Eesti Pank
Bank of Estonia

Can We Rely on Real Time Figures for Cyclically Adjusted Budget Balances?

Andrew Hughes Hallett, Rasmus Kattai and John Lewis

Working Paper Series
2/2009

Can We Rely on Real Time Figures for Cyclically Adjusted Budget Balances?*

Andrew Hughes Hallett[†], Rasmus Kattai[‡] and John Lewis[§]

Abstract

This paper analyses the reliability of real time estimates of cyclically adjust budget balances (CABs). We find that real time CABs are not better at forecasting the *ex post* figures than simpler benchmarks. Further, we find that real time CABs have low power in detecting fiscal slippages, and in correctly identifying fiscal improvements. Around half of the real time errors in CABs can be attributed to revisions in the cyclical component of the budget balance, and around one half to revisions in the deficit to GDP ratio across vintages. That means it will be difficult to use them to reliably monitor the health of public finances. Lastly, we find that CABs are systematically less reliable under conditions of poor or deteriorating public finances, which means they are at their most unreliable precisely when they are needed most.

JEL Code: H62, H87

Keywords: Real Time Data, Cylically Adjusted Budget Deficits, Measurement Error

Authors' e-mail addresses: ahughesh@gmu.edu, rkattai@epbe.ee and j.m.lewis@dnb.nl

The views expressed are those of the authors and do not necessarily represent the official views of De Nederlandsche Bank or Eesti Pank.

*The authors are grateful to the DNB's visiting scholar programme for enabling Rasmus Kattai to visit the DNB to work on this paper. This paper would not have been possible without the painstaking work of others in compiling the real time dataset. We thank Massimo Giuliodori for allowing us to use the dataset he assembled, and Peter Keus of De Nederlandsche Bank for his work in collating additional data. We thank seminar participants at the Universities of Amsterdam and Cape Town, the Estonian Economic Association, the CIRANO Workshop on Data revision, the KOF Workshop on data revision and at the Nederlandsche Bank for useful comments.

†George Mason University, School of Public Policy, 4400 University Drive, Fairfax, VA 22030, US.

‡Research Department, Bank of Estonia, Estonia Pst 15, 15095, Tallinn, Estonia.

§Economics and Research Division, De Nederlandsche Bank, Postbus 98, 1000AB, Amsterdam, The Netherlands.

Non-technical summary

Cyclically adjusted budget balances (CABs) are commonly used as an early warning indicator of fiscal loosening and its role has gradually increased with the reforms to the Stability and Growth Pact (SGP). CABs are useful for monitoring the underlying fiscal stance because they enable to strip out fiscal slippage also at the top of the business cycle, i.e. while the actual budget balance may still be in surplus because of extra revenues generated by good economic conditions. In order to perform as a "preventative arm" in avoiding major fiscal problems, as stated in the SGP, CABs have to be calculated based on the data available *at the time*.

Using the extensive real time dataset on the nineteen OECD countries we show in the paper that the real time CAB figures are very imprecise and perform extremely poorly in identifying fiscal slippages. It is mainly related to difficulties in measuring the current cyclical position of the economy correctly. The resulting problem is twofold — there are episodes of fiscal loosening that can only be verified *ex post* and not at the time; but there are also episodes of real time slippages, although it is not apparent when the *ex post* data is used. These wrong "alarms" are more likely to emerge from the real time data than "true" alarms. It brings us to a conclusion that CABs should be used with a great caution by policymakers.

Contents

1. Introduction

Cyclically adjusted budget balances (CABs) are routinely used in the analysis of fiscal policy. The basic rationale is well known — namely that by stripping out the cyclical effects of automatic stabilisers, the remaining adjusted balance is a truer picture of the underlying state of public finances.

However, the reported CAB for a given year may be subject to quite large revisions over time. Thus the figures available at the time the policy was being made may differ quite substantially from those published some years later in the official statistics. This is due not only to real time measurement error in the deficit itself, but also those in the cyclical adjustment process, which typically rely on a output gap measure (or other cyclical indicator) which may be known very imprecisely in real time (see Orphanides and Van Norden, 2002). The goal of this paper is to investigate the reliability of real time figures for cyclically adjusted budget balances. Our results are of importance for two of the most common uses of CABs.[1]

The first is the use of the CAB as a structural indicator of the sustainability of public finances. Cyclically adjusted figures are often considered an "early warning" device. During an upswing, a loosening of fiscal policy may not show up in the headline budget deficit figure, because strong growth boosts tax revenues. However, detecting deteriorations in the structural balance early enough may permit fiscal consolidations in the upswing of the cycle, and hence facilitate corrective measures which would be politically (and economically) more costly to achieve in the downswing when the actual budget deficit naturally increases (Buti et al., 2003b; Hughes Hallett et al., 2004). As a specific example, within Europe the "close to balance or in surplus" criterion of the Stability and Growth Pact (SGP) be explicitly assessed in terms of cyclically adjusted budget figures. Yet the effectiveness of the CAB as a timely indicator of fiscal health depends crucially on the accuracy of CAB figures available to policymakers *at the time*, rather than the *ex post* data published many years after the event.[2]

The second, is the use of the CAB as an indicator of discretionary fiscal policy. A substantial number of empirical papers treat the government's fiscal policy instrument as the cyclically adjusted (primary) balance, rather than the total balance. Since the operation of automatic stabilisers occurs in the absence of any discretionary interventions by the government, so the argument goes,

[1]See Blanchard (1990) for a comprehensive overview of different uses of CABs.

[2]This point was well illustrated by deterioration of public finances in the Netherlands during the early part of this decade. The European Commission's 2004 spring forecast recorded a cyclically adjusted deficit of 2.8% of GDP for 2002. However, the same forecast in Autumn 2001 (after the budget was passed) estimated the 2002 CAB to be in surplus by 0.8%.

these automatic effects need to be stripped out in order to assess the actions of the policymaker. However, most of the empirical work on this topic uses revised data which was not available to the policymaker at the time.[3] In such studies, the cyclically adjusted balance is usually taken to represent what the policymaker was trying to do at that time. If the real time figures are quite different from the revised data, then the latter may be a poor approximation of the fiscal stance fiscal policymakers thought they had adopted. For example, episodes which are classifed on basis of *ex post* data as fiscal loosenings, may have looked quite different to the policymaker at the time. Thus, the implicit assumption that the policymaker was deliberately loosening fiscal policy may not hold — the worsening of public finances may only have become apparent several years after the event.

The issue of cyclically adjusted budget balances is likely to grow in importance in the coming years. Many OECD governments face the need to consolidate public finances in the years ahead, in the wake of strongly expansionary stimulus packages. In general, simply relying on improved tax revenues from a pick-up in economic activity may not be enough to place public finances on a secure footing — additional discretionary measures will be needed. Clearly, quantifying the relative contributions of automatic stabilisers and of specific discretionary tightening measures will be an important task in determining the soundness of public finances.

This paper investigates how big the difference is between real time and *ex post* estimates of the CAB in three different ways. First, we see how the OECD's real time CAB figures match up against a simpler methodology. The latter is based on applying an HP filter to calculate our own output gap figures and using these to calculate our own CAB estimates (Section 2). Surprisingly we find that the simple method outperforms the accuracy of OECD's real time figures at "forecasting" the OECD's final figures. Second, we analyse how accurate the real time CABs are at identifying slippages in fiscal policy, and at detecting improvements (Section 3). We find that real time data is quite unreliable, and that a fiscal surveillance system based on this variable alone will result in large numbers fiscal slippages going undetected; and, in many instances, of fiscal policy being wrongly characterised as a slippage.

Third, we analyse whether there is a systematic pattern to the measurement errors associated with real time CABs (Section 4). We find that CABs are likely to be more unreliable, the worse the "true" state of public finances is (proxied by *ex post* data), and also in times when the "true" fiscal position is slipping. Thus indicating that CABs are systematically more inaccurate just at the time when they are needed most as an early warning.

[3] See for example, Galí and Perotti (2003), IMF (2008), Breuss and Roger (2005), Buti et al. (2003a).

There is an emerging literature on the use of real time data to evaluate fiscal policymakers behaviour. Forni and Momigliano (2004), Golinelli and Momigliano (2008), Cimadomo (2007) and Bernoth et al. (2008) all find that fiscal policy reaction functions using real time data show counter cyclical responses which do not show up when the same estimation is carried out with *ex post* data. On the fiscal monitoring side, Jonung and Larch (2006) investigate the role of errors in potential GDP forecasting, and find that for some countries, real time assessments of fiscal position are over optimistic due to a systematic upward bias in government produced forecasts of potential output. Using data on GDP revisions to analyse cyclically adjusted budget balances, González-Mínguez et al. (2003) show that revising previous values of GDP can generate the illusion of a change in government behaviour by altering the CAB, even though the actual deficit remains unchanged.

However, there is no systematic attempt to quantify the size of the error arising from the problem of inferring a CAB in real time, or how those errors may affect fiscal oversight itself. On the other hand, several authors have highlighted the importance of deliberate misreporting of headline deficit figures (see for example Koen and van den Noord, 2005; von Hagen and Wolff, 2006). Our focus here is rather different. First, we are not primarily concerned with deliberate data manipulations by governments as such but rather the total size of real time measurement errors in CAB figures. To this end, we make use of OECD data which, unlike European Commission data, is not directly supplied by the national governments themselves. As Beetsma and Guliodori (2008) note, this means that OECD data is less susceptible to strategic manipulation by national governments than that of other sources. Thus, our dataset is subject to measurement errors largely outside the scope of opportunistic accounting practices.

The information available to policymakers at the time may differ substantially from the information available after the event for a number of reasons. To begin with, contemperanoues data on deficits and GDP available to policymakers at the time are likely to be subject to many revisions in subsequent periods. Second, extra observations of output beyond time t can improve the estimation accuracy of the time t output gap. This is related to the so-called "endpoint problem", i.e. the more observations of output one has beyond year t, the better one can estimate potential output in year t itself. Third and more broadly, additional data may lead to revisions in the model of the economy, leading in turn to further changes in estimates of the output gap (Orphanides and Van Norden, 2002).

There are a number other factors which may generate a real time measurment error in the headline budget balance figures. The OECD (along with the IMF and the EU) uses an accrual rather than a cashflow basis for national accounts. This records transactions in the year in which the transaction itself occurred, rather than when the payment was actually transferred. Since many taxes are paid with a lag, the government recieves part of the year t tax revenue in year $t + 1$ or even $t + 2$. Similarly, some expenditures, although booked to year t may not be finalised until later years. Thus the at the year end, one must estimate the eventual expenditure and revenue flows which will eventually be booked to year t. In addition, there other data compilation lags — even if the expenditure has occurred within the same year as the transaction, the time needed to collate data may mean that by the year end, the exact figure is not known.

2. How well do the real time figures perform against a simple benchmark?

To quantify the magnitude of these errors, a real time dataset is used, compiled from successive issues of the OECD's *Economic Outlook* (EO). The dataset consists of the yearly published values of GDP, output gap and actual budget deficit series in each issue from December 1995 (Issue 58) to December 2008 (Issue 84), as well as the published values of cyclically adjusted budget balance. The compiled dataset includes 19 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, Norway, Portugal Spain, Sweden, United Kingdom and United States.[4] The time span for an individual time series is from 1985 to the year of publication.

We use OECD data rather than IMF or European Commission data because *Economic Outlook* has published cyclically adjusted budget balances since 1993, earlier than can be obtained from European Commission or IMF data. Using data from another source would mean shortening our already small sample.[5] The OECD has published its own real time database, the earliest vintage being 1999. We therefore we utilise our own (longer) dataset in order to take advantage of the larger number of observations.

Since the data is compiled by an independent body, it is reasonably in-

[4] The number of countries included in our dataset is determined by data availability. Countries which had missing values at least in one of the data categories, i.e. output, output gap, actual budget balance or CAB were left out. Those countries are the Czech Republic, Hungary, Iceland, Luxembourg, Mexico, Slovakia, South Korea and Turkey.

[5] The European Commission is yet to publish a real time data set of its own. The earliest vintage for cyclically adjusted budget balances we could find was from 1995, for 12 EU nations (European Economy 60, Broad Economic Policy Guidelines 1995), but then there is a gap until 1998.

sulated against "political" bias in provisional figures and forecasts compiled by national governments. For example, empirical work by Jonung and Larch (2006) suggests that, in some countries, estimates of potential output produced by national statistical agencies may have been biased systematically upwards, in order to present a more favourable picture of cyclically adjusted public finances.

Figure 1 depicts average CAB of 19 countries in our dataset and how it has changed over the time. It can be seen that data revisions have had minor effects on average CAB for certain years, 1996–1998 for example, whereas they are remarkably error prone for 2000 and onwards. As the average CAB figures to some extent cancels out revisions with opposite signs on a single country level, the imprecision of real time CAB estimates is even greater for a single country (elaborated in more detail later) and indicate clearly the difficulty of predicting the *ex post* data correctly in real time.

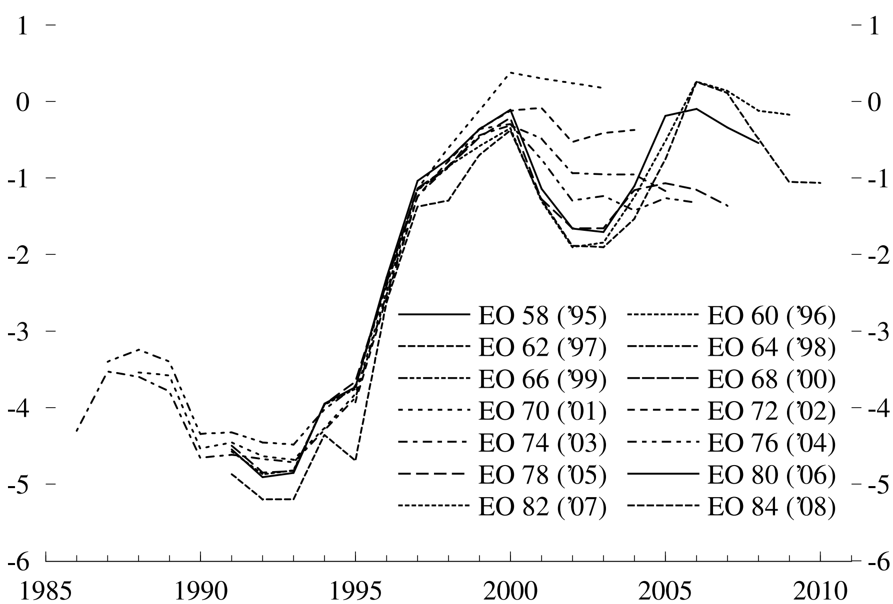

Figure 1: Average CAB Estimates of 19 OECD Countries Across Vintages 1995–2008 (% of GDP). Source: OECD *Economic Outlook* 58–84.

In what follows we compute our own real time CAB estimates and compare their power to predict the *ex post* realization with the one of the OECD's CAB estimates. Throughout this paper *ex post* data refers to the latest release of the *Economic Outlook*, which is the December 2008 issue.

Our simple benchmark is obtained as follows. First, we utilise real time GDP data to extract the output gap (although this is based on raw data taken from the OECD we refer to this as our "own estimate"). This measure of the cyclical position of the economy is then combined with the OECD's published budgetary elasticties and real time total budget balance data used to derive our own real time CAB estimates. The formal equation utilised in our calculations is the following:

$$\bar{b}_{t|t+s} = b_{t|t+s} - \varepsilon_t \tilde{y}_{t|t+s}, \tag{1}$$

in which $\bar{b}$ denotes CAB, b is the actual budget balance (both in percentages of GDP), ε stands for budget elasticity and $\tilde{y}$ denotes output gap. In our notation we use two time subscripts. The first, t, has a conventional interpretation — denoting the time period to which the observation refers. The second, expressed as $t+s$, refers to the vintage of data. For example, CAB for 2002 as reported in the 2006 edition of *Economic Outlook* would appear in our notation as $\bar{b}_{2002|2006}$, i.e. $t = 2002$ and $s = 4$.

Output gap estimates are obtained by applying the Hodrick-Prescott (HP) filter with the smoothing parameter set to 100 to estimate the trend output. This estimate is then subtracted from the actual real time GDP figure reported in the corresponding issue of EO.

A common problem related to the HP filter is the so called end-point problem, meaning that the filtered potential output time series tends to be biased toward the actual data in the beginning and at the end of the sample.[6] To mitigate the this, we add GDP forecasts for five years ahead before applying HP filter. Forecasts are produced with ARIMA models, which are automatically estimated by the algorithm built in the TRAMO program provided by EViews.[7]

We therefore get 14 vintages of output gap data for each country, which are based on the successive December issues of EO, that is December 1995 – December 2008 (similarly to the reported average CABs on Figure 1). We report the error in the real time output gap estimate with respect to the *ex post* realization. This is expressed in terms of the root mean squared errors (RMSE). The RMSE is calculated for 5 data vintages, formally referred to as $RMSE_{\tilde{y},s=i} = \left(\sum_t \left(\tilde{y}_{t|s=i} - \tilde{y}_{t|2008}\right)^2 / n\right)^{1/2}$, $i = \{0...4\}$. Thus, the earliest vintage $(s = 0)$ is the real time data, and subsequent vintages are those 1, 2, 3 and 4 years after time t.

[6]The issue is raised, for example, by Bjornland et al. (2005), Guay and St-Amant (1997) and St-Amant and van Norden (1997).

[7]Note that because we use real time GDP data to calculate this benchmark output gap, we respect the principle of using information which would have been available at the time.

RMSEs are based on a common sample length 1995–2004. 1995 is the earliest data vintage available, and given that our latest vintage is 2008, 2004 is the last year for which have data for $\tilde{y}_{t|s=4}$.

In practice, we calculate three sets of RMSEs — (a) we compare our own real time gap estimates with our own *ex post* data, (b) our own real time gap estimates are compared with the OECD's *ex post* data and (c) the OECD's real time data is compared with the OECD's *ex post* gap estimates.

The average RMSEs of all 19 countries are presented in Figure 2. First, it is visible that the size of the errors decrease in s. This is an expected result because the more information is available for the gap estimation, i.e. the further in the past lies the period of observation, the more accurate is the gap extraction. Second, our own gap estimates show the best results in terms of predictive accuracy when it is compared to our own *ex post* gap series. It is a bit worse in predicting OECD's *ex post* gap but surprisingly it still beats the accuracy of the OECD's own real time gap estimates (more detailed results at a country level are presented in Appendices 1 and 2).

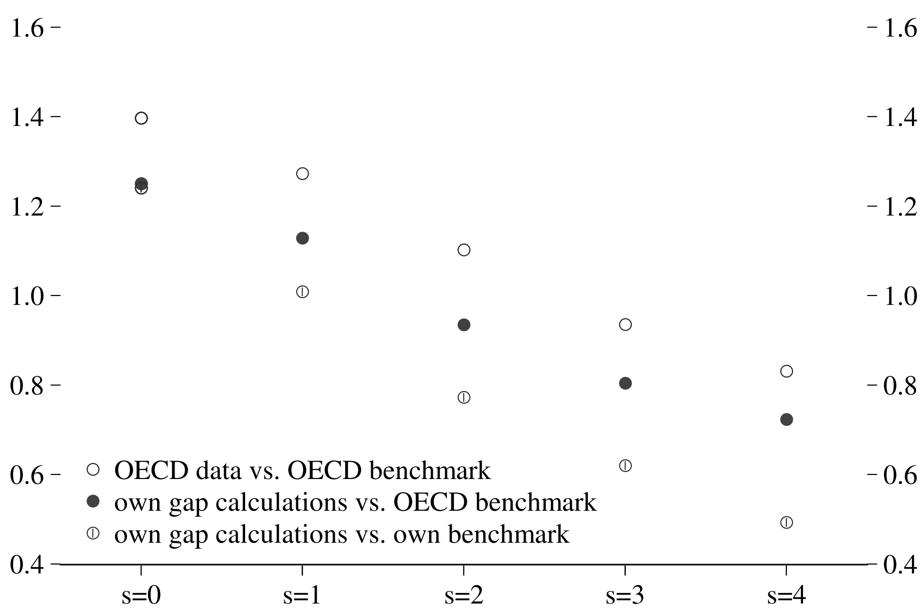

Figure 2: Revisions in Output Gap Estimates: Root Mean Squared Error. Source: OECD *Economic Outlook* 58–84, authors' own calculations.

Next we repeat the analysis for CAB figures. In this case, our own CAB figures are calculated by taking our own output gap figure, multiplying that by the OECD's estimate budget elasticity, then subtracting this figure from the corresponding OECD budget balance measure (see equation 1). This is a somewhat cruder methodology than the OECD uses, because the OECD also take into account potential lags in taxes, spending and also make adjustments for "one-off" effects on the primary balance. In doing this we make use of three sets of elasticities (this is why elasticity parameter ε has a time subscript in equation 1). Elasticities for 1985–1998 are taken from van den Noord (2000) and are based on 1998 weights; elasticities for 1999–2002 are based on 1999 weights (Ibid.); and elasticities for 2003–2008 are published by Girouard and André (2005) where they use 2003 weights.

The results are depicted in Figure 3 and they exhibit the same pattern as the gap estimates above.[8] We see our own real time CAB figures have the best forecasting power when they are compared to our own *ex post* data. They are moderately worse at predicting OECD's *ex post* CABs. However, they outperform the OECD's final data than the OECD's real time CABs do. (More detailed results at a country level are presented in Appendices 3 and 4).

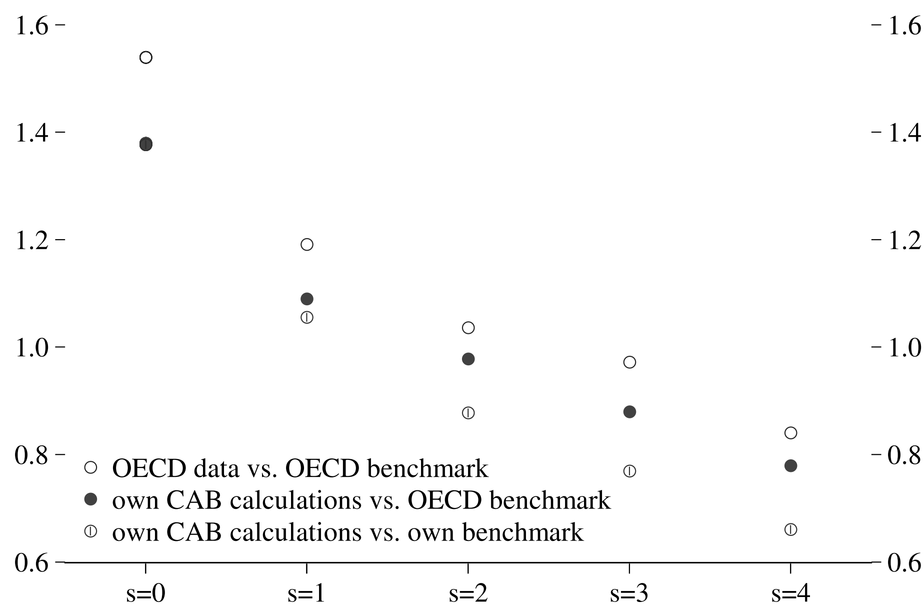

Figure 3: Revisions in CAB Estimates: Root Mean Squared Error. Source: OECD *Economic Outlook* 58–84, authors' own calculations.

Taken together, our analysis demonstrates that a simple HP filter approach combined with some pre-specified elasticities is does a better job of predicting the final OECD data than the OECD's own real time figures.

[8]Norway is excluded from the calculation of average RMSE because the OECD record oil revenues in the primary balance, but not in the CAB — our failure to do this in our own methodology generates a large RMSE.

3. The Effectiveness of CABs as an Early Warning Indicator

An important function of CABs is to serve as an early indicator of fiscal slippage, especially during the upper part of the business cycle when strong economic growth may mask the effect of fiscal loosening on the actual budget deficit.[9] To test the effectiveness of this warning[10], we construct two binary measures. First, we define a "fiscal slippage" to be a worsening in the CAB of $1.5pp$ over one year. This is similar to the definition adopted elsewhere in the literature (Blanchard et al., 2000; Hughes Hallett and Lewis, 2007). The second, we define a "fiscal loosening" to be, and that is defined a negative change in the CAB over one year. This has parallels in the European Commission's "significant improvement test" (discussed later). It can also be thought of as testing whether the real time data gives the correct sign for CAB changes.

If the change in the real time CAB exceeds some trigger value, then we assume a hypothetical "alarm" is sounded. If not, then no alarm is registered. Comparing these with the *ex post* data, we can then classify the CAB in one of four states — correct alarm, false alarm, missed alarm, correct all clear, depending on whether the alarm was correctly sounded or not (see Table 1).

Table 1: Classifying Budgetary Outcomes

		Real Time Data	
		CAB worsens more than $1.5pp$	CAB worsens less than $1.5pp$
Ex Post Data	CAB worsens more than $1.5pp$	Correct alarm	Missed alarm
	CAB worsens less than $1.5pp$	False alarm	Correct, all clear

The occurrence of a slippage and the early warning are both defined in terms of CABs. Thus, the issue is purely on of revisions in data across vin-

[9]More broadly, studies which analyse episodes of fiscal expansion or slippages, typically do so on the basis of *ex post* data (IMF, 2008; Buiter and Grafe, 2002; Hughes Hallett and Lewis, 2008) and do so in the belief that these slippages represented the deliberate choice of policymakers at the time. It is therefore instructive to see whether these slippages were indeed evident to policymakers at the time.

[10]Throughout this section we use the term "early warning" in a general sense. It does not refer to the much more specific concept of an "early warning" as laid out in the Stability and Growth Pact which EMU members are bound by.

tages. If there were no data revisions, and hence real time data was the same as *ex post* data, then the alarm would have a 100% success rate.

Using our own dataset, we obtain the results for the period 1995–2008. Table 2 shows that the CAB performs very poorly as an early warning indicator. In the baseline case (the first row, where the trigger is $-1.5pp$), only six out of sixteen fiscal slippages (*correct* plus *missed alarms*) are correctly picked out in real time. It could be that this result is sensitive to the trigger value and that a number of the missed slippages represent fiscal loosenings close to the $-1.5pp$ boundary. To investigate this we perform two robustness checks.[11]

Table 2: Success of the CAB as an Early Warning: Slippage of more than $1.5pp$ in one year

Trigger		Correct Alarm	False Alarm	Missed Alarm	Correct All Clear
	Frequency	6	3	10	209
$-1.5pp$	Average RT Change	-2.19	-2.12	-0.35	-0.01
	Average EP Change	-2.37	-0.01	-2.00	0.52
	Frequency	7	13	9	199
$-1.0pp$	Average RT Change	-2.03	-1.46	-0.28	-0.04
	Average EP Change	-2.26	-0.04	-2.04	0.58
	Frequency	12	34	4	178
$-0.5pp$	Average RT Change	-1.47	-0.96	0.239	-0.09
	Average EP Change	-2.26	-0.08	-1.78	0.68

Source: OECD *Economic Outlook* 58–84, authors' own calculations.

Firstly, we compute the average real time (RT) and *ex post* (EP) change in the CAB (expressed in percentage points). In the baseline case, the average change in the real time CAB for the "missed alarm case" is -0.35, quite far away from the trigger value. A small change in trigger value will not help to pick up these cases.

Second, we re-run the experiment for different trigger values (keeping the same definition of a slippage). Setting the trigger at -1.0 (fourth row) only picks up one extra slippage in real time. But this comes at a very high cost: 14 false alarms (that is eleven more than in the benchmark case). Notice also that average *ex post* change in the CAB is around zero. That means that not

[11]These results are robust to alternative defintions of fiscal slippage. In addition to definition of the next subsection, we also tried a definition based on the change in the CAB over two years (Hughes Hallett et al., 2007).

only are these governments (*ex post*) barely loosening fiscal policy, but a good number turn out *ex post* to have been tightening their fiscal policy. Setting the trigger at -0.5 (an even stricter measure) tells a similar story. In this case, 12 out of 16 slippages are picked up, but at the cost of 34 false alarms. That means that three out of every four alarms is a false one. The average value of the *ex* post change in the CAB is -0.1 (a long way away from the -1.5 percentage point definition of fiscal slippage) and hence these do not represent "near misses". Thus anybody attempting to issue cautions on the basis of real time figures is likely to have credibility problems.

One might take the position that a fiscal monitor might have a highly asymetric loss function and would thus be happy to tolerate a large number of false alarms, if that meant that all slippages would be picked out. However, in that case the monitor runs the risk of being seen "crying wolf". Knowing this, governments might reasonably chose to ignore an "early warning", arguing that there was a too high probability it would turn out to be false. More generally, a real time defintion which successfully captures all (or most) of the slippages also entails a very large number of wrongly identified slippages, which correspondingly weakens the power of the early warning in the first place.

More generally, this suggests caution in the interpretation of ex post data. The high number of missed alarms suggest that many fiscal slippages did not show up in real time, and hence may not have reflected a conscious desire to loosen fiscal policy. Simply labelling such episodes as fiscal slippages and assuming that they reflected deliberate policy measures may be misleading.

Can CABs Pick up Fiscal Consolidations in Real Time?

We now consider the power of real time CABs to pick up improvements in the fiscal position. For EMU countries, the European Commission now requires individual countries with excessive deficits to show a minimum improvement in their deficit figures as part of the national stability programmes, or face the threat of sanctions. These improvements have to be computed in terms of the CAB, and hence in real time, and the criterion is currently set at an improvement of at least $0.5pp$ of GDP each year. The logic underpinning this (Buti et al., 2003a; European Commission, 2002; Sapir, 2004), is that the (*ex post*) CAB should not deteriorate (i.e the change should be 0 or postive). To allow for possible real time measurement errors, a safety margin is added in. Thus the "real time" benchmark is that the (real time) CAB must improve by $0.5pp$. This test also has a more general relevance. Since the trigger value to denote a loosening is zero, this can also be interpreted as a test of whether we can correctly identify the sign of the change in the CAB in real time.[12] In

[12]Because we set the trigger to $0.5pp$, we are choosing a cautious value — the idea is that

what follows, we test the power of real time data to at least correctly indicate the sign of the change in the CAB.[13]

Given the recent deterioration in fiscal positions in most EU countries, a significant number of EU nations may be placed under the formal obligation to consolidate their public finances. Outside of the EU, many other OECD countries may also need to improve their structural budget position as they seek to withdraw stimulus measures. We analyse how accurate the data is at picking up these improvements in two samples. The first is the entire data sample (as in the previous table). For the second, we take only those countries whose budget deficit was more than 3% of GDP. For EMU countries this is the reference value at which excessive deficit procedure may be invoked. For other countries this is a simple way of restricting the sample to periods when deficits are relatively loose. The results for both samples are given in Table 3.

Table 3: Success of CAB at Detecting Improvements in Fiscal Position

Sample		Correct Alarm	False Alarm	Missed Alarm	Correct All Clear
	Frequency	73	59	13	83
Full	Average RT Change	−0.59	−0.85	0.79	0.55
	Average EP Change	−0.95	0.55	−0.84	1.3
	Frequency	11	10	12	23
Restricted	Average RT Change	−0.49	0.08	−0.32	0.36
	Average EP Change	−0.73	0.35	−0.92	1.24

Source: OECD *Economic Outlook* 58–84, authors' own calculations.

Looking at the full sample, Table 3 shows that in real time, the $0.5pp$ safety margin is insufficient to pick out all fiscal loosenings, although it does pick up most of them. However, it comes at a high cost in terms of "false alarms", almost 40% of alarms are incorrect.

The picture worsens considerably when we restrict our attention to countries whose (*ex post*) deficit exceeds 3%. Of the 23 instances of fiscal loosening in that group, only half are picked up in real time. Moreover, of those that are missed, the loosening actually turns out *ex post* to be even larger on average

tightenings will only be declared as such if there is significant evidence of them. This reflects the fact that one is usually more worried about failing to spot a deterioration in public finances, than misclassifying an *ex post* imrpovement. Setting the trigger to zero would lead to fewer false alarms, but more missed alarms. However, the results do not change substantially with respect to the trigger value.

[13] Note that we use OECD figures, rather than Commission figures on which the actual SGP decisions would be based.

than those which are picked up. That means that the $0.5pp$ safety margin is too small. Half of the countries who recorded an (*ex post*) fiscal loosening had shown a tightening of more than $0.5pp$ in real time. Equally worrying is that the fact that almost half of the alarms turn which are sounded turn out to be false.

For EMU members, the decision on whether to apply sanctions[14] is contingent on failure to improve by $0.5pp$. This data suggests that very often countries will be unfairly punished on the basis of real time data, whilst a number of "sinners" will go undetected, at least in real time.

How Much Better Are Later Data Vintages?

Table 4 extends the tests of Tables 2 and 3 using subsequent vintages of the CAB measure. Thus $s = 0$ denotes the real time case, and $s = 1, 2, 3$ refers to the outcomes when the tests are applied with CAB estimates made with the data available 1, 2 and 3 years later.[15] The results are shown in Table 4.

It is clear from Table 4 that although later vintages tend to perform somewhat better than real time data at picking out slippages, there are still substantial errors even three years after the event.

In the case of fiscal slippage, the $s = 1$ data actually performs more poorly than the $s = 0$ data — with three extra "false alarms". Even three years after the event ($s = 3$), half of alarms are false and two out of nine slippages go undetected.

For identifying the sign of the change in the CAB, subsequent vintages of data tend to do relatively better than in the slippage case, but significant errors remain even using data three years after the event.

In the case where the sample is restricted to high deficit countries, around three quarters of fiscal loosenings are detected in real time. Three years after the event, nearly all of them are detected. However, the number of false alarms is virtually unchanged between $s = 0$ and $s = 3$.

For the full sample the picture is similar. As s increases, the number of missed alarms falls quite sharply, suggesting extra vintages of data can help significantly in picking up slippages. As s increases, the number of false alarms falls, but not in a monotonic fashion — for the "fiscal slippage" case,

[14] If a country fails to consolidate enough, they can be required to make a non-interest bearing deposit, which can eventually be converted into a fine.

[15] Unlike the previous analysis our sample here ends in 2005. This is because we wish to test the performance of different vintages of data for the same sample period (the motivation here is the same as in Section 2 for calculating RMSEs across vintages). The December 2008 vintage constitutes our final data, so the last year for which we can have an $s + 3$ vintage of data (and which is distinct from the *ex post* data) is 2005. This differences in samples explains the discrepancy between the first set of results in Table 2, and the results in Table 4.

Table 4: Warnings by Vintage

Trigger	Correct Alarm	False Alarm	Missed Alarm	Correct All Clear
	Fiscal Slippage: CAB worsens by more than $1.5pp$			
$s = 0$	5	1	9	156
$s = 1$	5	4	9	153
$s = 2$	6	2	8	155
$s = 3$	7	2	7	155
	Fiscal Loosening (full sample): change in CAB is negative			
$s = 0$	55	34	13	69
$s = 1$	61	13	7	90
$s = 2$	64	24	4	79
$s = 3$	67	24	1	79
	Fiscal Loosening (restricted sample): change in CAB is negative			
$s = 0$	12	5	4	18
$s = 1$	14	3	2	20
$s = 2$	14	3	2	20
$s = 3$	15	4	1	19

Source: OECD *Economic Outlook* 58–84, authors' own calculations.

the $s = 1$ data is actually better than the $s = 2$ or $s = 3$ vintages. Three years after the event, more than one third of all alarms are still false.

Deficit Revisions or Problems with Cyclical Adjustment?

The above analysis demonstrates that CAB figures are unreliable. But it cannot determine whether the source of the measurement error is the actual deficit figures, or in the methodology used for the cyclical adjustment. To isolate the effect of the cyclical adjustment process we construct an artificially corrected real time data set, where we eliminate any errors in the budget deficit estimate.

Specifically, we take the *ex post* data for the unadjusted balance, and subtract from it the real time estimate of the cyclical component of budget balance. The only possible source of discrepancy remaining between this constructed time series and the *ex post* data, is errors in the real time cyclical adjustment process.[16]

[16]The intuition of this technique can be characterised by imagining that at the end of every

We then repeat the analysis of the previous section, substituting real time data for our hypothetical data. The results are shown below in Table 5.

Table 5: Hypothetical versus *Ex Post* Data: Slippage of more than $1.5pp$ in one year

	Correct Alarm	False Alarm	Missed Alarm	Correct All Clear
	Fiscal Slippage: CAB worsens by more than $1.5pp$			
Frequency	$10(+4)$	$5(+2)$	$6(-4)$	$207(-2)$
Average RT Change	-2.35	-2.07	-0.95	-0.03
Average EP Change	-2.38	-0.03	-1.73	0.29
	Fiscal Loosening (Full Sample): change in CAB is $+0.5pp$ *or less*			
Frequency	$84(+11)$	$34(-25)$	$2(-11)$	$108(+25)$
Average RT Change	-0.77	0.20	1.56	0.11
Average EP Change	-0.94	0.11	-0.46	0.61
	Fiscal Loosening (Restricted Sample): change in CAB is $+0.5pp$ *or less*			
Frequency	$23(+12)$	$9(-1)$	$0(-12)$	$24(+1)$
Average RT Change	-0.60	0.24	–	0.15
Average EP Change	-0.83	0.15	–	1.38

Source: OECD *Economic Outlook* 58–84, authors' own calculations.
Note: numbers in the brackets report the difference compared to the real time case as reported in Table 2.

For the slippage tests, the hypothetical data is better than the real time data. With the trigger set at $-1.5pp$, an extra four fiscal slippages are picked up, but still more than a third of slippages go undetected. However, the number of false alarms actually rises to five. Thus our hypothetical data actually performs even worse than the real time data.

The fact significant errors remain even after eliminating the effect of measurement error in the fiscal balance shows that the cyclical adjustment process is, in itself, a source of significant real time measurement errors. For the significant improvement test, a similar picture emerges. On the full sample, the hypothetical data detects an additional 11 slippages, virtually eliminating the problem of missed alarms; and the number of false alarms falls by 25. How-

year, a hypothetical statistician is handed the *ex post* data on the budget deficit to GDP ratio, which he/she must then cyclically adjust using the real time data available at the time.

ever, more than a quarter of alarms are still false. Restricting the sample to countries with a real time actual deficit of more than 3% tells a similar story. The number of missed alarms falls to zero, but only one false alarm is eliminated.

These results also demonstrate that significant real time data errors exist which are not the result of creative accounting by governments. Although the data on actual deficits are determined by the OECD in consultation with the national governments[17] the cyclical adjustment is done solely by the OECD, on the basis of the OECD's estimated output gap and budgetary elasticities. Thus, the only source of measurement error in these hypothetical data series is in the cyclical adjustment process, which by construction is entirely independent of national governments and hence cannot be due to creative accounting by governments.

4. Data Revisions and Fiscal Slippages: Is there a link?

We now consider how the error associated with real time CAB figures varies with the size of fiscal imbalance and the change in the *ex post* CAB figures. These asymmetries are potentially important, because from the point of view of fiscal surveillance, the authorities will be particularly interested in those countries who have CABs that are negative (or close to zero) and which are declining. If the CAB tends to be less efficient for, say, high and rising CAB surpluses, then this is may be less of a problem from the monitors' point of view than if the real time CAB performs more poorly for countries with worse fiscal health.

To check on this, we perform a regression analysis, to gauge which variables can explain the measurement error. The error, $u^{\bar{b}}$, is defined as a difference between the real time CAB estimate and the *ex post* measure, $u^{\bar{b}}_t = \bar{b}_{t|t} - \bar{b}_{t|2008}$. Since the dependent variable is the real time measurement error, a positive coefficient indicates an overly optimistic real time figure (i.e. a better CAB position in real time than *ex post*). We estimate five equations, each of them is estimated as a panel with fixed effects used to mop up country specific effects. The results are tabulated in Table 6.

The baseline result is regression I. The constant is positive and significant, indicating that on average the real time CAB is 47 basis points more positive

[17]Note however, the point raised earlier that the OECD still applies it's own judgements and does not take national governments figures at face value. See also Beetsma and Giuliodori (2008) for supporting arguments that OECD figures are independent.

than the *ex post* figure. The level of the *ex post* CAB, $\bar{b}_{t|2008}$, is not significant, indicating that the real time measurement error is not correlated with the *ex post* state of public finances. However, the *change* in the CAB, $\Delta\bar{b}_{t|2008}$, is significant, and has a negative sign. That means that the real time CAB figure tends to be overoptimistic when public finances are worsening, but over pessimistic when public finances are improving.

Table 6: Errors in the Real Time CAB

	Regression				
	I	II	III	IV	V
			Dependent variable: $u_t^{\bar{b}}$		
constant	0.476^{***}	0.334^{***}	0.394^{***}	0.209^{**}	0.207^{**}
$\bar{b}_{t\|2008}$	-0.000	-0.013	-0.038		
$b_{t\|2008}$				-0.056	-0.050
$c_{t\|2008}$				0.254^{***}	0.228^{***}
$\Delta\bar{b}_{t\|2008}$	-0.620^{***}	-0.514^{***}	-0.512^{***}		-0.473^{***}
$\Delta b_{t\|2008}$				-0.462^{***}	
$\Delta c_{t\|2008}$				0.382^{***}	
$\bar{b}^2_{t\|2008}$			-0.010		
$(\Delta\bar{b}_{t\|2008})^2$			-0.005		
$u_{t-1}^{\bar{b}}$		0.407^{***}	0.410^{***}	0.353^{***}	0.373^{***}
$Adj.R^2$	0.415	0.585	0.626	0.611	0.611
Nob	228	228	228	228	228

Source: OECD *Economic Outlook* 58–84, authors' own calculations.
Note: ** and *** denote statistical significance at 5% and 1% significance level respectively, Nob — number of observations.

The addition of a lagged dependent variable (regression II) results in a significant (positive) coefficient, indicating that measurement errors are serially correlated. It does not materially affect the sign, magnitude or statistical significance of the other coefficients. Regression III then tests for possible asymetries, by adding in the square *ex post* CAB, and the square of the change in the CAB. The coefficient on both variables is insignificant, thus providing no evidence of an asymmetry.[18]

[18]As a further check, these regressions were repeated with the modulus of the level and

It could be that the link between real time measurement error varies, depending on the extent to which the CAB shows up in the overall balance.[19] To test this, in regression IV, the *ex post* CAB (level and change) is decomposed into its constituents parts, i.e. actual budget balance, b, and cyclical component, c. The identities utilised here are $\bar{b}_{t|2008} = b_{t|2008} - c_{t|2008}$ for the level data and $\Delta\bar{b}_{t|2008} = \Delta b_{t|2008} - \Delta c_{t|2008}$ for the change in the CAB. If the decomposition is unnecessary, then the coefficient on the level or change of b and c should sum to zero.[20] This restriction is rejected for the level coefficients, but not for the change coefficients. Thus, the source (cyclical component or overall balance) only matters for the level of the CAB, but not for the change.

Thus regression V represents our preferred specification — a joint coefficient for the change in the CAB, but separate ones for the levels. Successive modifications to the specification reduce the value the constant term, but it still remains significant, implying that on average real time CAB figures underestimate the *ex post* figures by about 20 basis points.

The coefficient on $\Delta\bar{b}$ remains negative and significant, meaning that real time CAB figures are overly optimistic during fiscal loosenings and overly pessimistic during fiscal tightenings.

Finally the coefficient b is insignificant, but the coefficient on c is significant and negative. That implies that the overall balance does not affect the real time measurement error of the CAB. However, a strong cyclical position results in an overly pessimistic assessment of the CAB, and a weak cyclical position results in an overly negative assessment of the CAB.

These results are robust to various different treatments of panel effects. Including year effects, and/or removing country effects makes no difference to the sign or significance of the coefficients, nor does it make large changes to their magnitudes. The same goes for using random effects in place of fixed effects.

---

change in the CAB in place of their squared values. This alternative specification found no evidence of an asymmetry.

[19]Consider two equally sized changes to the CAB. In the first, the cyclical component of the budget balance does not change, and so the whole CAB change shows up one-for-one in the total balance. In the second, the change in the CAB is entirely offset by a change in the cyclical component, and hence the total balance is unchanged (for example, a government spending extra revenues during a boom time). Although the size of the changes is identical, it could be that the first one was easier to spot because it showed up in the overall balance.

[20]Or equivalently, the coefficient on b is equal to minus one times the coefficient on c.

5. Concluding Remarks

This paper investigates the accuracy of real time CAB figures in three ways. First, the OECD's own real time figures are actually slightly worse at forecasting the final OECD figure than a simple benchmark method is. Second, CAB figures perform poorly in identifying episodes of fiscal slippage and at correctly picking out the sign of the change in the *ex post* balance. Third, there is a systematic pattern to the real time measurement errors which implies that they are over optimistic during times of fiscal slippage (and over pessimistic in times of fiscal tightening).

Taken together, these provide strong evidence that in real time CABs are rather poor and inaccurate measures of fiscal stance. That suggests that caution should be applied in using them as a tools for fiscal monitoring. Broadly speaking, we find roughly equal contributions of deficit measurement and the cyclical adjustment process to the overall measurement error.

In terms of the existing literature, our results suggest that caution should be taken in the interpretation of an *ex post* CAB as a measure of the policymaker's desired fiscal stance. Many fiscal slippages which show up *ex post* were not apparent at the time; and conversely many apparent episodes of real time fiscal slippage turn out not to be so when *ex post* data is used. By demonstrating the large magnitude of the real time measurement error, our results support the argument of Orphanides (2001) that attempts to empirically estimate policymakers' reaction functions should be based on real time rather than *ex post* data.

On the policy front, our results suggest that attempts to monitor fiscal policies using CABs should be approached with even greater caution. Similarly, attempts to use CABs in a rule based framework (such as the SGP) may be subject to enforcement difficulties. Governments could legitimately claim that the data on the CAB were too preliminary, and could argue that more time was needed to accurately gauge the fiscal stance. A more aggressive stance by monitors and/or enforcers of such rules could be difficult, because it would lead to allegations of "crying wolf", i.e. frequently issuing alarmist statements which are later found to be unfounded.

6. References

Beetsma, R. and Giuliodori, M. (2008). On the Relationship Between Fiscal Plans in the European Union: An Empirical Analysis Based on Real-Time Data. *Journal of Comparative Economics*, 36, pp. 221–242.

Bernoth, K., Hughes Hallett, A. and Lewis, J. (2008). Did Fiscal Policymakers Know What They Were Doing? *CEPR Discussion Paper*, 6758.

Bjornland, H. C., Brubakk, L., and Jore, A.S. (2005). Measuring the Output Gap in Norway — an Assessment. *Working Paper Series of Norges Bank*, 2.

Blanchard, O. (1990). Suggestions for A New Set of Fiscal Indicators. *OECD Economics Department Working Paper*, 79.

Blanchard, O., Giavazzi, F. and Pagano, M. (2000). Searching for Non-Linear Effects of Fiscal Policy: Evidence from Industrial and Developing Countries. *NBER Working Papers*, 7460.

Breuss, F. and Roeger, W. (2005). The SGP fiscal rule in the case of sluggish growth: simulations with the Quest Model. *Journal of Policy Modelling*, 27.

Buiter, W. and Grafe, C. (2002). Patching Up the Pact — Some Suggestions for Enhancing Fiscal Sustainability and Macroeconomic Stability in an Enlarged European Union. *Centre for Economic Policy Research Discussion Paper*, 3496, reprinted in *Economics of Transition*, 27, pp. 67–102

Buti, M., Eijffinger, S. and Franco, D. (2003a). Revising the Stability and Growth Pact: grand design or internal adjustment? The European Economy. *Economic Papers*, 180, European Commission, Brussels.

Buti, M., Franco, D. and Ongega, H. (2003b). Revising the Stability and Growth Pact: Grand Design or Internal Adjustment. *CEPR Discussion Papers*, 3692.

Cimadomo, J. (2008). Fiscal Policy in Real Time. *ECB Working Paper*, 919.

Economic Outlook (1995). No. 58, Organisation for Economic Co-operation and Development.

Economic Outlook (1996). No. 60, Organisation for Economic Co-operation and Development.

Economic Outlook (1997). No. 62, Organisation for Economic Co-operation and Development

Economic Outlook (1998). No. 64, Organisation for Economic Co-operation and Development.

Economic Outlook (1999). No. 66, Organisation for Economic Co-operation and Development.

Economic Outlook (2000). No. 68, Organisation for Economic Co-operation and Development.

Economic Outlook (2001). No. 70, Organisation for Economic Co-operation and Development.

Economic Outlook (2002). No. 72, Organisation for Economic Co-operation and Development.

Economic Outlook (2003). No. 74, Organisation for Economic Co-operation and Development.

Economic Outlook (2004). No. 76, Organisation for Economic Co-operation and Development.

Economic Outlook (2005). No. 78, Organisation for Economic Co-operation and Development.

Economic Outlook (2006). No. 80, Organisation for Economic Co-operation and Development.

Economic Outlook (2007). No. 82, Organisation for Economic Co-operation and Development.

Economic Outlook (2008). No. 84, Organisation for Economic Co-operation and Development.

European Commission (2002). Public Finances in EMU. *European Economy*, 3, Office for Official Publications of the EC, Brussels.

Forni, L. and Momigliano, S. (2004). Cyclical Sensitivity of Fiscal Policies Based on Real-Time Data. *Termi di Discussione*, Banca d'Italia, Rome.

Galí, J. and Perotti, R. (2003). Fiscal Policy and Monetary Integration in Europe. *Economic Policy*, 37.

Girouard, N. and André, C. (2005). Measuring Cyclically-adjusted Budget Balances for OECD Countries. *OECD Economics Department Working Papers*, 434, OECD Publishing.

Golinelli, R., and Momigliano, S. (2008). The cyclical response of fiscal policy in the euro area. Why do results of empirical research differ so strongly? Paper presented at European Commission Workshop *Achieving and Safeguarding Sound Fiscal Positions*, January 17, Brussels, Web site: http://ec.europa.eu /economy_finance/events/event9197_en.htm.

González-Mínguez, J.M., Hérnandez de Cos, P. and Del Río, A. (2003). An Analysis of the Impact of GDP Revisions on Cyclically Adjusted Budget Balances. *Documento Ocasional*, 309, Banco De España, Madrid.

Guay, A. and St-Amant, P. (1997). Do the Hodrick-Prescott and Baxter-King Filters Provide a Good Approximation of Business Cycles? *CREFE Working*

Paper, 53, Center for Research on Economic Fluctuations and Employment, Université du Québec á Montréal.

Hughes Hallett, A., Kattai, R. and Lewis, J. (2007). Early Warning or Just Wise After the Event? The problem of Using Cyclically Adjusted Budget Deficits for Fiscal Surveillance. *CEPR Discussion Paper*, 6213.

Hughes Hallett, A. and Lewis, J. (2008). European Fiscal Discipline Before and After EMU: Crash Diet or Permanent Weight Loss. *Macroeconomic Dynamics*, 12, pp. 404–428

Hughes Hallett, A., Lewis, J. and von Hagen, J. (2004). Fiscal Policy in Europe 1991–2003: An Evidence Based Analysis. Centre for Economic Policy Research, London.

IMF (2008). Financial Stress, Downturns and Recoveries. *World Economic Outlook*, International Monetary Fund, Washington.

Jonung, L. and Larch, M. (2006). Fiscal Policy in the EU: Are the Official Output Forecasts Biased? *Economic Policy*, pp. 493–533.

Koen, P. and van den Noord, P. (2005). Fiscal Gimmickry in Europe: One-Off Measures and Creative Accounting. *OECD Economics Department Working Papers*, 417.

Orphanides, A. (2001). Monetary Policy Rules Based on Real Time Data. *The American Economic Review*, 91(4), pp. 964–985.

Orphanides, A. and van Norden, S. (2001). The Unreliability of Output Gap Estimates in Real Time. *The Review of Economics and Statistics*, 84(4), pp. 569–583.

Sapir, A. (2004). *An Agenda for a Growing Europe*. Oxford University Press, Oxford.

St-Amant, P. and van Norden, S. (1997). Measurement of the Output Gap: A Discussion of Recent Research at the Bank of Canada. *Technical Report*, 79, Bank of Canada.

van den Noord, P. (2000). The Size and Role of Automatic Fiscal Stabilizers in the 1990s and Beyond. *OECD Economics Department Working Papers*, 230, OECD Publishing.

von Hagen, J. and Wolff, G. (2006). What do deficits tell us about debt? Empirical evidence on creative accounting with fiscal rules in the EU. *Journal of Banking and Finance*, Elsevier, 30(12), pp. 3259-3279.

Appendix 1. Revisions in gap estimates: Root Mean Squared Error (own gap estimates are compared to OECD's *ex post* data)

	s=0		s=1		s=2		s=3		s=4		mean		p-value
	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	of t-test*
Australia	0.86	0.73	0.97	0.91	1.07	1.12	0.95	1.08	0.93	1.04	0.96	0.98	0.79
Austria	1.27	1.12	1.32	1.07	0.88	0.88	0.55	0.56	0.46	0.48	0.90	0.82	0.74
Belgium	0.98	1.67	0.63	1.32	0.44	1.10	0.35	1.09	0.22	0.84	0.52	1.20	0.01
Canada	1.21	0.87	1.18	0.85	1.02	0.75	0.97	0.71	0.75	0.56	1.03	0.75	0.02
Denmark	0.73	0.99	0.78	0.85	0.61	0.61	0.53	0.45	0.41	0.69	0.61	0.72	0.38
Finland	2.10	1.83	2.04	1.74	1.60	1.58	1.20	1.25	0.99	1.07	1.59	1.49	0.74
France	0.71	1.28	0.67	1.06	0.60	1.04	0.56	1.07	0.48	1.15	0.60	1.12	0.00
Germany	1.19	1.60	0.76	1.27	0.32	0.97	0.30	0.70	0.39	0.52	0.59	1.01	0.14
Greece	0.79	0.63	0.60	0.73	0.52	1.11	0.47	1.25	0.40	1.02	0.56	0.95	0.02
Ireland	1.38	1.36	1.93	1.99	1.51	1.61	1.52	1.20	1.59	0.88	1.59	1.41	0.42
Italy	0.83	1.33	0.97	0.77	1.21	0.43	1.45	0.75	1.55	0.94	1.20	0.84	0.11
Japan	1.85	2.97	1.33	2.16	1.24	1.65	0.95	1.02	0.80	0.98	1.23	1.76	0.24
Netherlands	1.45	1.48	1.23	1.30	0.86	0.90	0.60	0.68	0.38	0.74	0.90	1.02	0.66
Norway	1.65	1.09	1.17	1.01	1.20	1.00	1.17	0.94	1.10	0.62	1.26	0.93	0.03
Portugal	2.39	2.94	1.74	2.48	1.14	1.91	0.75	1.39	0.55	0.91	1.31	1.93	0.25
Spain	1.28	1.02	1.25	1.17	1.18	1.17	1.13	1.09	1.15	0.97	1.20	1.08	0.05
Sweden	0.88	1.09	0.83	1.18	0.67	0.94	0.43	0.76	0.42	0.85	0.65	0.96	0.03
United Kingdom	0.86	0.93	0.80	0.85	0.73	0.78	0.70	0.52	0.62	0.30	0.74	0.68	0.61
United States	1.35	1.61	1.25	1.48	0.96	1.40	0.70	1.27	0.56	1.24	0.96	1.40	0.03
Mean	1.25	1.40	1.13	1.27	0.93	1.10	0.80	0.94	0.72	0.83	0.97	1.11	

Source: OECD Economic Outlook 58–84, authors' own calculations. * The null hypothesis of the test is the equality of means, p-value of 0.01 refers to the rejection of the null at 1% significance level.

Appendix 2. Revisions in gap estimates: Root Mean Squared Error (own gap estimates are compared to own *ex post* data)

	s=0		s=1		s=2		s=3		s=4		mean		p-value
	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	of t-test*
Australia	0.77	0.73	0.64	0.91	0.76	1.12	0.56	1.08	0.30	1.04	0.61	0.98	0.01
Austria	0.98	1.12	1.03	1.07	0.58	0.88	0.29	0.56	0.31	0.48	0.64	0.82	0.39
Belgium	0.78	1.67	0.52	1.32	0.38	1.10	0.41	1.09	0.34	0.84	0.49	1.20	0.00
Canada	1.01	0.87	0.96	0.85	0.79	0.75	0.72	0.71	0.49	0.56	0.79	0.75	0.69
Denmark	0.78	0.99	0.73	0.85	0.55	0.61	0.46	0.45	0.38	0.69	0.58	0.72	0.29
Finland	1.35	1.83	1.36	1.74	0.96	1.58	0.51	1.25	0.35	1.07	0.91	1.49	0.05
France	0.89	1.28	0.67	1.06	0.40	1.04	0.27	1.07	0.16	1.15	0.48	1.12	0.00
Germany	1.33	1.60	0.88	1.27	0.59	0.97	0.47	0.70	0.23	0.52	0.70	1.01	0.29
Greece	0.92	0.63	0.66	0.73	0.47	1.11	0.40	1.25	0.31	1.02	0.55	0.95	0.04
Ireland	3.40	1.36	3.34	1.99	2.49	1.61	1.88	1.20	1.33	0.88	2.49	1.41	0.04
Italy	1.19	1.33	0.64	0.77	0.37	0.43	0.38	0.75	0.40	0.94	0.60	0.84	0.28
Japan	1.67	2.97	1.19	2.16	1.19	1.65	0.85	1.02	0.63	0.98	1.11	1.76	0.16
Netherlands	1.41	1.48	1.12	1.30	0.71	0.90	0.53	0.68	0.41	0.74	0.84	1.02	0.47
Norway	1.18	1.09	0.59	1.01	0.61	1.00	0.64	0.94	0.62	0.62	0.73	0.93	0.18
Portugal	2.02	2.94	1.30	2.48	0.73	1.91	0.74	1.39	0.89	0.91	1.14	1.93	0.11
Spain	0.94	1.02	0.88	1.17	0.85	1.17	0.84	1.09	0.83	0.97	0.87	1.08	0.00
Sweden	0.82	1.09	0.67	1.18	0.61	0.94	0.52	0.76	0.39	0.85	0.60	0.96	0.01
United Kingdom	1.00	0.93	0.89	0.85	0.85	0.78	0.82	0.52	0.73	0.30	0.86	0.68	0.18
United States	1.14	1.61	1.10	1.48	0.80	1.40	0.50	1.27	0.27	1.24	0.76	1.40	0.01
Mean	1.24	1.40	1.01	1.27	0.77	1.10	0.62	0.94	0.49	0.83	0.83	1.11	

Source: OECD *Economic Outlook* 58–84, authors' own calculations. * The null hypothesis of the test is the equality of means, p-value of 0.01 refers to the rejection of the null at 1% significance level.

Appendix 3. Revisions in CAB estimates: Root Mean Squared Error (own CAB estimates are compared to OECD's *ex post* data)

	s=0		s=1		s=2		s=3		s=4		mean		p-value
	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	of t-test*
Australia	1.04	1.10	0.88	0.93	0.99	1.13	0.98	1.04	0.78	0.73	0.93	0.99	0.57
Austria	1.17	1.12	1.17	1.20	1.09	1.31	1.03	1.18	0.45	0.34	0.98	1.03	0.84
Belgium	0.64	0.99	0.48	0.90	0.40	0.90	0.40	0.95	0.21	0.67	0.43	0.88	0.00
Canada	1.01	1.19	0.68	0.94	0.39	0.64	0.30	0.54	0.27	0.28	0.53	0.72	0.40
Denmark	1.13	1.57	1.20	1.49	1.04	1.30	0.83	1.08	0.69	0.96	0.98	1.28	0.08
Finland	1.61	1.14	1.17	0.85	0.90	0.67	0.66	0.65	0.70	1.35	1.01	0.93	0.73
France	0.52	0.82	0.33	0.73	0.32	0.60	0.38	0.61	0.28	0.42	0.37	0.64	0.01
Germany	2.19	2.33	2.07	2.02	2.12	2.08	2.12	2.11	2.02	2.04	2.10	2.12	0.88
Greece	2.51	2.43	1.77	1.77	1.47	1.55	1.13	1.34	1.07	1.43	1.59	1.70	0.73
Ireland	1.58	1.45	0.70	0.56	0.47	0.54	0.38	0.48	0.43	0.44	0.71	0.69	0.95
Italy	1.16	1.17	0.79	0.67	0.67	0.30	0.78	0.55	0.94	0.66	0.87	0.67	0.26
Japan	2.16	2.79	2.10	2.57	2.12	2.37	1.75	2.00	1.83	1.94	1.99	2.33	0.10
Netherlands	2.20	2.38	1.61	1.63	1.69	1.62	1.66	1.55	1.54	1.50	1.74	1.74	0.97
Norway	10.34	2.31	10.53	1.10	9.97	0.42	9.83	0.75	9.93	0.36	10.12	0.99	0.00
Portugal	2.18	2.24	1.65	1.79	1.14	1.35	0.94	1.06	0.65	0.73	1.31	1.43	0.75
Spain	0.72	0.82	0.56	0.67	0.39	0.65	0.54	0.78	0.44	0.53	0.53	0.69	0.07
Sweden	1.47	1.73	1.51	1.79	1.16	1.40	0.83	0.97	0.76	0.95	1.15	1.37	0.37
United Kingdom	0.91	1.04	0.50	0.56	0.82	0.41	0.74	0.33	0.74	0.20	0.74	0.51	0.18
United States	0.63	0.63	0.45	0.47	0.43	0.45	0.39	0.50	0.23	0.44	0.43	0.50	0.35
Mean**	1.38	1.54	1.09	1.19	0.98	1.04	0.88	0.97	0.78	0.84	1.02	1.12	

Source: OECD Economic Outlook 58–84, authors' own calculations. * The null hypothesis of the test is the equality of means, p-value of 0.01 refers to the rejection of the null at 1% significance level. ** Without Norway.

Appendix 4. Revisions in CAB estimates: Root Mean Squared Error (own CAB estimates are compared to own *ex post* data)

	s=0		s=1		s=2		s=3		s=4		mean		p-value
	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	own	OECD	of t-test*
Australia	0.78	1.10	0.65	0.93	0.74	1.13	0.72	1.04	0.57	0.73	0.69	0.99	0.01
Austria	1.11	1.12	1.09	1.20	1.11	1.31	1.06	1.18	0.26	0.34	0.93	1.03	0.68
Belgium	0.59	0.99	0.42	0.90	0.40	0.90	0.40	0.95	0.21	0.67	0.40	0.88	0.00
Canada	1.03	1.19	0.72	0.94	0.41	0.64	0.28	0.54	0.26	0.28	0.54	0.72	0.44
Denmark	1.30	1.57	1.33	1.49	1.19	1.30	1.01	1.08	0.89	0.96	1.14	1.28	0.37
Finland	1.11	1.14	0.75	0.85	0.41	0.67	0.34	0.65	0.65	1.35	0.65	0.93	0.18
France	0.54	0.82	0.35	0.73	0.28	0.60	0.23	0.61	0.14	0.42	0.31	0.64	0.01
Germany	2.42	2.33	2.12	2.02	2.21	2.08	2.20	2.11	2.10	2.04	2.21	2.12	0.26
Greece	2.47	2.43	1.71	1.77	1.38	1.55	1.09	1.34	1.09	1.43	1.55	1.70	0.64
Ireland	1.98	1.45	1.27	0.56	1.03	0.54	0.81	0.48	0.66	0.44	1.15	0.69	0.17
Italy	1.26	1.17	0.79	0.67	0.43	0.30	0.28	0.55	0.23	0.66	0.60	0.67	0.77
Japan	2.16	2.79	2.07	2.57	2.07	2.37	1.70	2.00	1.78	1.94	1.96	2.33	0.08
Netherlands	2.16	2.38	1.70	1.63	1.68	1.62	1.61	1.55	1.45	1.50	1.72	1.74	0.94
Norway	2.06	2.31	0.91	1.10	0.28	0.42	0.27	0.75	0.17	0.36	0.74	0.99	0.63
Portugal	1.77	2.24	1.13	1.79	0.76	1.35	0.64	1.06	0.52	0.73	0.96	1.43	0.21
Spain	0.60	0.82	0.49	0.67	0.34	0.65	0.49	0.78	0.37	0.53	0.46	0.69	0.01
Sweden	1.36	1.73	1.33	1.79	1.03	1.40	0.67	0.97	0.61	0.95	1.00	1.37	0.16
United Kingdom	0.87	1.04	0.77	0.56	0.52	0.41	0.48	0.33	0.45	0.20	0.62	0.51	0.53
United States	0.60	0.63	0.46	0.47	0.41	0.45	0.34	0.50	0.15	0.44	0.39	0.50	0.22
Mean	1.38	1.54	1.06	1.19	0.88	1.04	0.77	0.97	0.66	0.84	0.95	1.12	

Source: OECD *Economic Outlook* 58–84, authors' own calculations. * The null hypothesis of the test is the equality of means, p-value of 0.01 refers to the rejection of the null at 1% significance level.

Working Papers of Eesti Pank 2009

No 1
Martti Randveer, Tairi Rõõm
The Structure of Migration in Estonia: Survey-Based Evidence