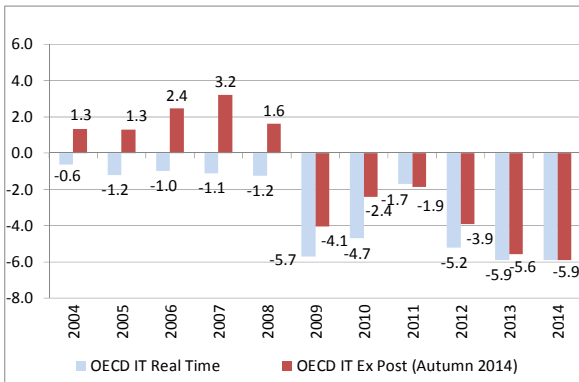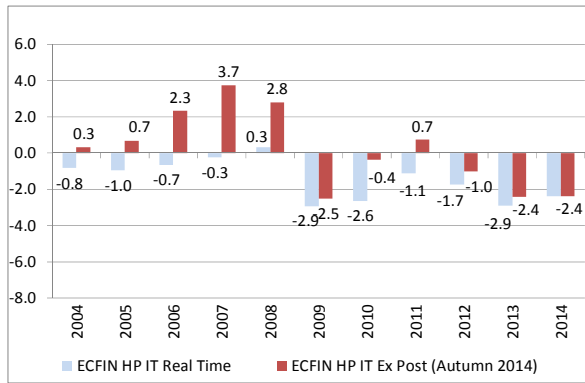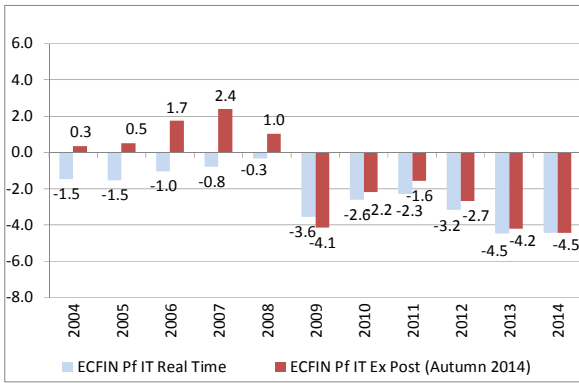
NL

AT

PT

FI

SE

UK

CZ

EE

HU

LV

LT

Chart (left): ECFIN Pf LT Real Time / ECFIN Pf LT Ex Post (Autumn 2014)

Year	Real Time	Ex Post
2004	2.1	1.9
2005	2.2	3.8
2006	1.4	5.1
2007	1.0	9.1
2008	4.4	6.4
2009	-8.2	-10.1
2010	-7.1	-8.5
2011	-3.1	-3.6
2012	-2.0	-1.0
2013	0.0	0.0
2014	0.2	

Chart (right): ECFIN HP LT Real Time / ECFIN HP LT Ex Post (Autumn 2014)

Year	Real Time	Ex Post
2004	1.6	1.4
2005	1.8	3.8
2006	2.2	6.5
2007	2.2	13.7
2008	7.2	12.8
2009	-6.1	-6.6
2010	-7.3	-7.4
2011	-4.0	-4.8
2012	-3.8	-2.5
2013	-1.7	-1.5
2014	-0.9	

	ECFIN Pf LT Real Time	ECFIN HP LT Real Time	OECD LT Real Time	IMF LT Real Time
Avg ABS differences 2004-2014	1.9	2.4	#N/A	#N/A
Avg ABS differences 2006-2008	4.6	7.1	#N/A	#N/A
Avg ABS differences 2009-2014	0.8	0.5	#N/A	#N/A

PL

SK

SI

CY

MT

BG

RO

HR

EA12

ANNEX 3 : THE POTENTIAL FOR IMPROVING THE BIASEDNESS PERFORMANCE OF THE METHOD BY RE-OPENING THE NAWRU CENTERING / NON-CENTERING DEBATE

The relatively good performance of the method has occurred despite the fact that the Spring 2014 changes to the NAWRU methodology unfortunately did not address the centering issue. This annex provides evidence that non-centering of the NAWRU has had negative implications in terms of the biasedness record of the method. As stated at the time of the discussions in the EPC, the Commission was in favour of imposing centering & underlined their intention to get the OGWG to examine this issue again in the future since it continues to represent a serious distortion in the overall method.

Non-centering of the NAWRU generates an explicit, negative, bias in the output gap estimates : The EU's method has done a relatively good job despite the bias introduced in the method by not applying a zero mean restriction to the NAWRU calculations. Graph 1 provides an estimate of the magnitude of the bias introduced by applying this approach to the NAWRU estimations. Based on the Winter 2015 ECFIN forecast, the graph looks at the average output gap for the EU15 countries for the period 1980-2014 and for the new Member States for the period 1995-2014. On average, and for a long enough period, we would assume that the output gap is close to zero. Since the period for the new Member States is rather short, we can expect the average output gap to deviate from zero, especially since many of these countries have experienced strong booms in this period.

- The first bar in the graph reports the average output gap over the relevant time period based on an HP filtered potential output. We find that the HP filtered output gap is slightly negatively biased for the EU15 and often positively biased for the new Member States.

- Next to this, we report the average output gap based on the commonly agreed production function approach. We see a much stronger negatively biased output gap for the EU15 and often a positive bias for the new Member States.

- The production function approach includes a non-centred NAWRU, since there is a constant in the cycle specification for the TKP (Traditional Keynesian Phillips Curve) countries and a non-centering adjustment made for the NKP (New Keynesian Phillips Curve) countries.

- In the third bar of the graph we therefore show the output gap based on a potential output estimate using a centred NAWRU. To calculate this we subtract the constant from the TKP NAWRU and the adjustment factor from the NKP NAWRU, and re-estimate potential output and the output gap. We find that the bias can be reduced strongly for many EU15 countries. Since the adjustment factor and constant are always negative, and the average output gap for many new Member States was already positive, including the centered NAWRU leads to a larger positive bias.

Graph 1 : The graph below provides an estimate of the size of the bias introduced by not applying a zero mean restriction to the NAWRU calculations
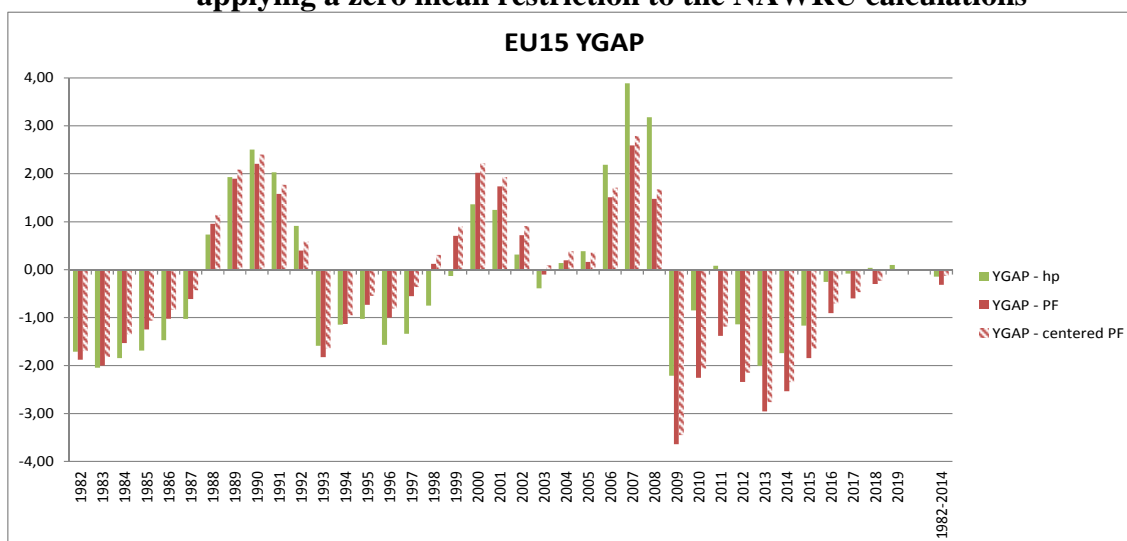
The main conclusions to be taken from Graph 1, on the impact of centering on the output gaps produced with the PF & HP filter methodologies, are as follows :

- The HP filter (which by definition has a zero mean), for all countries, produces a mean output gap which is closer to zero compared with both the uncentered & centered PF methods.

- For the PF method, for those countries with an average negative output gap over the period 1980-2014, centering always leads to a reduction in the negative gap.

- Finally, for the PF method, for those countries, with an average positive output gap over the same period (all new Member States which is probably due to the fact that we are working with a small sample of cycles, & a lot of the new Member States started off with a positive output gap), centering makes the mean of the positive output gap larger.

How do the above real-time bias estimates compare with the equivalent estimates quoted in the 2012 Discussion Paper by G. Kempkes[13] ? : The Kempkes paper analysed real-time output gaps for the EU-15 countries over the 1996-2011 period, as estimated by the EU, the IMF & the OECD & found that these estimates were negatively biased. The magnitude of the bias, according to Kempkes, is considerable : "on average, real-time cyclical components as a percentage of GDP are biased downwards by about 0.5 percentage points per year" (implying that all three institutions have an equivalent optimistic bias of ½ a percentage point with respect to potential growth rates)[14]. In order to provide a comparison with the Kempkes

---

[13] Kempkes, G; (2012) "Cyclical adjustment in fiscal rules : some evidence on real-time bias for EU-15 countries", Discussion Paper No 15/2012, Deutsche Bundesbank.

[14] At the Sept 2015 Joint EPC-ECFIN-JRC Workshop on "Assessment of the real time reliability of different output gap calculation methods", G. Kempkes (Bundesbank) argued that symmetric estimation of OG's is crucial for cyclical adjustment of government fiscal balances. If unbiased real time estimation of output gaps are not feasible, fiscal control accounts for cyclical components should be considered by policy makers (at the moment there is a large & significant bias towards negative OG's in real time which, if implemented in a fiscal rule, would have resulted in significant increases in debt ratios in many EU countries over the last 20 years – 15 percentage points on average).

results, we calculated a weighted average of the adjusted and unadjusted output gaps for the same sample of countries used by Kempkes, (i.e. the "old" EU15 Member States - DE FR IT ES IRL AT NL BE FI SE PT UK DK EL & LU) & for the same time period used for the calculations in Graph 1. The important point to stress is not that the overall bias is lower compared with our estimates (which is not that surprising since the time period covered by the two datasets are not comparable – the Kempkes paper covers the period 1996-2011, whereas our calculations cover a much longer period 1982-2014). What is more interesting to highlight is that if one were to impose the centered option, the overall bias would be only about 1/3 of the level of the uncentered option, comparable in fact to the bias associated with the HP filter. This also implies that the real-time optimistic bias for potential growth rates would correspondingly be reduced to an average of only 0.1%.

Graph 2 :. The graph below provides an estimate of the size of the bias introduced by not applying a zero mean restriction to the NAWRU calculations.
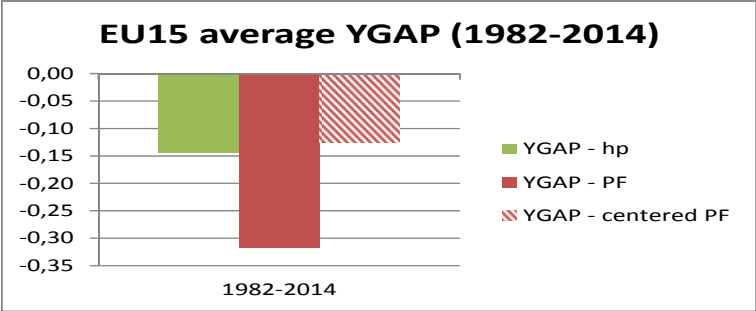
ANNEX 4: REPLICATING THE BUNDESBANK ANALYSIS (example of France over the period 1998-2010)

France - YGAP (PF)	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010		
2005_II.	-0,3	0,8	2,3	1,9	1,1	-0,2	-0,2								
2006_I.	-0,3	0,8	2,4	2,1	1,2	-0,1	0,0	-0,7							
2006_II.	-0,3	0,8	2,4	1,9	0,9	-0,1	0,0	-0,8							
2007_I.	-0,3	0,8	2,4	2,0	0,9	-0,1	0,1	-0,8	-1,0						
2007_II.	-0,3	0,8	2,2	1,8	0,7	-0,2	0,1	-0,2	-0,2						
2008_I.	-0,4	0,7	2,2	1,8	0,7	-0,2	0,3	0,0	0,1	-0,2					
2008_II.	-0,6	0,5	2,0	1,6	0,6	-0,2	0,4	0,4	0,8	1,0					
2009_I.	-1,3	-0,1	1,7	1,6	0,9	0,2	1,0	1,2	1,7	2,2	1,7				
2009_II.	-1,0	0,2	1,9	1,8	1,0	0,3	0,9	0,9	1,3	1,9	0,8				
2010_I.	-1,0	0,2	1,9	1,8	1,0	0,3	0,9	1,0	1,4	1,9	0,8	-2,7			
2010_II.	-0,1	1,2	3,1	2,9	2,1	1,3	1,8	1,7	1,8	2,1	0,4	-3,6			
2011_I.	-0,1	1,3	3,2	3,1	2,3	1,5	1,9	1,8	1,9	2,0	0,3	-3,9	-3,9		
2011_II.	-0,3	1,1	2,8	2,8	2,0	1,1	1,8	1,7	2,3	2,8	1,1	-2,8	-2,6		
2012_I.	0,0	1,3	2,9	2,9	2,0	1,1	1,8	1,7	2,3	2,7	1,1	-2,8	-2,5		
2012_II.	-0,1	1,2	2,9	2,8	1,9	1,1	1,8	1,7	2,4	2,9	1,4	-2,8	-2,2		
2013_I.	0,1	1,4	3,0	2,9	2,0	1,1	1,7	1,7	2,4	3,0	1,5	-2,7	-2,0		
2013_II.	-0,1	1,2	2,8	2,7	1,9	1,0	1,7	1,6	2,4	3,0	1,4	-2,7	-2,0		
2014_I.	-0,4	1,0	2,7	2,7	2,0	1,2	2,0	2,0	2,8	3,4	1,8	-2,4	-1,8		
2014_II.	-0,5	0,8	2,5	2,6	2,0	1,1	2,1	1,8	2,5	3,1	1,7	-2,2	-1,3		
sign chang	1	1	0	0	0	1	1	1	1	1	0	0	0	7	
Min. value	-1,3	-0,1	1,7	1,6	0,6	-0,2	-0,2	-0,8	-1,0	-0,3	0,3	-3,9	-3,9		
Max. value	0,1	1,4	3,2	3,1	2,3	1,5	2,1	2,0	2,8	3,4	1,8	-2,2	-1,3		
Spread	1,4	1,5	1,5	1,5	1,6	1,7	2,2	2,8	3,8	3,7	1,5	1,7	2,6	2,1	

Note 1: _I. Spring forecast, _II. Autumn forecast

Note 2: Quote from Bundesbank study "Only data on historical gaps are included in order to exclude revisions that are obviously the result of uncertainties in connection with macroeconomic forecasts. That means that looking for instance at the WEO of April 1999, the last estimate included in the analysis refers to the output gap in 1998".

Population of Working Age Growth (for 2009-2012 age group 15-64, afterwards 15-74)

Population of Working age, Avg 2009-2013

Total Hours Worked

Differences Total Hours Worked, Avg 2009-2013

Employment Growth

ABS Differences Employment, Avg 2009-2013
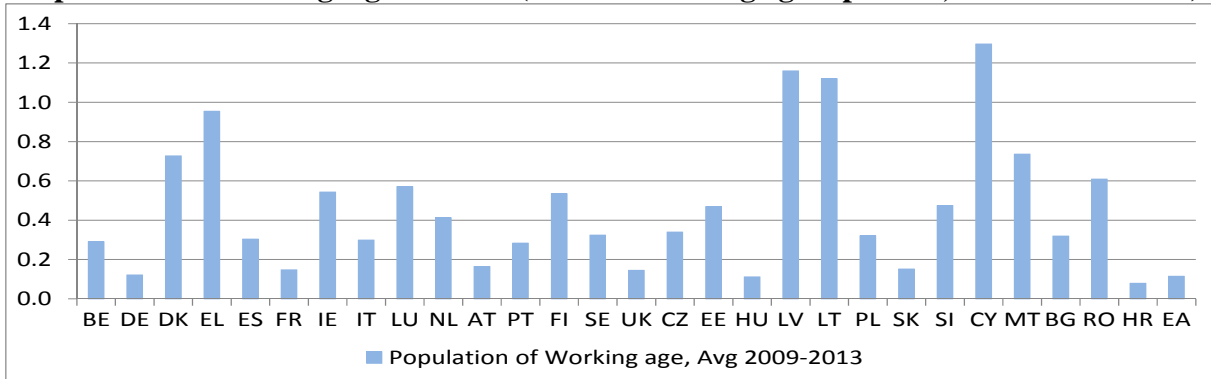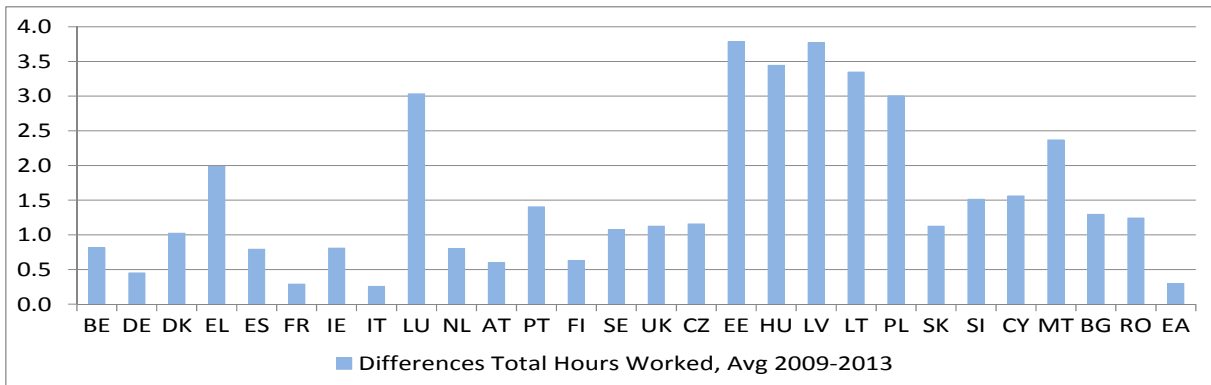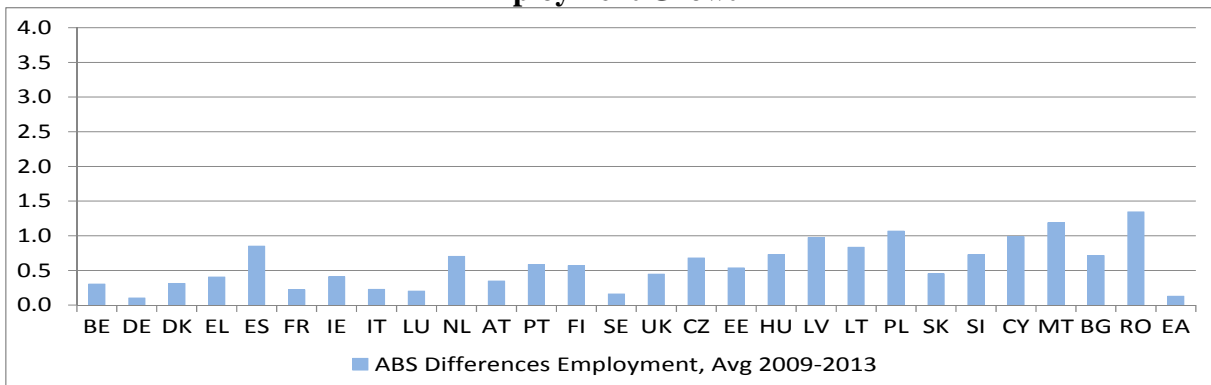
Annual Average Hours per Employee Growth

ABS Differences Annual Hours per Employee, Avg 2009-2013
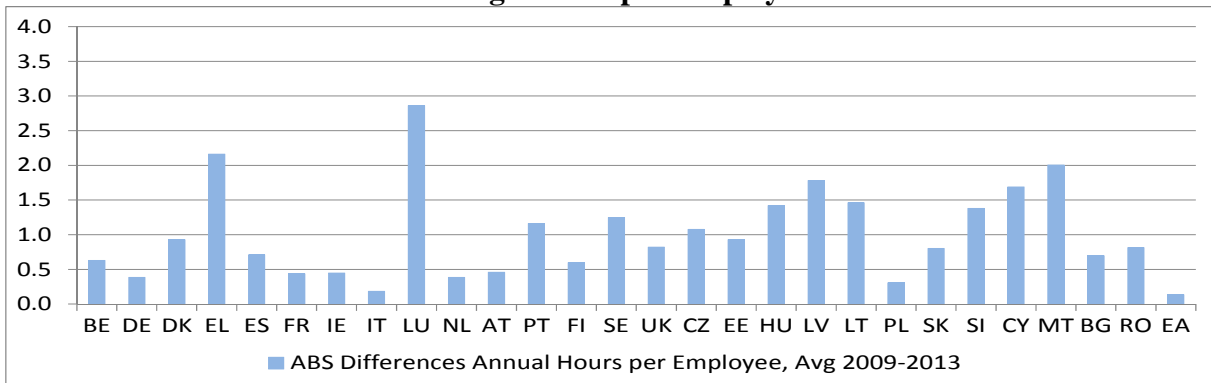
ANNEX 6: WHAT DRIVES THE OUTPUT GAP REVISIONS IN THE EU'S PF METHODOLOGY – IS IT DATA/FORECAST REVISIONS OR METHODO-LOGICAL REVISIONS?

This Annex looks at the source of the revisions in the Output Gap estimates and assesses the contribution of data and forecast revisions on the one hand & of large & small methodological changes on the other. It focuses on the key question as to whether the revisions are being driven by data / forecast revisions or methodological changes. Section 1 examines the impact of the 4 large methodological changes made to the method since 2002 on the revisions properties of the method. Section 2 then goes on to describe the impact of small methodological changes on revisions made by the Commission services during each forecasting exercise. The analysis in section 2 is carried out on the basis of the Spring 2015 forecasting exercise.

1 Impact on Output Gap Revisions of Large Methodological Changes : There have been four significant methodological changes made to the method since its inception in 2002. Before providing an assessment of the impact of these changes on revisions, the text below gives a short description of each of the changes :

a) Hours worked (Autumn 2005) : Total hours worked is the preferred measure of labour input in the national accounts but its measurement has been challenging due to the growing importance of service activities, self-employed jobs and the emergence of a range of new, often irregular, working patterns. Due to these measurement issues, its use in the PF methodology was delayed until the Autumn 2005 forecasts since there was an absence of datasets of sufficient quality for a large number of the Member States. Following the resolution of all of the country specific data issues, the hours worked series for the respective countries were successfully introduced in the Autumn 2005 forecasting exercise. In addition, given the associated joint OECD / Eurostat decision to use the national accounts (as opposed to the labour force survey) as the preferred source of labour input data, the method was also modified to take both the employment and hours worked input variables from this single source.

b) TFP (Autumn 2010) : The key driving force behind the adoption of the new Kalman Filter trend TFP approach was the persistent concerns, expressed over many years, with the operation of the previous HP filter methodology for calculating trend TFP, with the latter's limitations & problematic nature being particularly exposed during the financial crisis. The basic problem with the existing HP filter method was that such univariate techniques tend to produce imprecise estimates at the end of the sample period (& especially close to turning points / "boom-bust" episodes). Consequently, preliminary HP trend TFP estimates were frequently & sizeably revised over time. Whilst it was accepted that revisions would never be fully eliminated, the new TFP methodology was expected to offer a number of important advantages relative to the existing HP approach, including firstly, less trend TFP & TFP gap revisions; secondly, a more realistic pattern for short and medium term trend TFP developments; & finally, less end-point bias problems which can produce highly misleading signals around turning points (leading of course to large subsequent revisions).

c) Population of Working Age - POPW (Spring 2013) : In Spring 2013 it was decided to widen the definition of "working age" in the PF calculations to include people up to the age of 74. Since then POPW is defined as 15 to 74 (compared with the 15 to 64 definition used up

until Spring 2013). This change was introduced to align the PF method with the approach used by the EPC's Ageing Working Group for their long run ageing exercise.

d) NAWRU (Spring 2014) : Two NAWRU changes were introduced for the first time in the Spring 2014 forecasting exercise :

- **1. New technical extension rules for the NAWRU** : Instead of the previous extension rule for the medium term NAWRU of taking 50% of the change in the previous year, the new approach now takes 50% of the most recent NAWRU change in T+3, followed by a flat extension rule in T+4 and T+5.

- **2. New NAWRU specifications :** Following the proposal to introduce a non-centered NAWRU, based on the notion of an "all encompassing Phillips Curve", the EPC endorsed a new Keynesian Phillips Curve (NKP) specification for 21 of the 28 Member States, and the traditional Keynesian Phillips Curve (TKP) specification for the remaining 7 countries, namely Belgium, Germany, Italy, Luxembourg, Malta, the Netherlands and Austria. Bearing in mind the importance of the stability principle, the Commission committed itself to using these EPC endorsed NKP / TKP country preferences for a period of 3 years.

The timing of these 4 large methodological changes is shown visually in graph 1 below. The first point to stress is that, contrary to the widely held public perception, the method has not been continuously changed over the last 13 years – in fact, only 4 out of the 29 vintages stretching from Autumn 2002 until Spring 2015 were affected (note the 3 Winter vintages are not indicated in graph 1).

Graph 1 : Revisions in the PF Output Gap from forecast to forecast – focus on large methodological changes (mean over 14 years of data and standard deviations, Euro Area)
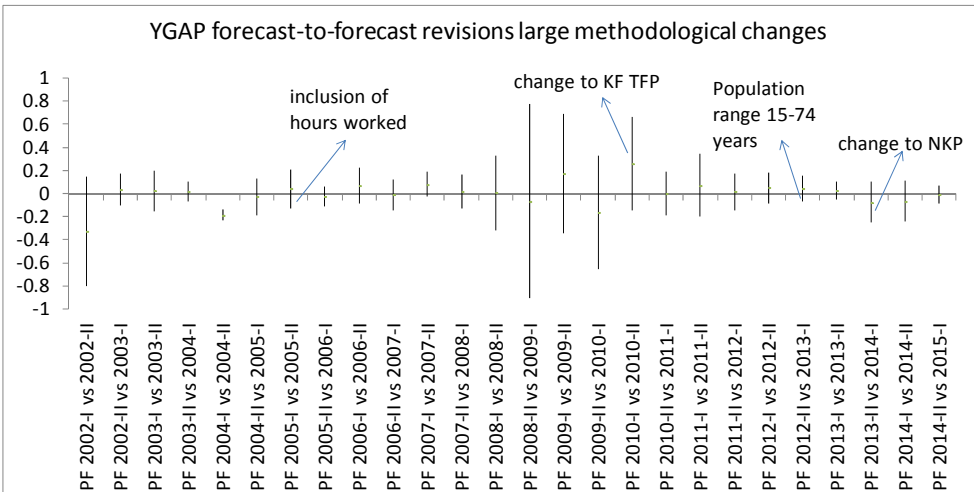
More importantly, as shown in graph 1, these methodological changes did not have much of an impact on the Output Gap revisions at the EA level for the vintage in which they occurred. In fact, as graph 1 stresses, the impact on revisions is only noticeable with respect to 2 of the "large changes" (i.e. the TFP & NAWRU changes). The introduction of hours worked in 2005 & the change to 15-74 in 2013 did not lead to any dramatic change in the forecast-to-forecast revisions. In addition, for the TFP and NAWRU changes, the result in

terms of revisions goes in conflicting directions, with the new TFP method in 2010 leading to a temporary deterioration in the (forecast to forecast) revisions performance, while the opposite occurred for the 2014 NAWRU change.

Graph 2 : Revisions in the PF Output Gap – impact of the 4 large methodological changes (mean and standard deviations, Euro Area)
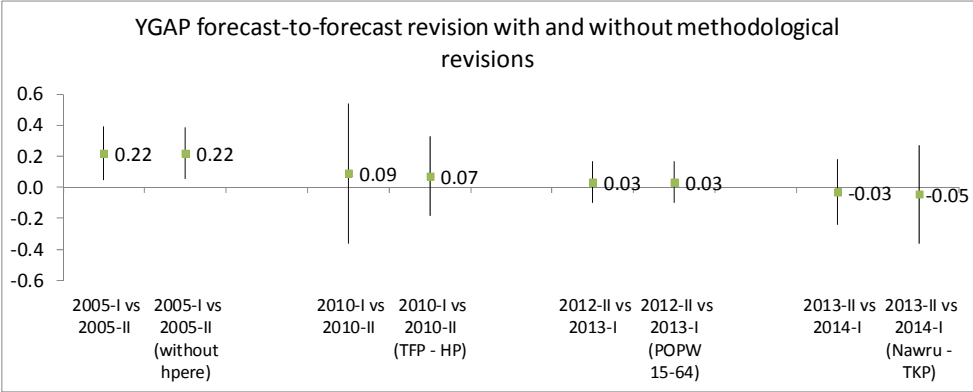
These conflicting results are further corroborated in graph 2 where the impact on changes in the overall Output Gap are shown by comparing the real time Output Gap for the year in which the change was introduced with the Output Gap estimate for those years taken from the Autumn 2014 vintage. If the KF estimation of TFP were not to have been introduced in Autumn 2010, the estimated Output Gap at that time would have been closer to the estimated Output Gap in Autumn 2014. However, the revision would again have increased, if in Autumn 2014 TFP still had been estimated using an HP filter. In the case of the NAWRU, graph 2 shows that if the NAWRU would have been calculated using the TKP for all EA12 members, the revisions compared to Autumn 2014 would have been larger.

Graph 3 : Output Gap estimates for EA12 - current year, current year without the methodological change and Autumn 2014 (2005, 2010, 2013 and 2014)
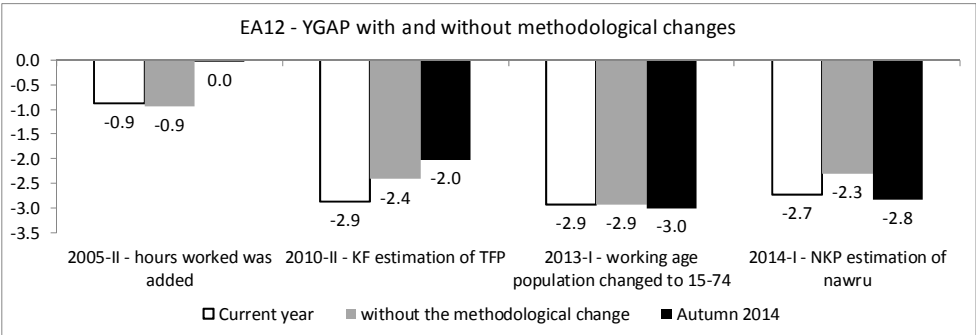
2 Impact on Output Gap Revisions from Small Methodological Changes (An assessment based on the Spring 2015 Forecast)

The previous section has shown the impact of large methodological changes on the Output Gap. Using the Spring 2015 forecast, it is possible to draw some conclusions with respect to the impact of small methodological changes on Output Gap revisions, especially since between Winter 2014-2015 and Spring 2015 no large methodological changes have occurred.

Table 1 provides a short summary of the overall revisions, in the Spring 2015 forecasting exercise, to the Output Gap, actual GDP & potential GDP forecasts for 2015 & 2016

(compared with those from Winter 2014 / 2015). It shows that the Spring 2015 Output Gap revisions for 2015 & 2016, compared with the previous Winter 2015 forecasts for those years, were generally quite small. The Output Gap forecast for the Euro Area only changed by .07 % points (2015) & .04 % points (2016). If one looks at the actual & potential GDP drivers of those small changes in Output Gaps, one observes similarly small revisions. Over the two forecast years, the average change in GDP was 0.07 in the Euro Area, with the average change in potential being about 60% smaller at 0.03.

As a general rule, changes to Output Gaps which are driven by actual GDP growth rates being higher than those of potential are not problematic. If however the change in Output Gaps is being driven by a change in potential output which is higher than that of actual output, an additional explanation is warranted to ensure that the PF methodology is doing its job correctly & is not playing an excessive role in driving the Output Gap revision. In this respect, the best criterion for establishing that the method is working correctly is to compare the extent to which the change in average potential growth rates over a number of years are greater than the equivalent average change in actual GDP. Using an average, rather than a single year, criterion is important since we do not want to pick up temporary variations in actual GDP forecasts which are then subsequently revised in the opposite direction at the next forecasting exercise, thereby ensuring that the average GDP forecast for the period as a whole does not change. We are more interested in ensuring that the PF method is reacting correctly to the change in average actual GDP (i.e. trend actual GDP) over the forecast period as a whole, which for the Spring 2015 forecasting exercise would be the period 2014-2016.

Table 1 : Spring 2015 Output Gap, actual & potential GDP growth rate differences with the Winter 2015 Forecast Round

	Spring 2015 Output Gap, difference with Winter			Spring 2015 GDP growth, difference with Winter 2015				Spring 2015 Potential GDP growth, difference with Winter 2015			
	2014	2015	2016	2014	2015	2016	Avg 2014-2016	2014	2015	2016	Avg 2014-2016
BE	-0.1	-0.1	0.0	0.0	-0.1	0.1	**0.0**	0.1	0.0	0.0	**0.0**
DE	-0.1	0.2	0.2	0.1	0.4	0.0	**0.2**	0.1	0.1	0.0	**0.1**
EL	0.9	-0.3	-0.5	-0.2	-1.9	-0.7	**-0.9**	-0.3	-0.6	-0.5	**-0.5**
ES	-0.3	0.1	0.1	-0.1	0.6	0.1	**0.2**	0.0	0.1	0.1	**0.0**
FR	0.0	0.1	-0.1	0.0	0.1	-0.1	**0.0**	0.0	0.0	0.0	**0.0**
IE	0.2	0.3	0.1	0.0	0.1	-0.1	**0.0**	-0.1	0.0	0.0	**0.0**
IT	0.1	0.0	0.1	0.1	0.0	0.1	**0.1**	-0.1	0.1	0.0	**0.0**
LU	-0.4	0.2	0.5	0.1	0.9	0.5	**0.5**	0.2	0.2	0.2	**0.2**
NL	0.0	-0.1	-0.1	0.2	0.2	0.0	**0.1**	0.2	0.2	0.1	**0.2**
AT	-0.1	-0.1	-0.2	0.1	0.0	0.0	**0.0**	0.0	0.0	0.0	**0.0**
PT	-0.2	-0.1	0.0	-0.1	0.0	0.1	**0.0**	-0.2	-0.1	0.0	**-0.1**
FI	0.2	-0.1	-0.3	-0.1	-0.5	-0.4	**-0.3**	-0.1	-0.2	-0.2	**-0.1**
EA12	-0.04	0.07	0.04	0.05	0.18	-0.02	**0.07**	0.02	0.07	0.01	**0.03**

What is the impact of "Small Methodological Changes" on Revisions : Tables 2 & 3 provide a quantification of the extent to which the "small changes"[15] made to the PF's methodology by the Commission services during the Spring 2015 forecasting exercise have driven the Output Gap revisions. These tables make a detailed assessment of the sources of

---

[15] These changes include changes necessitated by revisions to the historical & short term forecast data input series, and refer mainly to changes in the priors and bounds for TFP and NAWRU as well as to changes in the ARMA specifications for the extension of the hours worked, participation rate and investment to potential output series.

the Output Gap changes in 2015 & 2016, assessing to what extent the Output Gap changes in the forecast years are being driven by:

- new time series information (i.e. the changes to the historical data series & to the desk's short term forecasts) which has become available since the Winter forecasting exercise;
- large changes to the PF methodology; or
- the series of "small methodological changes" which the Commission services make to accommodate the changes in the underlying data input series (i.e. either the historical or forecast changes), with those small methodological modifications being motivated by the need to minimise historical revisions & maximise goodness of fit[16].

Table 2 Methodological and non-methodological sources of revisions to the Output Gap in 2015 (2015 Spring)

	Revision to output gap for 2015	1. Historical and short term forecast data revisions	2. Large changes to the PF methodology	3. Small methodological changes			
				Total	a. NAWRU	b. TFP	c. Other
AT	-0.10	-0.10	0.00	0.00	0.00	0.00	0.00
BE	-0.09	-0.09	0.00	0.00	0.00	0.00	0.00
DE	0.18	0.18	0.00	0.00	0.00	0.00	0.00
EL	-0.30	-0.30	0.00	0.00	0.00	0.00	0.00
ES	0.12	0.13	0.00	-0.02	0.00	0.00	-0.02
FI	-0.12	-0.12	0.00	0.00	0.00	0.00	0.00
FR	0.06	0.07	0.00	-0.01	0.00	0.00	-0.01
IE	0.27	0.27	0.00	0.00	0.00	0.00	0.00
IT	0.01	0.02	0.00	-0.01	-0.01	0.00	0.00
LU	0.23	0.23	0.00	0.00	0.00	0.00	0.00
NL	-0.06	-0.07	0.00	0.01	0.00	0.00	0.01
PT	-0.05	-1.18	0.00	1.13	1.12	0.00	0.00

---

[16] There are a number of principles / criteria which underlie the, stability-oriented, revisions policy which is used during the Commission services' Winter, Spring & Autumn forecasting exercises. These principles / criteria are applied when estimating the NAWRU, trend TFP, as well as the investment to potential output ratio, trend participation rate & trend hours worked series. There is a clear desire to ensure methodological stability, with small methodological revisions only been introduced if fully justified by changes in the underlying data input series (i.e. historical data series or forecast revisions). The methodology for estimating the NAWRU implies a NAWRU model that complies with three criteria: (1) minimizing historical changes in the NAWRU (unless due to consequential changes in the historical data), (2) a highly significant beta coefficient (measuring the effect of the unemployment gap on the wage indicator in the Phillips curve) and (3) a good Phillips curve fit. Priors on the parameters in the TFP model are only adjusted in order to decrease the size of the revisions in TFP trend growth rates relative to the previous forecast round or to improve the fit of the model. In assessing the latter, attention is in particular paid to the significance of the parameter beta, measuring the strength of the relation between the TFP cycle and the indicator of capacity utilization, and the fit of the TFP cycle. Finally, when deciding on which AR specifications to use for exogenously forecasting the participation rate, hours worked and investment to potential output series, the objective again is to limit historical revisions. Therefore it was decided not to use an automatic selection of the best specification but to look at individual countries separately. In finding the best model / specification for each of the individual Member States, consideration is given to the model fit (adjusted R²) and the significance of the coefficients.

Table 3 Methodological and non-methodological sources of revisions to the Output Gap in 2016 (2015 Spring)

| | Revision to output gap for 2016 | What is driving the 2016 output gap revision? | | | | | |
| | | 1. Historical and short term forecast data | 2. Large changes to the PF methodology | 3. Small methodological changes | | | |
				Total	a. NAWRU	b. TFP	c. Other
AT	-0.16	-0.16	0.00	0.00	0.00	0.00	0.00
BE	-0.03	-0.03	0.00	0.00	0.00	0.00	0.00
DE	0.17	0.17	0.00	0.00	0.00	0.00	0.00
EL	-0.48	-0.48	0.00	0.00	0.00	0.00	0.00
ES	0.11	0.16	0.00	-0.05	0.00	0.00	-0.05
FI	-0.35	-0.35	0.00	0.00	0.00	0.00	0.00
FR	-0.07	-0.07	0.00	0.00	0.00	0.00	0.00
IE	0.15	0.15	0.00	0.00	0.00	0.00	0.00
IT	0.11	0.12	0.00	-0.01	-0.01	0.00	0.00
LU	0.55	0.56	0.00	-0.01	0.00	0.00	-0.01
NL	-0.10	-0.12	0.00	0.02	0.00	0.00	0.02
PT	0.01	-1.27	0.00	1.27	1.27	0.00	0.00

Based on the evidence presented in tables 2 and 3, we can conclude that:

- Firstly, the only real driver of the Output Gap changes in the Spring 2015 forecasts are changes to the data input series (i.e. the new forecast information which has emerged since the Winter forecast exercise and historical data revisions) rather than methodological changes.

- Secondly, since there have been no "large" methodological changes to the PF method, the contribution to Output Gap revisions in this forecast round was zero.

- Finally, with respect to the "small" methodological changes made by the Commission services, for 11 out of 12 member states, the impact of these changes in parameter settings, bounds, priors and ARMA specifications is only noticeable at the second decimal place, the exception being Portugal. In the case of Portugal, the NAWRU parameters had to be changed to better fit the new data and to minimise historical revisions in the NAWRU.

- Tables 2 & 3 furthermore provide a three-way breakdown of the impact of these "small changes" on the output gap into the impact from changes to the NAWRU, trend TFP & other specifications (with this "other" category summarising the effects of changes to hours worked, investment & participation rate specifications). Since neither the priors nor any other parameters had been changed in the trend TFP estimations, the impact on revisions is zero. The NAWRU bounds and priors had only been changed for PT and IT, leading to very small revisions in the output gap for IT and somewhat larger revisions for PT. In the case of PT, keeping the parameters at their Winter 2014-2015 levels would have led to historical and current changes in the NAWRU in every year across the sample horizon varying from -2 to up to 1.5 percentage points. This would have led to much larger output gap revisions than what we see now. The other changes contribute to the revisions in Spain, Luxembourg and the Netherlands.

Overall, we can conclude that the widely held belief that methodological changes are a key driver of output gap revisions is not supported by the evidence presented[17]. By far the most significant driver of output gap revisions over time are historical data revisions & forecasting errors[18]. The impact of the four significant methodological changes made to the PF method since its inception in 2002, have been essentially neutral. If anything, the introduction of the NAWRU NKP methodology has decreased the revisions from its introduction, while the change towards a KF TFP calculation seems to have led to smaller revisions over the whole post-crisis period. Secondly, with respect to "small methodological changes" made by the Commission services in carrying out the potential growth rate calculations, it is clear that these changes were not a substantive factor in driving the output gap revisions in the Spring 2015 forecasts. The only real driver of those revisions were data input changes to both the historical, national accounts, time series as well as changes to the forecasts made by desk officers compared with the Winter 2015 exercise.

---

[17] At the Sept 2015 Joint EPC-ECFIN-JRC Workshop on "Assessment of the real time reliability of different output gap calculation methods", D. Turner (OECD) stressed that whilst GDP data revisions & methodological changes have contributed to real time revisions, the large & systematic revisions around the crisis have been driven by end-point problems (with the latter being the key issue to focus on). G. Kempkes (Bundesbank) replied that he was not convinced by the OECD suggestion that output gap revisions are mainly associated with end-point filtering problems. The main source of output gap revisions were, in his opinion, caused by revisions to actual GDP forecasts (he mentioned that this viewpoint was supported by the fact that the revision results from the EU's PF & HP filter methods were actually quite similar, indicating that methodological drivers of revisions were secondary).

[18] Revisions to the level of actual GDP (due to forecast errors or historical data revisions) are a bigger source of revisions in the level of potential output than changes to the production function methodology. Policy makers also need to accept that once a real time assessment is linked to forecasts (from either the Commission or the national authorities) revisions are inevitable due to the natural uncertainty surrounding all forecasting exercises.

+ Why HP100 is preferred to HP6.25 for Fiscal Policy Surveillance Purposes

In April 2014, the Bundesbank produced a study entitled "On the reliability of international organisations' estimates of the output gap", with the objective of assessing the reliability of the output gap estimates produced by international organisations such as the OECD & IMF, compared with the estimates produced by simple statistical procedures such as the HP filter (with a smoothing parameter of 6.25 rather than 100 which is the preferred parameter value for use in fiscal surveillance)[19]. The key results from the Bundesbank's analysis are given in Table 1, with the main conclusions stressed in the Bundesbank paper being the following :

- Firstly, all output gap estimates are associated with a high degree of uncertainty since an economy's potential output is unknown. In the past, simple as well as more complex statistical procedures have provided very unreliable output gap estimates, with later revisions often being as large as the previously identified gap itself.

- Secondly, as shown in the results given in Table 1 (showing the dispersion of the output gap estimates for each of the G7 countries, published in the IMF's & the OECD's regular publications since the spring of 1999 onwards, compared with appropriate real-time output gap estimates derived from a HP filter), the IMF & OECD output gap estimates are significantly less reliable compared with those of the HP filter. Not only are revisions substantially larger but also the IMF & OECD output gaps change their sign much more frequently than those of the HP filter.

- Thirdly, the Bundesbank analysis also suggests that IMF & OECD estimates of the output gap for the years just ended are initially frequently too unfavourable (driven by an excessively favourable view of the associated potential output growth rates). In subsequent rounds, these output gap estimates are generally revised up (& potential down). In overall terms the Bundesbank paper concludes that the potential growth rates of the G7 economies were probably, considerably, overestimated during the boom at the turn of the millennium and in the years immediately preceding the global financial crisis, suggesting a systematic optimistic bias.

- Fourthly, the paper stresses 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.

- Finally, the Bundesbank paper concludes that given the size of the IMF & OECD[20] output gap revisions, such estimates have to be treated with caution in sensitive

---

[19] The Bundesbank analysis employs a smoothing parameter of 6.25 for the HP filter– as shown later in this Annex, such a low smoothing parameter leads to a large degree of pro-cyclicality in the estimates – consequently for fiscal surveillance purposes, a smoothing parameter of 100 is generally used.

[20] A 2012 Bank of England paper analyzed whether the accuracy of the real time estimates of the output gaps produced by the OECD's estimation methodology had improved over time. The paper hypothesized that improvements should be expected given the availability of enhanced computing power for incorporating a greater amount of contemporaneous information, as well as the fact that new business surveys on spare capacity had become more widely available which also could be

economic policy areas such as in analysing cyclically adjusted fiscal deficits or in a monetary policy context.

Table 1 : Assessment of the real time reliability of OECD, IMF & HP Filter 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 Gap Estimates (Maximum value for a year over the different forecast vintages less the minimum value)			Number of years in which the sign of the Output Gap changed		
		IMF	OECD	HP Filter	IMF	OECD	HP Filter
United States	1980-1997	4.0	1.8	0.3	13	7	2
	1998-2010	2.9	2.4	1.0	7	7	3
Japan	1980-1997	2.4	2.3	0.7	4	8	2
	1998-2010	1.9	2.5	1.2	0	4	6
Germany	1980-1997	3.6	3.1	0.3	7	8	1
	1998-2010	1.9	1.9	1.1	5	4	4
France	1980-1997	1.6	2.2	0.5	5	5	3
	1998-2010	2.6	2.7	0.9	11	8	5
UK	1980-1997	2.6	2.5	0.5	7	3	2
	1998-2010	2.3	2.2	1.0	9	7	8
Italy	1980-1997	4.2	2.0	0.4	10	5	2
	1998-2010	3.1	3.3	1.0	10	11	6
Canada	1980-1997	10.9	1.9	0.3	16	4	1
	1998-2010	2.1	2.0	1.0	8	5	7

Why HP100 is preferred to HP6.25 for fiscal policy surveillance purposes ?[21]

Graph 1 shows the level of Euro Area GDP from 1981 to 2016, with the forecast years taken from the Commission services Autumn 2014 forecasting exercise. The graph shows clearly that the HP6.25 trend GDP line is much more pro-cyclical compared with the HP100 trend line i.e. the HP6.25 trend line more closely tracks the actual GDP line compared with the HP100 trend line[22]. Consequently, **when one assesses the real time reliability of the HP6.25**

---

exploited. Unfortunately, the hypothesis was rejected, with the analysis indicating, rather surprisingly, that the size of the OECD's output gap revisions had actually increased, not decreased, over time.

[21] When the HP filter was used officially for EU budgetary surveillance purposes in the period up to 2002, it was agreed that an inverse signal to noise ratio of 100 (similar to Backus and Kehoe (1992)) would be applied since fiscal policy makers were more interested in having smoother, medium term oriented, cycles (Note : this was not the case however for monetary policy institutions, which tended to use a smaller inverse signal to noise ratio. A smaller ratio is more appropriate for these institutions since less smooth series, covering shorter periods of time, are more suitable as an information input for short term oriented, monetary policy decision making purposes). One of the key drivers for moving away from the HP filter was the hoped-for gains regarding a reduction in pro-cyclicality. Policy makers called for a new method to be developed for assessing structural budget balances since it was felt that past surveillance exercises, based on the HP filter, had on a number of occasions produced an excessively optimistic picture of the degree of budgetary improvement in the upswing phase of previous cycles. This optimism was linked, to some extent, to the cyclicality of the trend GDP estimates which had been calculated using the HP filter statistical method and via which the estimates of structural budget balances had been generated. Consequently one of the key objectives laid out for the PF methodology was to reduce the degree of cyclicality of the potential growth estimates to an absolute minimum in order to avoid the mistakes of the past.

[22] As stressed in IMF (2013) regarding the smoothness properties of potential GDP "Central banks, whose primary purpose is the prevention of inflationary tensions, will tend to favour shorter horizons & relatively flexible potential outputs. By

method, it is not surprising to find that the size of the output gap revisions are much smaller compared with those of the HP100 method since the absolute size of the HP6.25 gaps are so much smaller. The corollary of this however is that the trend GDP revisions are much higher for HP6.25 compared with HP100. Consequently one must be careful in making judgements about the relative merits of different output gap estimation methods – whilst real time reliability is clearly important, there are other considerations to be borne in mind namely the absolute size of the gaps and whether the estimates display pronounced pro-cyclical features.

Graph 1 : Absolute difference between Output and Trend Output in levels (EuroArea, Autumn 2014)
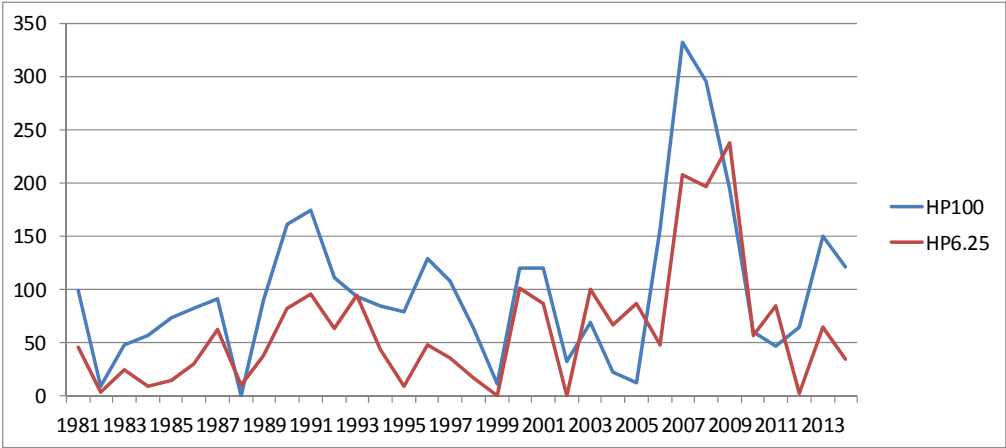
Note: the smaller the difference, the closer Trend follows Actual and the more the problem of procyclicality occurs

---

contrast, if the main focus is to target structural deficit targets, the IMF would give preference to a longer time horizon & a low volatility trend".

REFERENCES

Backus, D., & Kehoe, P. (1992), "International evidence on the historical properties of business cycles", *The American Economic Review, 864-888.*

Benes J, Clinton K, Garcia-Saltos R, Johnson M, Laxton D, Manchev P, & Matheson T (2010), "Estimating potential output with a multivariate filter", *IMF Working Paper 10 / 285.*

Ball, L.M. (2014), "Long term damage from the great recession in OECD countries", *NBER Working Paper 20185.*

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.*

Chiu, A., & Wieldek, T (2012), "Did output gap measurement improve over time ?", *Bank of England Discussion Paper No.36.*

Cohen-Setton, J (2013), "Blogs review : the structural balance controversy", *Bruegel Blog (October).*

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).*

Darvas, Z (2015), "Filling the gap : open economy considerations for more reliable potential output estimates", Bruegel Paper (Nov).

Denis, C., McMorrow, K., & Röger, W. (2002), "Production function approach to calculating potential growth and output gaps – estimates for the EU Member States and the US". *Directorate-General for Economic and Financial Affairs. Brussels: European Commission.*

Denis, C., Grenouilleau, D., McMorrow, K., & Röger, W. (2006), "Calculating potential growth rates and output gaps - a revised production function approach". *Directorate-General for Economic and Financial Affairs. Brussels: European Commission.*

Dolls, M., Fuest, C., Kock, J., Peichl, A., Wehrhofer, N., & Wittneben, C, (2014), "Automatic stabilizers in the Eurozone : analysis of their effectiveness at the Member State level & Euro area level & in international comparison", *ZEW 5/14.*

ECB (2000), "Potential output & output gaps : concept, uses & estimates", *ECB Monthly Bulletin (October).*

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

European Commission (2013), "Building a strengthened fiscal framework in the European Union : A guide to the Stability & Growth Pact", *Occasional Papers 150.*

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

Fioramanti, M, Padrini, F & Pollastri, C (2015), "Estimation of potential output & output gaps : an analysis of some critical issues", *UPB Italy.*

Gerlach, P (2011), "The global output gap : measurement issues & regional disparities"; *BIS Quarterly Review (June).*

Havik, K., Mc Morrow, K., Orlandi, F., Planas, C., Raciborski, R., Röger, W., Rossi, A., Thum-Thysen, A. & Vandermeulen, V. (2014), "The production function methodology for calculating potential growth rates and output gaps". *Directorate-General for Economic and Financial Affairs. Brussels: European Commission.*

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

Hodrick, R.J. & Prescott, E. C (1997), "Postwar US business cycles : an empirical investigation", *Journal of Money, Credit & Banking 29: 1-16.*

IMF (2013), "France – Article IV Report", IMF Country Report No13/251.

IMF (2015), "Ireland – Selected Issues", IMF Country Report No15/78.

IMF (2015), "France, Germany, Italy & Spain – Selected Issues", (July).

Kempkes, G (2012), "Cyclical adjustment in fiscal rules : some evidence on real-time bias for EU-15 countries", *Discussion Paper Deutsche Bundesbank, No 15.*

Marcellino, M & Musso, A (2011), "The reliability of real-time estimates of the Euro area output gap", Economic Modelling 28, 1842-1856.

Mc Morrow, K, Roeger, W, Vandermeulen, V & Havik, K (2015), "An assessment of the real time reliability of the Output Gap estimates produced by the EU's Production Function Methodology", Paper presented at the Joint EPC-ECFIN-JRC Workshop "Assessment of the real time reliability of different output gap calculation methods" (28 Sept 2015)

Minsky, H (1986), "Stabilizing an unstable economy", *Yale University Press*.

Mourre, G, Astarita, C & Princen, S (2014), "Adjusting the budget balance for the business cycle : the EU methodology", *European Economy Economic Papers 536.*

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 (1999), "The reliability of output gap estimates in real time", *Board of Governors of the Federal Reserve System, Finance & Economics Discussion Series 99/38.*

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***.**

Proietti, T, Musso, A & Westermann, T (2007), "Estimating potential output & the output gap for the Euro Area : A model based production function approach", *Empirical Economics, 33, 85-113.*

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

Ravn, M., & Uhlig, H. (2002), "On adjusting the Hodrick-Prescott filter for the frequency of observations". *Review of Economics and Statistics, 84(2), 371-376.*

Saint-Amant, P & van Norden, S (1997), "Measure of the output gap : A discussion of recent research at the Bank of Canada", *Technical Report No. 79.*

Scott, A (2000), "Stylised facts from output gap measures", *Reserve Bank of New Zealand, Discussion Paper Series.*

Slevin, G (2001), "Potential output & the output gap in Ireland", *Central Bank of Ireland Technical Paper.*

Summers, L (2014), "US economic prospects : Secular stagnation, Hysteresis & the Zero Lower Bound", *Business Economics, Vol 49, No 2.*

Summers, L (2014), "Reflections on the new secular stagnation hypothesis", *VOX (October).*

Watson, M (2007), "How accurate are real-time estimates of output trends & gaps", *Federal Reserve Bank of Richmond Economic Quarterly (Spring).*

EUROPEAN ECONOMY DISCUSSION PAPERS

European Economy Discussion Papers can be accessed and downloaded free of charge from the following address:
http://ec.europa.eu/economy_finance/publications/eedp/index_en.htm

Titles published before July 2015 under the Economic Papers series can be accessed and downloaded free of charge from:
http://ec.europa.eu/economy_finance/publications/economic_paper/index_en.htm

Alternatively, hard copies may be ordered via the "Print-on-demand" service offered by the EU Bookshop: http://bookshop.europa.eu.

HOW TO OBTAIN EU PUBLICATIONS

Free publications:
- one copy:
  via EU Bookshop (http://bookshop.europa.eu);

- more than one copy or posters/maps:
  - from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  - from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  - by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
    calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*)   The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:
- via EU Bookshop (http://bookshop.europa.eu).