Eesti Pank
Bank of Estonia

Early Warning or Just Wise After the Event? The Problem of Using Cyclically Adjusted Budget Deficits for Fiscal Surveillance

Andrew Hughes Hallett,
Rasmus Kattai and John Lewis

Working Paper Series
2/2007

Early Warning or Just Wise After the Event? The Problem of Using Cyclically Adjusted Budget Deficits for Fiscal Surveillance*

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

Abstract

The effectiveness of cyclically adjusted balances (CABs) as an indicator of the health of public finances depends on the accuracy with which cyclically adjusted figures can be calculated in real time. This paper measures the accuracy of such figures using a specially constructed real time dataset containing published values of deficits, output gaps and cyclically adjusted deficits from successive issues of OECD economic outlook. We find that data revisions are so great that *real-time* CABs have low power in detecting fiscal slippages as defined by the *ex post* data.

JEL Code: H62, H68

Keywords: fiscal surveillance, cyclically adjusted budget balance, real time data

*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 of the University of Amsterdam for allowing us to use the dataset he assembled, and Peter Keus of De Nederlandsche Bank for his work in collating an additional dataset. We thank seminar participants at the University of Amsterdam for useful comments. The views expressed in this paper are those of the authors and not necessarily those of De Nederlandsche Bank or the Bank of Estonia.

[†]George Mason University, School of Public Policy, 4400 University Drive, Fairfax, VA 22030, US; ahughesh@gmu.edu.

[‡]Research Department, Bank of Estonia, Estonia Pst 15, 15095, Tallinn, Estonia; rkattai@epbe.ee.

[§]Economics and Research Division, De Nederlandsche Bank, Postbus 98, 1000AB, Amsterdam, The Netherlands; j.lewis@dnb.nl.

Non-technical summary

The cyclically adjusted budget balance (CAB) is becoming more and more important tool for monitoring the health of public finances. Extracting cyclical factors form the actual budget balance sheds light on the underlying state of public finances. The idea of using the cyclically adjusted budget balance instead of the actual balance, as it is stated in the Stability and Growth Pact, is to get an early signal, which would enable to prevent the fiscal slippages.

The effectiveness of the CAB as an indicator of the state of public finances depends on the accuracy of CAB figures available to policymakers and to those who carry out the fiscal surveillance *at the time*, rather than the *ex post* data published several years after the event, which may differ substantially from the figures available in real time.

In the current paper a real-time dataset for OECD countries is utilised to quantify the error in the real time CAB estimates. We find that CAB's perform really poorly in real-time and that CAB's have extremely low power in detecting fiscal slippages as defined by the ex post data. Also we find that 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.

Contents

1. Introduction

The cyclically adjusted budget balance (CAB) is an increasingly important tool for monitoring the state of public finances. For example, in March 2003 EcoFin resolved that the "close to balance or in surplus" requirement of the SGP be explicitly assessed on the basis of cyclically adjusted budget figures. Their subsequent analysis of fiscal stability in Europe likewise makes extensive use of cyclically adjusted figures (European Commission, 2006).

The 2005 reforms to the Stability and Growth Pact placed a greater emphasis on the "preventative arm" of the pact, implying a shift in attention from actual, to cyclically adjusted budget deficits. In addition, the reforms specifically required that the adjustment paths for countries under the Excessive Deficit Procedure be specified in cyclically adjusted terms.[1]

The rationale for this emphasis is straightforward and well known. First, stripping out the influence of cyclical factors yields a measure of fiscal stance, which better captures the underlying state of public finances. Moreover, according to European Commission (2002), the stabilising role of fiscal policy should in general function through the operation of automatic fiscal stabilisers (as opposed to a discretionary change in fiscal stance by the authorities). Any loosening of the cyclically adjusted position is typically regarded as a slippage in fiscal discipline rather than as an appropriate policy response to a negative shock.[2]

Second, as the European Commission (2006) notes, fiscal loosening often occurs at the top of the cycle, rather than as a discretionary response in recessions. If governments expand their fiscal stance rather than the actual deficit during good times, this may show up more clearly in the cyclically adjusted rather than the actual deficit. Moreover, 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, 2003; Hughes Hallett et al, 2004).

The effectiveness of the CAB as an indicator of fiscal health by those entrusted with fiscal surveillance 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, which may differ substantially from the

[1] For a full account of the current fiscal surveillance framework in EMU, see Wierts et al (2006).

[2] The official position, laid out in 2002, is more nuanced. However, given that most EMU members have CABs of less than zero, it seems plausible to assume that the vast majority of decreases in the CAB would be considered unfavourable developments.

figures available in real time. For this reason, Berger, De Haan and Jansen (2003) argue that cyclically adjusted budget balances should play no role at all in a reformed SGP. This argument was well illustrated by the recent experience of the Netherlands, which was judged in June 2004 to be in excessive deficit for the year 2003, on account of a budget deficit of 3.2% of GDP. Looking at the European Commission's 2004 spring forecast which recorded a cyclically adjusted deficit of 2.8% of GDP for 2002, it is tempting to conclude that an obvious early warning sign had been overlooked. However, the data available *at the time* contained no such early warning — the European Commissions Autumn 2001 forecast estimated the CAB to be in surplus by 0.8% (see Appendix A). In this case, the excessive deficit was not the result of policymakers ignoring an early warning, but the result of the failure of real time CAB figures to sound the alarm in the first place.

Evaluating the usefulness of the CAB as an "early warning" mechanism requires an analysis based on real-time as opposed to *ex post* data. The information available to policymakers at the time may differ substantially from the information available after the event for two key reasons. First, data on deficits and GDP available to policymakers at the time are likely to be preliminary, and be subject to many revisions in subsequent periods. Second, extra observations of output beyond time t can improve the accuracy with with the time t output gap can be estimated: the so-called "endpoint problem". Therefore, in the same vein as Orphanides (2001), in order to evaluate the decisions of policymakers, we must analyse their behaviour using the data they had available to them at the time, rather than on the basis of *ex post* data which the policymaker did not have.

There is a small but growing literature on the use of real time data to evaluate fiscal policymakers behaviour. Forni and Momigli-ano (2004) estimate a fiscal policy reaction function using real time and find counter cyclical responses which do not show up when the same estimation is carried out with *ex post* data. On fiscal monitoring, Jonung and Larch (2006) investigate the effect 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, Hernández de Cos and Del Río (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 deficit remains unchanged. However, to our knowledge, 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.

Accordingly, the primary goal of this paper is to investigate how big is the difference between *real time* and *ex post* estimates of the CAB, and consequently assess its effectiveness as an “early warning" of fiscal loosening. To do this we first construct a formal model of fiscal monitoring using CABs, and then we quantify the effect of data revisions using a specially constructed real-time dataset taken from successive issues of OECD’s *Economic Outlook*.

2. The model

We commence with a brief description of the notation for time. In what follows, each variable has 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, suppose a researcher opens the 2006 edition of *Economic Outlook* and looks up the printed value of variable X for 2002 — the value that he finds would in our notation be denoted by $X_{2002|2006}$.

For ease of interpretation we define the second subscript relative to the first. If $s > 1$, then this corresponds the figure published s years after the event; if $s = 0$, then we have "real time" data in the sense that this corresponds to the information which the policymaker had at time t; and if $s < 0$, then we have a forecast of the future value of a variable, in $-s$ years time.

2.1. Actual and potential output

Actual output, Y is equal to potential output, Y^*, plus the output gap, $\tilde{Y}$:

$$Y_t = Y_t^* + \tilde{Y}_t. \tag{1}$$

Final output data, Y_t, is not available in real time, and so the authorities must use preliminary output data. The preliminary estimate Y_t based on the information known at $t+s$ is subject to an error term $\eta_t|t+s$:

$$Y_{t|t+s} = Y_t + \eta_{t|t+s}. \tag{2}$$

The level of potential output is not directly observable in real time either. It is estimated using data available at time $t+s$, subject to an error term ξ:

$$Y^*_{t|t+s} = Y_t^* + \xi_{t|t+s}. \tag{3}$$

Combining (1), (3) and (2) yields an expression for the error with which the output gap is estimated in real time:

$$\tilde{Y}_{t|t+s} = Y_{t|t+s} - Y^*_{t|t+s} \tag{4}$$

$$\Rightarrow \tilde{Y}_{t|t+s} = \tilde{Y}_t + \eta_{t|t+s} - \xi_{t|t+s}. \tag{5}$$

Depending on how potential output figures are derived, η and ξ may be positively correlated. That correlation is likely to be higher if purely statistical detrending methods are used: that is those which only identify the cyclical position of the economy based on the actual output data. Output gap estimates based on multivariate models, which are not solely driven by the dynamics of output time series, are expected to show lower correlation between η and ξ. In principle equation 5 implies that if actual and potential output are both underestimated to a similar extent, the error in the gap estimate may still be close to zero. Therefore the real time output gap data used for extracting the cyclical component of a budget balance may be close to its true value, although both actual and potential output have been wrongly estimated.

2.2. Setting fiscal policy

Revenues R, and expenditures E are given by,

$$R_t = (\tau_t + \alpha_t)Y^*_{t|t-1} + \beta_t Y_t + u_t, \tag{6}$$

$$E_t = (\gamma_t + \theta_t)Y^*_{t|t-1} + \phi_t Y_t + v_t, \tag{7}$$

where u_t and v_t are white noise error terms. The governments published plans are expressed in terms of τ, β, γ and ϕ. In addition this model introduces a distinction between the governments published plans and the "true" underlying structural budget balance captured by α and θ.

To develop an intuition of the model, these two equations can be divided by the forecast value of potential output, $Y^*_{t|t-1}$ ($s = -1$ here because we assume that the budget is set the year before it is enacted, and hence when setting the budget policymakers must use forecasts for the following year):

$$r_t = \tau_t + \alpha_t + \beta_t Y_t/(Y^*_t + \xi_{t|t-1}) + u^{'}_t, \tag{8}$$

$$e_t = \gamma_t + \theta_t + \phi_t Y_t/(Y^*_t + \xi_{t|t-1}) + v^{'}_t, \tag{9}$$

where $u^{'}_t = u_t/Y^*_{t|t-1}$ and $v^{'}_t = v_t/Y^*_{t|t-1}$. The cyclical elasticity is given by $\beta - \phi$, and expenditures/revenues which do not vary with the level of economic activity are given by $\tau_t + \alpha_t$ and $\gamma_t + \theta_t$ respectively.

The parameters α and θ are needed in order to motivate the role of a fiscal monitor. If these parameters are set equal to zero, then the only source of deviation in the CAB from its published values is the error in estimating potential output, ξ_t. In such case, the role of the monitor is simply reduced to that of auditing potential output.

We motivate the assumption that α and θ are not equal to zero in two ways. The first, more innocent interpretation is that these two parameters capture the errors in estimating the spending/revenue effects of a given fiscal programme, or the effects of acyclical shocks to spending/revenue which occur after the fiscal programme is enacted. The second, more cynical interpretation is that these parameters represent the fiscal authority's private information, which is concealed from the public (and the authorities monitoring them) and which make the published plans appear more prudent than they really are.[3]

Setting the model up in this way yields the crucial property that the budget balance responds in the same way to a given change in GDP regardless of whether the causal factor was a change in the output gap or a change in potential output. Thus the automatic component fiscal policy operates in response to any change in output, not just changes in the output gap.

The primary fiscal balance is defined as:

$$B_t = R_t - E_t. \tag{10}$$

We may substitute in equations (1) to (7) and re-arrange to yield the following expression for the primary balance:

$$\begin{aligned} B_t = \begin{pmatrix} \tau_t + \alpha_t + \beta_t \\ -\gamma_t - \theta_t - \phi_t \end{pmatrix} & Y^*_{t|t-1} && \text{active fiscal policy} \\ +(\phi_t - \beta_t)&\xi_{t|t-1} && \text{error in pot. GDP forecast} \\ +(\beta_t - \phi_t)&\tilde{Y}_t && \text{true cyclical component} \\ +u_t - &v_t. \end{aligned} \tag{11}$$

This enables a decomposition of the budget balance into its various components. The first is the structural budget generated by the spending and revenue parameters. The second term captures the effect of the governments error in estimating (forecasting) potential GDP. This creates the possibility that even

[3]This private information could take either sign, but in the context of the problem of monitoring the behaviour of fiscal authorities, it seems more sensible to restrict our attention to the case where the fiscal authority withholds private information in order to present an overly optimistic view of public finances.

with α and θ switched off, a government which aims to run a balanced budget in the long-run may fail to do so because it overestimates potential output growth. The third term captures the pure cyclical effects.

The true cyclically adjusted budget balance is defined as:

$$\bar{B}_t = (\tau_t + \alpha_t + \beta_t - \gamma_t - \theta_t - \phi_t)Y^*_{t|t-1} + (\phi_t - \beta_t)\xi_{t|t-1} + u_t - v_t. \quad (12)$$

In an ideal world world where $\xi = 0$ and the expected values of u_t and v_t equal zero, the fiscal monitor can, using aggregate deficit data and the announced fiscal policy, back out the values of α and θ.

In reality, however, the ideal will not be realised. Informational constraints mean that the fiscal monitor cannot directly uncover the values of α and θ, and hence cannot know the true cyclically adjusted budget balance. The best the monitor can do is to estimate the cyclically adjusted budget balance.

Re-writing (11) and (12) yields:

$$\bar{B}_t = B_t - (\beta_t - \phi_t)\tilde{Y}_t. \quad (13)$$

Taking expectations of (13), conditional on information at time $t+s$ yields:

$$\bar{B}_{t|t+s} = B_{t|t+s} - (\beta_t - \phi_t)\tilde{Y}_{t|t+s}. \quad (14)$$

This is the best estimate of the cyclically adjusted deficit available to the monitor at time $t+s$. Subtracting (12) from (14), defining error in the preliminary budget deficit data as $B_{t|t+s} = B_t + \epsilon_{t|t+s}$ and substituting in (5), allows us to write the error in estimating the cyclically adjusted budget deficit as:

$$\bar{B}_{t|t+s} - \bar{B}_t = (\phi_t - \beta_t)(\eta_{t|t+s} - \xi_{t|t+s}) + \epsilon_{t|t+s}. \quad (15)$$

This equation says that error in estimating the CAB will be equal to the error in estimating the actual budget balance, plus the error in the output gap times the cyclical sensitivity of the budget balance. This generates four key propositions.

Proposition 1: Errors in the actual deficit have a one-for-one effect on the CAB
Proof: Differentiating (15) with respect to $\epsilon_{t|t+s}$ yields:

$$\frac{\partial(\bar{B}_{t|t+s} - \bar{B}_t)}{\partial \epsilon_{t|t+s}} = 1. \quad (16)$$

Proposition 2: Overestimating potential output improves the CAB
Proof: Differentiating (15) with respect to $\xi_{t|t+s}$ yields:

$$\frac{\partial(\bar{B}_{t|t+s} - \bar{B}_t)}{\partial \xi_{t|t+s}} = (\beta_t - \phi_t). \tag{17}$$

If $\beta_t > 0$ (i.e. revenues rise with output) and $\phi_t < 0$ (i.e. expenditures fall as output rises) then the sign of this effect is positive. This formalises the claim of Jonung and Larch (2006) that if the CAB, rather than the actual balance, is targeted then policymakers can improve the CAB figure by overestimating potential (but not actual) output when submitting budgetary plans.

Proposition 3: Overestimating actual output worsens the CAB
Proof: Differentiating (15) with respect to $\eta_{t|t+s}$ yields:

$$\frac{\partial(\bar{B}_{t|t+s} - \bar{B}_t)}{\partial \eta_{t|t+s}} = (\phi_t - \beta_t). \tag{18}$$

If $\beta_t > 0$ and $\phi_t < 0$ then this is negative.

Proposition 4: The sensitivity of CAB estimates with respect to output gap errors depends on the elasticities of revenues and expenditures
Proof: Differentiating (15) with respect to $(\eta_{t|t+s} - \xi_{t|t+s}$ yields:

$$\frac{\partial(\bar{B}_{t|t+s} - \bar{B}_t)}{\partial (\eta_{t|t+s} - \xi_{t|t+s})} = (\phi_t - \beta_t). \tag{19}$$

In other words, the sensitivity of this error is equal to the size of the automatic fiscal stabilisers. In the limiting case where there are no automatic stabilisers, then the actual budget balance is equal to the CAB, and so the error in estimating the output gap has no effect at all on the CAB figure.

3. Real time versus ex post data: what's the difference?

A real time dataset must capture the actual information available to policymakers at a certain time. Successive vintages of data contain two principal pieces of information- new observations of variables, and revised figures for previous periods. Conditioning on information available at time t does not simply mean taking the final vintage of the data, and discarding observations made after time t, since this truncated dataset would also include the information about variables before time t which came from data revisions that occurred after time t.

3.1. The real time dataset

To quantify the magnitude of these errors, a new real time dataset was compiled using data published in successive issues of the OECD's *Economic Outlook*, (EO). The dataset consists of the published values of GDP, output gap and actual budget deficit series in each issue from December 1995 (Issue 58) to December 2005 (Issue 78), as well as the published values of cyclically adjusted budget balance.

When the writing of this paper commenced there was no publicly available OECD dataset. Therefore our datasets had to be compiled by taking successive issues of EO and collating them into a single file.[4] The starting point was the real dataset of various fiscal policy and output variables compiled by Massimo Giuliodori of the University of Amsterdam for Beetsma and Giuliodori (2006), who manually inputted values from paper copies of EO. This was supplemented by the work of Peter Keus, a statistician at De Nederlandsche Bank, using Optimal Character Recognition software to create a database from paper copies of EO.

Strictly speaking this methodology could still produce a dataset which departs from the actual information available at the time and/or the official" real time dataset which may be published by the OECD for several reasons:

- i) typographical errors in the original hard copies of *Economic Outlook*,
- ii) information lost due to rounding figures to one decimal place,
- iii) methodological changes in the compilation of statistics.[5]

However, the magnitude of these errors is likely to be small in practice, particularly when set against its specific advantages.

[4] Subsequently the OECD published a real time database, its earliest vintage was 1999, therefore we continued to use our own dataset on account of its larger number of observations.

[5] Whilst methodological changes may be important from the point of view of generating conceptually consistent series across vintages, they are of much less importance in this paper. For example, output gap methodologies may not be consistent from vintage to vintage. But it is still permissible to collate published values across vintages, since these represent the best available information on the output gap given the information (i.e. the raw data and the compilation methodology) available at time t.

First, it is, to our knowledge, the most comprehensive real time dataset of OECD data currently in use, and has more observations than the OECD's recently release real time dataset . This implies that our results are the first of their kind to use a comprehensive real time datset based on OECD data which, in contrast with other work, tends to use publicly available real time datasets. Only a handful of national statistical offices have published real time GDP data, and even fewer have published budget deficit data.

Second, given real time GDP data, the issue arises as to how to calculate the output gap. Simply applying a mechanistic procedure such as an HP filter, suffers from the fact that in reality, a policymaker would have had more information (outside of the real time dataset) and could have used a more sophisticated approach. Taking the figure reported by the OECD in *Economic Outlook* circumvents this problem and gives a figure which corresponds to the "best guess" of the output gap at the time, given all the information available. Moreover, it permits a direct comparison of mechanistic methods versus more complicated techniques of estimating the output gap in real time.

Third, since the data is compiled by an independent body, it is partially insulated against "political" bias in provisional figures and forecasts compiled by national governments.[6] 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, it is suggested, to present a more favourable picture of cyclically adjusted public finances. As well as providing a potentially more accurate picture information available at the time decisions were made, our dataset permits the possibility of testing for such a bias.

3.2. Output gaps across vintages

Since the true value of the output gap is never known, $\eta - \xi$ can never be known with certainty. Therefore it is proxied by comparing the estimate at the end of year $t + s$ with the final reported value. We take the root mean squared error (RMSE) of the output gap's deviation from the latest currently available time series:

$$\sigma_{\eta-\xi,s} = \sqrt{\frac{1}{n}\sum_{t=t_0}^{f-s}(\tilde{y}_{t|t+s} - \tilde{y}_{t|f})^2}, \tag{20}$$

where $\tilde{y}$ is the size of output gap as a ratio to potential GDP, t_0=1995 and

[6] This problem is not entirely eliminated by using OECD data, since OECD figures are typically generated following consultation with national governments.

time f represents the final observation available, this case from the December 2005 edition of *Economic Outlook*. We focus on five consecutive vintages, s = 0...4, which means that the maximum number of observations, n, is 11 if s = 0 and 7 if s = 4. In the following n is fixed at 7 for the sake of comparability across all values of s, which gives the sample period 1995–2001.

Table 1 shows the mean squared deviation between the output gap estimated at time $t + s$ and the final figures for all OECD countries for which a complete set of observations existed across the sample period (the Czech Republic, Hungary, Iceland, Luxembourg, Mexico, Slovakia, South Korea and Turkey were excluded on this basis).

Table 1: Revisions OECD output gap figures: RMSE

s=	0	1	2	3	4	Mean
Australia	0.95	0.87	0.93	0.75	0.83	0.87
Austria	1.71	1.69	1.33	0.91	0.61	1.25
Belgium	1.06	0.82	0.60	0.77	0.58	0.76
Canada	0.76	0.75	0.59	0.58	0.56	0.65
Denmark	1.36	1.07	0.82	0.78	1.03	1.01
Finland	2.50	2.48	2.25	1.65	1.25	2.03
France	0.57	0.48	0.41	0.50	0.61	0.51
Germany	1.70	1.48	1.32	1.01	0.75	1.25
Greece	1.00	0.88	1.11	1.29	1.00	1.06
Ireland	2.36	2.54	2.45	1.91	1.32	2.12
Italy	2.39	1.74	1.20	0.68	0.45	1.29
Japan	3.50	2.51	1.66	0.84	0.93	1.89
Netherlands	1.49	1.49	1.16	0.75	0.47	1.07
Norway	1.61	1.52	1.24	0.95	0.52	1.17
Portugal	2.18	1.67	1.19	0.75	0.31	1.22
Spain	1.57	1.74	1.69	1.48	1.17	1.53
Sweden	1.50	1.66	1.49	1.18	0.92	1.35
United Kingdom	0.89	0.82	0.72	0.47	0.27	0.64
United States	1.62	1.31	1.07	0.99	1.09	1.22
Mean	1.62	1.45	1.22	0.96	0.77	1.20

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

Estimating the output gap in real time (i.e. $s = 0$), yields an average RMSE of 1.6pp. Additional data reduces the RMSE, but even after four extra years of data, the error is still around 0.8pp of potential GDP. Equally striking is the marked variation across countries. In some cases, such as Denmark France and United States, the RMSE actually rises with s — largely due to the presence of a single large revision for one particular year. Conversely, the real time RMSE for countries such as Finland, Ireland, Italy and Japan is particularly large. The latter two suggest that output gap calculations are particularly vulnerable to revisions in the face of a structural break, and that it may take many years before this is corrected.

3.3. Budget balances across vintages

The revisions in budget balance, as a ratio to GDP, may be due to methodological changes in compiling governmental or national accounts statistics but also due to *ex post* corrections in both variables. We compute the root mean squared errors of actual budget balance:

$$\sigma_{\epsilon,s} = \sqrt{\frac{1}{n}\sum_{t=t_0}^{f-s}(b_{t|t+s} - b_{t|f})^2}, \tag{21}$$

where b stands for budget balance, given in per cent of GDP. As in equation (20), t_0 denotes the year 1995 and time f represents the final observation available, in this case the December 2005 edition of *Economic Outlook*. Analogously, the number of observations, n, is fixed at 7 and the sample period is 1995–2001. These are tabulated in table 2.

Table 2: Revisions in OECD budget balance figures: RMSE

s=	0	1	2	3	4	Mean
Australia	1.01	0.90	1.02	0.82	0.12	0.77
Austria	0.43	0.24	0.32	0.29	0.26	0.31
Belgium	0.54	0.28	0.30	0.28	0.15	0.31
Canada	1.16	0.88	0.67	0.59	0.39	0.74
Denmark	0.94	1.14	1.16	1.15	0.97	1.07
Finland	1.53	0.63	0.63	0.55	0.29	0.73
France	0.28	0.27	0.20	0.23	0.06	0.21
Germany	1.16	0.24	0.13	0.02	0.05	0.32
Greece	2.90	2.59	2.40	1.83	1.35	2.21
Ireland	1.37	0.40	0.34	0.29	0.20	0.52
Italy	0.79	0.49	0.36	0.33	0.00	0.40
Japan	0.87	0.83	0.55	0.57	0.54	0.67
Netherlands	0.94	0.29	0.20	0.14	0.09	0.33
Norway	1.46	0.86	0.24	0.24	0.10	0.58
Portugal	1.27	0.95	0.74	0.69	0.43	0.82
Spain	0.54	0.41	0.25	0.45	0.31	0.39
Sweden	0.93	1.02	1.02	0.87	0.62	0.89
United Kingdom	0.63	0.76	0.20	0.17	0.10	0.37
United States	0.94	0.75	0.68	0.48	0.08	0.59
Mean	1.04	0.73	0.60	0.53	0.32	0.64

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

The revisions for budget deficits tend to be smaller than for output gaps, but remain nevertheless significant. The RMSE in real time is just over unity, and remains at 0.7 after one year. The RMSE also appears to be more even across countries, although some differences still remain — with Greece, Ireland and Finland having markedly higher RMSEs.

3.4. Robustness of CAB estimates over time

As in the previous section, the magnitude of revisions in cyclically adjusted balance is measured in terms of a root mean squared error — comparing year $t+s$ estimate, $\overline{b}_{t|t+s}$, to the final reported value, $\overline{b}_{t|f}$ (lower case denotes that the cyclically adjusted budget balance is expressed as a ratio to potential output)[7]:

$$\sigma_{\eta-\xi,\epsilon,s} = \sqrt{\frac{1}{n}\sum_{t=t_0}^{f-s}(\overline{b}_{t|t+s} - \overline{b}_{t|f})^2}. \tag{22}$$

[7]Both the CAB figure and potential output are taken from the same vintage.

As earlier, the final "true value" is the one reported in 2005 *Economic Outlook* ($f = 2005$ with the exception of Germany, where $f = 2004$). The initial period is $t_0 = 1995$; and, as before, n = 7 (with the exception of Germany in case of which n = 6). The results are shown below in table 3.

Table 3: Revisions in OECD's CAB estimates: RMSE

s=	0	1	2	3	4	Mean
Australia	1.04	0.72	0.87	0.78	0.48	0.78
Austria	0.91	0.86	0.75	0.58	0.34	0.69
Belgium	0.57	0.33	0.43	0.68	0.54	0.51
Canada	1.27	1.02	0.78	0.67	0.40	0.83
Denmark	1.58	1.62	1.53	1.41	1.28	1.49
Finland	2.18	1.83	1.70	1.11	0.53	1.47
France	0.52	0.50	0.38	0.40	0.17	0.39
Germany	0.94	0.73	0.54	0.36	0.21	0.56
Greece	3.06	2.70	2.72	2.10	1.54	2.42
Ireland	2.05	1.08	0.94	0.93	0.84	1.17
Italy	1.56	1.01	0.63	0.49	0.30	0.80
Japan	1.94	1.56	1.12	0.85	0.71	1.23
Netherlands	1.30	0.95	0.45	0.30	0.51	0.70
Norway	2.13	1.17	0.35	0.82	0.35	0.96
Portugal	2.05	1.49	1.04	0.80	0.50	1.18
Spain	0.83	0.80	0.97	1.16	0.88	0.93
Sweden	1.55	1.52	1.39	1.35	1.04	1.37
United Kingdom	0.96	0.46	0.31	0.26	0.12	0.42
United States	0.53	0.49	0.50	0.51	0.42	0.49
Mean	1.38	1.09	0.95	0.82	0.60	0.97

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

During 1995–2001 the OECD's real time estimate of the CAB (i.e. s = 0) deviated from its final value by 1.4pp. The variance of the error across countries is significant, ranging from between between 0.5 pp in France and the United States and 3pp in Greece. The year after, the average error decreases to 1.1 pp and falls gradually to about 0.6pp four years later (see table 3). Compared to revisions in actual budget deficits (table 2), the error in CAB is roughly 50% higher across the vintages, indicating that revisions to the output gap are important source of error (for a full tabulation of the cyclical component of the budget deficit, see Appendix B).

4. Assessing the OECDs figures against a simple benchmark

In addition to the issue of how robust estimates are across vintages, it is also instructive to see how the OECD's real time figures for the output gap and CAB compare with a more simple real time benchmark.

To make the comparison we calculate our own measures of output gaps of different vintages using real time output time series data from different OECD *Economic Outlook* editions. To detrend the data, we apply the most commonly used univariate Hodrick-Prescott (HP) filter, which minimises the following loss function with the smoothing parameter λ set to 100:

$$L = \sum_{t=1}^{T} (Y_t - Y_t^*)^2 + \lambda \sum_{t=2}^{T-1} (\Delta Y_{t+1}^* - \Delta Y_t^*)^2. \tag{23}$$

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 in the end of sample[8]. This is a crucial shortcoming in the current context — dealing with the real time data means we are especially interested in the reliability of output gap estimates at the end of the sample. In order to mitigate the end-point problem, we add GDP forecasts for five fears ahead before applying HP filter[9]. Forecasts are produced with ARIMA models, which are automatically estimated by the algorithm built in the TRAMO-SEATS seasonal adjustment program, provided by EViews. The revisions across vintages implied by this technique are shown in table 4.

The root mean squared error of our own calculated gap estimates in real time (i.e. s = 0) exceeds the OECD's on average (compare tables 4 and 1). But, for s $= 1$ the errors are roughly equal; and the accuracy of our estimates increases and even surpasses the OECD figures if s is larger than one. These results are shown in greater detail in Appendix C.

Although the latter seems to support the reliability of HP filtered output gap, the comparison may be distorted because of differences in the final gap data. Our own calculated real time series are compared to our own final output gap estimates and the OECD's real time estimates, and the OECDs final output

---

[8] The issue is raised, for example, by Guay and St-Amant (1997) and St-Amant (1997).

[9] Although Kattai and Vahter show, based on Estonian GDP data, that even without extrapolating the available time series a simple univariate HP filter has at least as good or even better real time properties as the more sophisticated methods, such as the Baxter-King, Watson, Harvey-Clark, Kuttner or Gerlach-Smets models.

gap respectively. To check on the possibility of distortions, we define the final gap data as the the OECD's latest estimates in both cases.

Table 4: Revisions in own output gap: RMSE

s=	0	1	2	3	4	Mean
Australia	1.31	1.10	1.05	0.59	0.24	0.86
Austria	1.02	1.00	0.58	0.24	0.18	0.60
Belgium	0.98	0.96	0.36	0.41	0.24	0.59
Canada	2.57	1.87	1.34	0.85	0.36	1.40
Denmark	1.17	1.01	0.59	0.46	0.45	0.73
Finland	5.41	3.75	2.05	1.08	0.39	2.54
France	1.11	0.82	0.63	0.35	0.16	0.62
Germany	1.02	0.71	0.48	0.50	0.17	0.57
Greece	2.89	2.04	1.24	0.81	0.42	1.48
Ireland	4.83	3.77	2.42	1.38	0.47	2.58
Italy	1.11	0.90	0.57	0.43	0.30	0.66
Japan	1.96	1.51	0.97	0.38	0.56	1.08
Netherlands	1.93	1.54	1.02	0.51	0.23	1.05
Norway	1.47	1.21	0.81	0.71	0.47	0.93
Portugal	2.05	1.50	0.85	0.70	0.66	1.15
Spain	1.69	1.16	0.61	0.45	0.41	0.87
Sweden	2.56	1.58	1.00	0.61	0.27	1.20
United Kingdom	1.75	1.18	0.88	0.71	0.54	1.01
United States	1.82	1.35	0.97	0.62	0.39	1.03
Mean	2.03	1.52	0.97	0.62	0.37	1.10

Source: OECD *Economic Outlook* 57-69, authors' own calculations.

By their nature potential output and the output gap are unobservable. Whatever method is used to extract potential output from the actual output data, there is no way to say which measure is closer to the "true value". However, more sophisticated methods may give a result which is closer to what is believed to be the "true value". Therefore we take OECD's final gap estimate as a benchmark for the HP filtered gap. The results are shown in table 5.

Using the OECD's final output gap data as the final "true" value increases the average error across the countries in the own calculated gap, as one would expect. Comparing tables 4 and 5 we see that the values in table 5 are typically higher than the corresponding figures in 4.

Table 5 details how well our own real time output gap estimates can predict that final OECD figures for output gaps. Comparing this with table 1, which shows how well the OECD's own real-time output gaps match up with the OECD's final figures, yields the interesting result that the OECD's gap clearly outperforms our own calculations only in the case where $s = 0$ or s≤1. The average error (bottom row in both tables) is actually lower for our own estimates, when $s > 1$.

Table 5: Revisions in output gap: RMSE. Real time data: our own methodology. Final data: OECD.

s=	0	1	2	3	4	Mean
Australia	1.46	1.27	1.20	0.76	0.36	1.01
Austria	1.48	1.60	1.37	1.19	1.10	1.35
Belgium	1.21	1.20	0.76	0.73	0.66	0.91
Canada	2.80	2.03	1.49	0.99	0.54	1.57
Denmark	1.04	0.93	0.50	0.39	0.42	0.66
Finland	7.08	5.47	3.75	2.76	2.09	4.23
France	1.09	0.98	1.12	0.88	0.68	0.95
Germany	1.33	0.95	0.51	0.44	0.64	0.77
Greece	2.99	2.25	1.55	1.20	0.99	1.80
Ireland	4.10	3.31	2.23	1.69	1.37	2.54
Italy	1.37	1.30	1.00	0.90	0.81	1.08
Japan	1.87	1.55	1.17	0.55	0.66	1.16
Netherlands	2.00	1.89	1.76	1.46	1.35	1.69
Norway	2.01	1.94	1.76	1.51	1.34	1.71
Portugal	1.78	1.53	1.23	0.91	0.87	1.26
Spain	2.80	2.32	1.84	1.64	1.57	2.03
Sweden	3.61	2.77	2.16	1.75	1.45	2.35
United Kingdom	1.66	1.11	0.84	0.66	0.47	0.95
United States	1.66	1.04	0.70	0.61	0.70	0.94
Mean	2.28	1.87	1.42	1.11	0.95	1.52

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

In other words, our simple method provides, on average, a better estimate of the OECD's final output gap figures than the OECD's real time figures for vintages $s > 1$. This shows that although the production function method employed by the OECD is able to pick up changes in production inputs (and therefore in potential output) more quickly, the relative advantage in terms of accuracy compared to the HP filter is not very significant (also see Appendix D).

Having calculated the output gap measure, we continue with calculating our own estimates of the cyclically adjusted balance by equation 14. The CAB is found for five consecutive vintages and then compared to OECD's final CAB estimate (OECD *Economic Outlook* 2005). Elasticities of revenues and expenditures, β and ϕ, are the ones used by the OECD and based on weights for 2003. The results are tabulated below in table 6:

Table 6: Revisions in CAB estimates: RMSE. Real time CAB data: our own methodology. Final CAB data: OECD .

s=	0	1	2	3	4	Mean
Australia	1.06	0.79	0.81	0.72	0.37	0.75
Austria	0.94	0.96	0.67	0.61	0.51	0.74
Belgium	1.23	0.88	0.54	0.50	0.57	0.75
Canada	0.92	0.70	0.54	0.39	0.37	0.59
Denmark	1.08	1.22	1.29	1.15	1.00	1.15
Finland	4.54	3.44	2.53	1.91	1.38	2.76
France	0.87	0.59	0.60	0.57	0.49	0.62
Germany	0.79	1.28	0.93	1.11	1.11	1.04
Greece	2.68	2.33	2.42	1.91	1.33	2.13
Ireland	2.90	1.61	0.86	0.54	0.44	1.27
Italy	1.46	0.99	0.71	0.70	0.70	0.91
Japan	1.38	1.19	0.70	0.43	0.62	0.87
Netherlands	1.40	1.17	0.74	0.63	0.54	0.90
Norway	8.73	9.23	8.90	8.87	9.03	8.95
Portugal	1.89	1.46	1.08	0.92	0.67	1.20
Spain	1.54	1.24	1.11	1.24	1.12	1.25
Sweden	2.74	2.31	1.90	1.77	1.52	2.05
United Kingdom	1.03	0.54	0.89	0.84	0.87	0.83
United States	0.82	0.64	0.64	0.54	0.26	0.58
Mean*	1.63	1.30	1.05	0.92	0.77	1.13

* Excluding Norway.
Source: OECD *Economic Outlook* 58-78, authors' own calculations.

The cyclically adjusted budget balance conditional on the data available at time $t+s$, $\bar{b}_{t|t+s}$, shows that in real time (i.e. $s = 0$) the root mean squared error we make across the countries is about 1.6pp, which is not significantly higher compared to the average error in OECD's revisions, which is 1.4. The difference between our and OECD's calculations stays around 0.2 pp across all vintages, i.e. when $s = 0...4$. These results are shown in Appendix E.

In summary, we have used a very simple method to estimate the CAB in real time in order to estimate the robustness of the official estimates to different estimation techniques- as opposed to robustness to new or revised information. First, we employed one of the most standard filtering techniques, the Hodrick-Prescott filter, to extract the output gap series. Second, we do not disaggregate expenditures or revenues in order to take into account differences in cyclical elasticities of their subcomponents. Despite this, our own calculated real time CAB estimate does not indicate a significantly bigger deviation from the OECD's final data than OECD's real time CAB estimates. On average, both departure from the final values by about 1.5 pp and the error decreases to a bit less than 1 pp after adding four more data points. In other words, even four years after the event the error made in CAB estimate remains quite significant.

5. The effectiveness of CABs as an early warning indicator

An important function of CABs is to serve as an early warning 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.

5.1. Can CABs sound the alarm in real time?

To test the effectiveness of this warning, we construct a simple binary measure of fiscal slippage using *ex post* data, and analyse how many times the real-time CAB figure correctly indicated the presence, or absence, of a fiscal loosening. In keeping with the literature[10], we first define a fiscal slippage as a worsening, in the *ex post* CAB, of 1.5 pp of potential GDP over one year. Thus a value of -1.5 pp corresponds to a worsening in the CAB of 1.5 pp of potential GDP.

If the change in the real time CAB is more than some trigger value, then we assume a hypothetical ”alarm" is sounded. If it is lower, then no early warning is registered. Comparing these with the *ex post* data, we can 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.[11]

If we were to mount formal statistical tests at this point, the false alarm and missed alarm outcomes would be associated with type II and type I errors respectively, under the null hypothesis that the CAB deficit was small enough not to trigger an alarm or any need for remedial action. Thus:

[10]Our measure is similar to that of Blanchard, Giavazzy and Pagano (2000).

[11]Note that both 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 vintages. 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.

H_0: Trigger value (T) $<$ the true CAB value, where T ≤ 0 is a some value which is set by the policy makers.

Type I error: we measure the change in the CAB $<$ T with real time data, when the true (ex post) change in the CAB $>$ T. This generates a false alarm.

Type II error: we measure the change in CAB $>$ T, when the true (ex post) change in the CAB $<$ T. This implies a missed alarm.

Table 7: Classifying budgetary outcomes

		Real time data	
		CAB worsens more than 1.5 pp	CAB worsens less than 1.5 pp
Ex post data	CAB worsens more than 1.5 pp	Correct alarm	Missed alarm
	CAB worsens less than 1.5 pp	False alarm	Correct, all clear

A low number of false alarms would indicate a high degree of confidence in our real time CAB deficit numbers. If this is the case, real time data discriminate well in that those countries that are picked out are genuinely problem cases, and few are falsely accused of weak fiscal discipline when in fact their structural deficits are no threat to themselves or others. A high number of missed alarms, on the other hand, are a sign that not all problem cases are being detected. In that case, real time data have very little power: a significant number of problem or emerging cases are being missed that the true or *ex post* data would have identified successfully. Such an imbalance in the likelihood of the errors would pose a risk for the policy makers. Not only does it imply that relying on real time data may be a risky strategy in the context of fiscal surveillance. More controversially, it supplies an opportunity to unscrupulous national policy makers to massage their fiscal projections favourably in order to minimize scrutiny or public criticism of their plans (as Jonung and Larch (2006) claim). An imbalance in the errors of this kind does not supply a political motive as such: but it certainly makes such a strategy much more feasible since the probability of being caught with a bad CAB deficit is a lot lower than the probability of being thought not to have one when in fact you do have on.

To test this empirically, the experiment was conducted for a variety of different trigger values.

The trigger value is initially set at -1.5 pp of potential GDP, and then, as a robustness check, the process is repeated with the trigger set at -1.0 pp and -2.0 pp of potential GDP.

The results are reported below. *Average Revision* is the average revision between real time and final data for the change in the CAB; and *Average change in CAB* is the average *ex post* change in the CAB for all the observations in each category.

Table 8: Success of the CAB as an early warning: slippage of more than 1.5 pp in one year

Trigger		Correct alarm	False alarm	Missed alarm	Correct all clear
	Frequency	4	13	10	163
-1.0 pp	Average revision	-0.38	0.86	-2.09	-0.02
	Actual change in CAB	-1.64	-0.65	-2.16	0.54
	Frequency	3	3	11	173
-1.5 pp	Average revision	-0.33	0.58	-1.03	-0.05
	Actual change in CAB	-1.64	-1.28	-2.11	0.48
	Frequency	2	1	12	175
-2.0 pp	Average revision	-0.23	0.73	-0.96	-0.06
	Actual change in CAB	-1.41	-1.47	-2.11	0.46

Source: OECD *Economic Outlook* 58-78, authors' own calculations

From table 8, the CAB performs very poorly as an early warning indicator. Of the 14 instances of slippage in our sample, only four are picked up when the trigger is -1.5 pp. Making the trigger more sensitive does little to help. Even if it is set at -1 pp, only one extra slippage is picked up. It is also clear that that the bulk of the problem lies with missed alarms than with false alarms. For trigger values of -1 pp and -1.5 pp most of the "wrong verdicts" are missed alarms rather than false alarms. Moreover, the average value of the change in the CAB in the *false alarm* is -1.28 pp and thus fairly close to threshold value of -1.5 pp, implying that we can characterise most of these are "near misses"; on the other hand the corresponding value for the *missed alarms* is -2.11, meaning a number of sizeable slippages go undetected.

As a further robustness check, we repeat the analysis for a second definition of a fiscal slippage — defined as a worsening of the CAB by 2 percentage points, over two years, from $t-2$ to t, for trigger values of -1.5, -2, and -2.5 percent of potential GDP. The results are shown in table 9. This table presents a very similar story. Of the 20 recorded fiscal slippages, we can pick out at best just under one half. There are also a substantial proportion of false alarms, relative to the number of correct alarms.

Table 9: Success of the CAB as an early warning: slippage of more than 2 pp over 2 years

Trigger		Correct alarm	False alarm	Missed alarm	Correct all clear
	Frequency	9	7	11	163
-1.5 pp	Average revision	0.45	-1.792	2.65	-0.15
	Actual change in CAB	-3.22	-0.83	-3.05	1.05
	Frequency	7	3	13	167
-2.0 pp	Average revision	0.40	-1.78	2.33	-0.13
	Actual change in CAB	-3.47	-0.81	-2.94	1.01
	Frequency	4	1	16	169
-2.5 pp	Average revision	0.51	-1.82	1.94	-0.11
	Actual change in CAB	-4.13	-1.75	-2.87	0.99

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

A different approach would be to set the trigger value so as to optimise some criterion. To test the merits of this approach we consider three possible criteria. The first is simply to minimise the number of wrong verdicts (*Min WV*) — thus placing equal weight on false alarms and missed alarms. The second is to minimise a weighted sum of false verdicts (*Min WSWV*), where missed alarms are given twice the weight of false alarms. The third and fourth (*Capture 50%, Capture 75%*) are to set the trigger value so that it identifies at least 50%, and 75% of fiscal slippages respectively. That is to reduce the proportion of type I errors to some specified level. In each case, the optimal value (to one decimal place) of the trigger is identified by trial and error. The results are reported in table 10.

Table 10: Optimising the trigger value

Slippage Definition	Criterion	Trigger	Correct alarm	False alarm	Missed alarm	Correct all clear
	Min WV	-2.2	2	0	12	176
1.5 pp	Min WSWV	-2.2	2	0	12	176
in 1 yr	Capture $\geq$50%	-0.6	7	25	7	151
	Capture $\geq$75%	0.3	11	77	3	99
	Min MV	-2.1	7	1	13	169
2 pp	Min WSWV	-2.1	7	1	13	169
in 2 yrs	Capture 50%	-1.3	10	9	10	161
	Capture 75%	-0.5	15	23	5	147

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

These results make it clear that optimising the trigger value does not improve the usefulness of real time CABs. In the first definition of a slippage, minimising either the simple or weighted sum of wrong verdicts gives a very small number of false alarms, at a cost of many missed alarms. To capture even 50% of slippages, the trigger must be set at -0.6%, but at this value, over 75% of the alarms raised are false ones. To capture 75% the trigger must be set at +0.3%, with only 1 in 8 alarms being correct. In practice it would be difficult for a monitor to establish any credibility with a strategy that "cries wolf" quite so often.

Under the second definition, minimising the sum (simple or weighted) of wrong verdicts still results in over 50% of slippages going undetected. Capturing at least 50% of slippages does yield some slightly more encouraging results, but still 50% of alarms are false ones. To capture 75% of slippages, the trigger has to be set at +0.2%, but at this level well over half of alarms are false.

5.2. Deficit revisions or problems with cyclical adjustment?

Our previous calculations (tables 3 and 6) establish that there are substantial discrepancies in CABs across vintages. These discrepancies come partly from revisions in the deficit ratio, and partly from revisions to the cyclically adjusted component (recall equation (11)). To quantify the relative contributions of these two sources of error, we re-run our previous analysis with the real-time data artificially "corrected" for the effect arising from revisions to the deficit ratio.

Using the same notation as in section 2 and with lower case letters denoting ratios to potential output, we may write the real time CAB as follows:

$$\overline{b}_{t|t} = b_t + \epsilon^{'}_{t|t} - (b_t - \overline{b}_t + \nu_{t|t}), \tag{24}$$

where $b_t + \epsilon^{'}_{t|t}$ is the real time estimate of the actual budget balance ($\epsilon^{'}_{t|t} = \epsilon_{t|t}/Y^*_{t|t}$), $b_t - \overline{b}_t$ is the true cyclical component of the budget deficit and $\nu_{t|t}$ is the error made in estimating the total cyclical component of the budget balance at time t.

The real time change in the CAB is thus given by:

$$\overline{b}_{t|t} - \overline{b}_{t-1|t} = \overline{b}_t + \epsilon^{'}_{t|t} - \nu_{t|t} - \overline{b}_{t-1} - \epsilon^{'}_{t-1|t} + \nu_{t-1|t}. \tag{25}$$

Trivial manipulation of the above two equations yields the following expression for the total error in the real time change in the CAB:

$$\Delta\bar{b}_t - \Delta\bar{b}_{t|t} = \nu_{t|t} - \nu_{t-1|t} - (\epsilon'_{t|t} - \epsilon'_{t-1|t}). \tag{26}$$

We construct an artificially corrected real time data set, where we eliminate any error in the budget deficit estimate. Specifically, we add the term $\epsilon'_{t|t} - \epsilon'_{t-1|t}$ to the real time change in the CAB series. In such case, the only source of discrepancy between the modified real time series and the *ex post* data, will be errors in the cyclical adjustment process.

The intuition of this technique can be roughly characterised by imagining that at the end of every 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. The resultant series can be though of as a hypothetical real time CAB series.

The difference between the final data and the hypothetical series is the error which is purely attributable to problems with the cyclical adjustment process.

Comparing the original real time data with the hypothetical data, any differences must only reflect differences in the data on actual deficits, since the information used for the cyclical adjustment was identical in both cases.

Visual inspection of the scatterplot of the hypothetical versus *ex post* data shows that the hypothetical data is considerably more correlated with the *ex post* data than the original real time data, implying that a significant part of the data revision problem relates to the actual budget deficit figures (see Appendix F). Equally, the fact that the points are not very tightly clustered around the 45 degree line (dotted line) implies that cyclical adjustment still plays a significant role. One can notice that the slopes of the fitted trend lines on scatter plots with the hypothetical real time data (solid lines) are sharper than 45 degree lines (also compared to the graphs with the non-hypothetical real time data), which indicates that the pure cyclical effect has been systematically overestimated in real time.

We then repeat the analysis of the previous section, substituting real time data for our hypothetical data. The results are shown below in table 11.

Table 11 shows that for the first definition of slippage — a fall of 1.5 pp over the space of one year, the hypothetical data is better than the real time data. Setting the trigger at -1.5 captures all slippages but produces four false alarms. The trigger value at -1.0 pp also captures all slippages, albeit at the cost of 19 false alarms.

For the second definition — a fall of more than 2 pp over two years, the hypothetical data performs better, but not by a great margin. Comparing table 11 with tables 8 and 9, we see that for all trigger values the number of correct alarms rises, and consequently the number of missed alarms falls. Whilst this means that most of the slippages can now be detected, we also see that the with the hypothetical data, the number of false alarms actually rises.

Table 11: Hypothetical versus *ex post* data

Slippage of more than 1.5 pp in one year

Trigger		Correct Alarm	False Alarm	Missed Alarm	Correct All Clear
	Frequency	14	19	0	157
-1.0 pp	Average Revision	1.61	0.52	-	-0.14
	Average CAB	-2.21	-1.11	-	0.58
	Frequency	14	4	0	172
-1.5 pp	Average Revision	1.61	0.06	-	-0.07
	Average CAB	-2.21	-1.01	-	0.46
	Frequency	10	2	4	174
-2.0 pp	Average Revision	1.66	0.29	1.49	-0.07
	Average CAB	-2.42	-0.99	-1.70	0.53

Slippage of more than 2.0 pp in two years

Trigger		Correct Alarm	False Alarm	Missed Alarm	Correct All Clear
	Frequency	13	10	1	164
-1.5 pp	Average Revision	1.56	0.72	2.29	-0.12
	Average CAB	-2.25	-0.87	-1.69	0.65
	Frequency	12	5	2	170
-2.0 pp	Average Revision	1.64	0.84	1.42	-0.09
	Average CAB	-2.31	-0.90	-1.64	0.59
	Frequency	8	2	6	174
-2.5 pp	Average Revision	1.77	0.29	1.40	-0.07
	Average CAB	-2.54	-0.99	-1.77	0.55

Source: OECD *Economic Outlook* 58-78, authors' own calculations.

Comparing these results with those obtained using real time data, it is clear that errors in budget deficit figures must shoulder a significant amount of the blame for the poor performance of real time CAB figures. Nevertheless, the hypothetical figures still perform relatively poorly, indicating that significant problems also exist on the cyclical adjustment side. The relatively better now-casting power and the smaller number of missed alarms of the hypothetical real time CAB is largely due to the overestimation of the cyclical position at the time. Although being biased in the "right" direction, it is an important source of impreciseness.

5.3. Data revisions and fiscal slippages: is there a link?

The analysis presented so far in this section suggests that there may be an important asymmetry in data revisions — namely that the fit of the model tends to be poorer, the worse is the *ex post* CAB. We now explore this property in more depth. Specifically, we consider how the error associated with real time figures, defined as the difference between real time and *ex post figures*, varies with the level and the change in the *ex post* CAB figure. These asymmetries are potentially important, because from the point of view of fiscal surveillance, the authorities will be particularly interested, inter alia, in those countries who have CABs which 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 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.

First we regress the error ,$\upsilon_{t|t} = \bar{b}_{t|t} - \bar{b}_{t|f}$, the square of the error and the absolute value of error on the level and change in the *ex post* CAB. To control for possible inertia in error we also optionally add a lagged value of the dependent variable. Estimation results for $\upsilon_{t|t}$ show that there is a tendency to overestimate (underestimate) CAB if the true value of it has recently decreased (increased). Also there is a little evidence on CAB being overestimated (underestimated) the bigger the structural deficit (surplus) known *ex post*. These two findings taken together imply that the extreme cases are not that easily detected. This is also supported by the results of regressions for the variance of error, $\upsilon^2_{t|t}$ and $|\upsilon_{t|t}|$, which indicate that *ex post* worsening of the CAB ($\Delta\bar{b}_{t|f} < 0$) increases the impreciseness on the real time CAB estimate (see table 12).

Secondly we try to pick up the possible asymmetries by estimating a pooled regression with the fixed cross-country effects and two dummy variables, which describe four states of CAB — negative and falling ($\bar{b}_{t|f}{}^{-,\downarrow}$), negative and rising ($\bar{b}_{t|f}{}^{-,\uparrow}$), positive and falling ($\bar{b}_{t|f}{}^{+,\downarrow}$) and positive and rising (the effect is captured by the constant term). The experiment yields the very similar results to the first set of regressions. The CAB tends to be overestimated if the true CAB has been in deficit and falling. We also find that the magnitude of error is higher if the true CAB is positive and rising but also when it is negative and falling. As all specifications indicate high persistence in the error related to the real time CAB estimates, the monitor is likely to carry any misjudgments over several periods.

Table 12: Errors in the real time CAB

	Dependent Variable					
	$\upsilon_{t\|t}$	$\upsilon_{t\|t}$	$\upsilon^2_{t\|t}$	$\upsilon^2_{t\|t}$	$\|\upsilon_{t\|t}\|$	$\|\upsilon_{t\|t}\|$
			Regression I			
constant	0.261***	0.282***	1.817***	0.978***	0.972***	0.464***
$\overline{b}_{t\|f}$	-0.159***	-0.049	-0.289*	-0.189	-0.046	-0.024
$\Delta\overline{b}_{t\|f}$	-0.473***	-0.478***	-0.636	-0.769**	-0.135*	-0.147*
$\upsilon_{t-1\|t-1}$		0.466***				
$\upsilon^2_{t-1\|t-1}$				0.424***		
$\|\upsilon_{t-1\|t-1}\|$						0.471***
$\overline{R}^2$	0.316	0.549	0.060	0.286	0.038	0.286
Nob	209	190	209	190	209	190
			Regression II (fixed cross-country effects)			
constant	-0.174	-0.101	1.913***	1.210***	0.857***	0.574***
$\overline{b}_{t\|f}^{+,\downarrow}$	0.980***	0.986***	-0.554	0.157	-0.004	0.125
$\overline{b}_{t\|f}^{-,\uparrow}$	0.266	-0.025	-0.562	-0.932	0.022	-0.138**
$\overline{b}_{t\|f}^{-,\downarrow}$	1.284***	0.995***	1.782**	1.881**	0.550***	0.469***
$\upsilon_{t-1\|t-1}$		0.372***				
$\upsilon^2_{t-1\|t-1}$				0.273***		
$\|\upsilon_{t-1\|t-1}\|$						0.289***
$\overline{R}^2$	0.298	0.443	0.191	0.32	0.207	0.320
Nob	198	180	198	180	198	180

Notes: $\overline{b}_{t|f}^{+,\downarrow} = \Phi_1(1-\Phi_2)$, $\overline{b}_{t|f}^{-,\uparrow} = \Phi_2(1-\Phi_1)$ and $\overline{b}_{t|f}^{-,\downarrow} = \Phi_1\Phi_2$, where Φ_1 = 1 if $\Delta\overline{b}_{t|f} < 0$, 0 otherwise; Φ_2 = 1 if $\overline{b}_{t|f} < 0$, 0 otherwise. *, ** and *** stand for statistical significance at 10%, 5% and 1% significance level respectively. $\overline{R}^2$ denotes the adjusted fit of the model. Δ is the difference operator. Source: OECD *Economic Outlook* 58-78, authors own calculations.

Taken together these results imply that real time CAB figures tend to be more reliable for countries with healthy and stable public finances, and relatively poorer for countries with worse or changing public finance situations- and particularly when public finances are in deficit. In the context of fiscal monitoring, this means that real time CAB figures are less reliable in precisely the circumstances in which they are most needed as an "early warning" mechanism.

6. Conclusions

This paper is based on the observation that the CAB can only function as an "early warning" of future fiscal problems if the CAB can be reliably estimated in real time. Accordingly, we set out to analyse the properties of real time CAB data in relation to available *ex post* data.

The descriptive statistics of section 3 showed clearly that both output gap and government deficit figures are revised substantially for many years following their official publication. Section four showed that the real time CAB figures deviate substantially from the *ex post* data, to the extent that a simple HP filter based method of cyclical adjustment predicts the final OECD CAB figure nearly as well as the real time OECD CAB figures do.

Analysing the ability of CABs to sound the alarm in real time when they pass some trigger value is fraught with difficulty. Our analysis in section 5, showed that in order to capture a reasonable proportion of fiscal slippages the trigger value must be set in such a way that the majority of "alarms" turn out to be false alarms. The experience of the first incarnation of the Stability and Growth Pact suggests it is hard enough to enforce sanctions when the transgressor has verifiably violated a clear numerical rule. Attempting to punish (or even threatening to punish) a government on the basis of an unreliable real time figure will be even tougher, since the government in question can claim that it is too early to convincingly verify whether or not the threshold has been crossed. But if the monitor waits for several years more additional data in order to verify the true situation, then the whole "early warning" rationale for using the CAB will disappear.

The experiments with hypothetical data suggest that if the monitor were given the final deficit data, but had to cyclically adjust it on the basis of real time figures, a significant proportion of wrong errors would be eliminated. This implies that revisions in the deficit figures may account for a large part of the errors in real time data, rather than the cyclical adjustment itself. Thus, better real time data on actual deficits would considerably improve the estimates of real time cyclically adjusted budget deficits.

Furthermore, the econometric analysis of section 5 shows that the performance of real time figures is not consistent across economic circumstances. In particular they tend to be more unreliable for countries for countries who have more loose and volatile fiscal policies, that is, they tend to be the least reliable when they are needed most.

References

Beetsma, R., Giuliodori, M., 2006. On the Relationship between Fiscal Plans in the European Union: An Empirical Analysis Based on Real-Time Data. Mimeo, University of Amsterdam.

Berger, H., De Haan, J., Jansen, D.J., 2003. The End of the Stability and Growth Pact? *CESifo Working Papers*, 1093.

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

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

European Commission, 2002. Public Finances in EMU. European Economy, 3.

European Commission, 2006. Public Finances in EMU. European Economy, 3.

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

González-Mínguez, J.M., Hérnandez de Cos, P., Del Río, A., 2003. An Analysis of the Impact of GDP Revisions on Cyclically Adjusted Budget Balances. *Documento Ocasional, Banco De España*, 0309.

Guay, A., St-Amant, P., 1997. Do the Hodrick-Prescott and Baxter-King Filters Provide a Good Approximation of Business Cycles? *CREFE Working Paper*, 53.

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

Jonung, L., Larch, M., 2006. Fiscal Policy in the EU: Are the Official Output Forecasts Biased? *Economic Policy*.

OECD Economic Outlook, 1995–2005. 58–78.

Kattai, R., Vahter, P., Potential Output and Output Gap in Estonia. *Working Papers of Eesti Pank*. Forthcoming.

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

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

Wierts, P., Deroose, S., Flores, E., Turrini, A., 2006. Fiscal Policy Surveillance in Europe. McMillan Palgrave.

Appendix A. The cyclically adjusted budget balance of the Netherlands across vintages

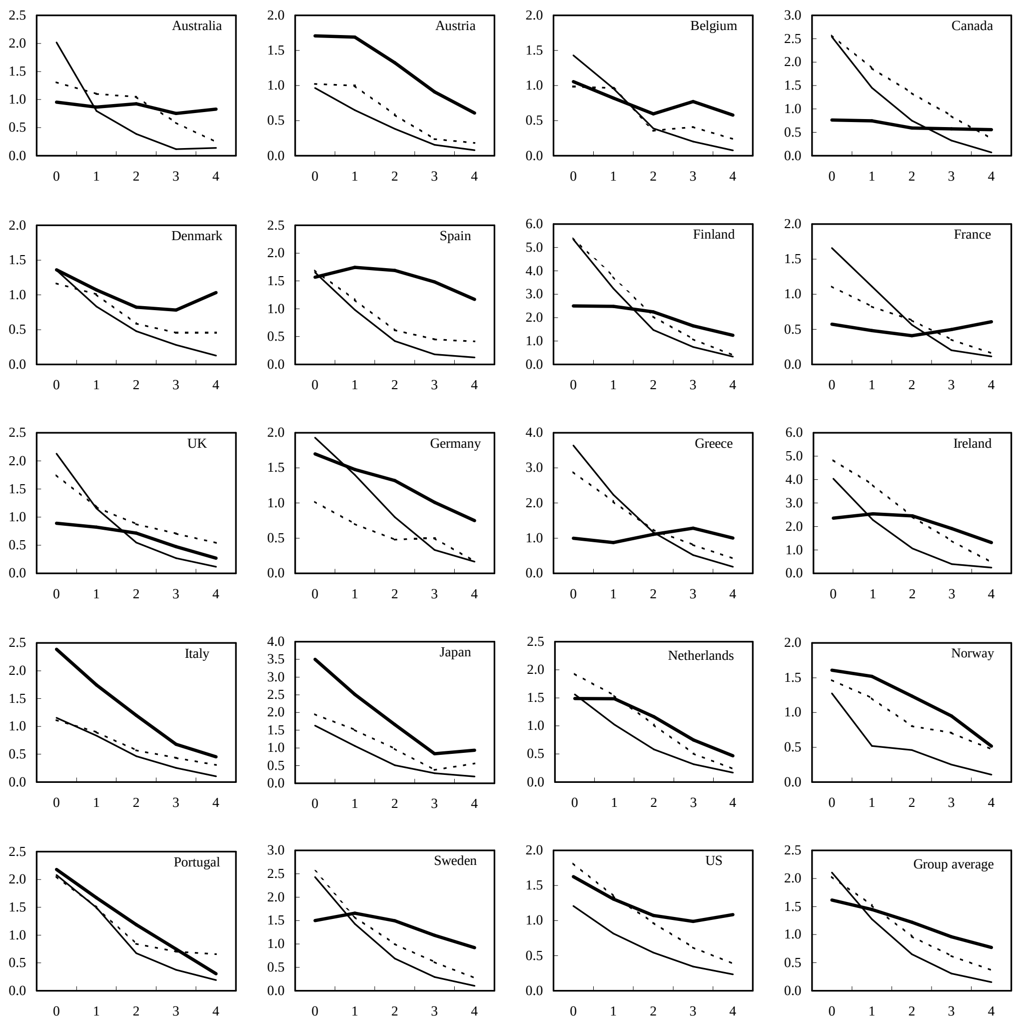
Year	1999	2000	2001	2002	2003
Estimated CAB, spring 2004	-1.4	-1.0	-1.8	-2.4	-1.3
Estimated CAB, autumn 2001	-0.6	0.1	0.2	0.8	0.8

Appendix B. Summary statistics of budgets cyclical components

	Mean	Root Mean squares	Standard deviation	Min	Max
Australia	-0.57	1.00	0.84	-2.45	0.54
Austria	-0.11	0.77	0.78	-1.27	1.71
Belgium	-0.68	1.16	0.97	-2.23	1.04
Canada	-0.33	0.94	0.90	-2.07	1.19
Denmark	-0.06	0.93	0.95	-2.06	1.49
Finland	-1.02	2.61	2.46	-6.51	2.52
France	-0.59	0.91	0.71	-1.71	0.61
Germany	-0.12	1.14	1.17	-1.26	3.35
Greece	-0.31	0.99	0.97	-1.86	1.03
Ireland	-0.56	1.57	1.50	-2.79	1.60
Italy	-0.50	1.03	0.93	-2.06	1.65
Japan	-0.14	0.65	0.65	-1.21	1.05
Netherlands	0.26	1.18	1.18	-1.91	2.50
Norway	7.69	9.02	4.83	1.63	18.30
Portugal	-0.49	1.63	1.59	-4.04	1.71
Spain	-0.55	1.39	1.31	-2.37	1.53
Sweden	-0.83	1.86	1.70	-5.13	1.15
United Kingdom	0.04	1.04	1.07	-1.62	2.67
United States	-0.21	0.46	0.41	-0.77	0.51
Mean	0.05	1.59	1.31	-2.19	2.43

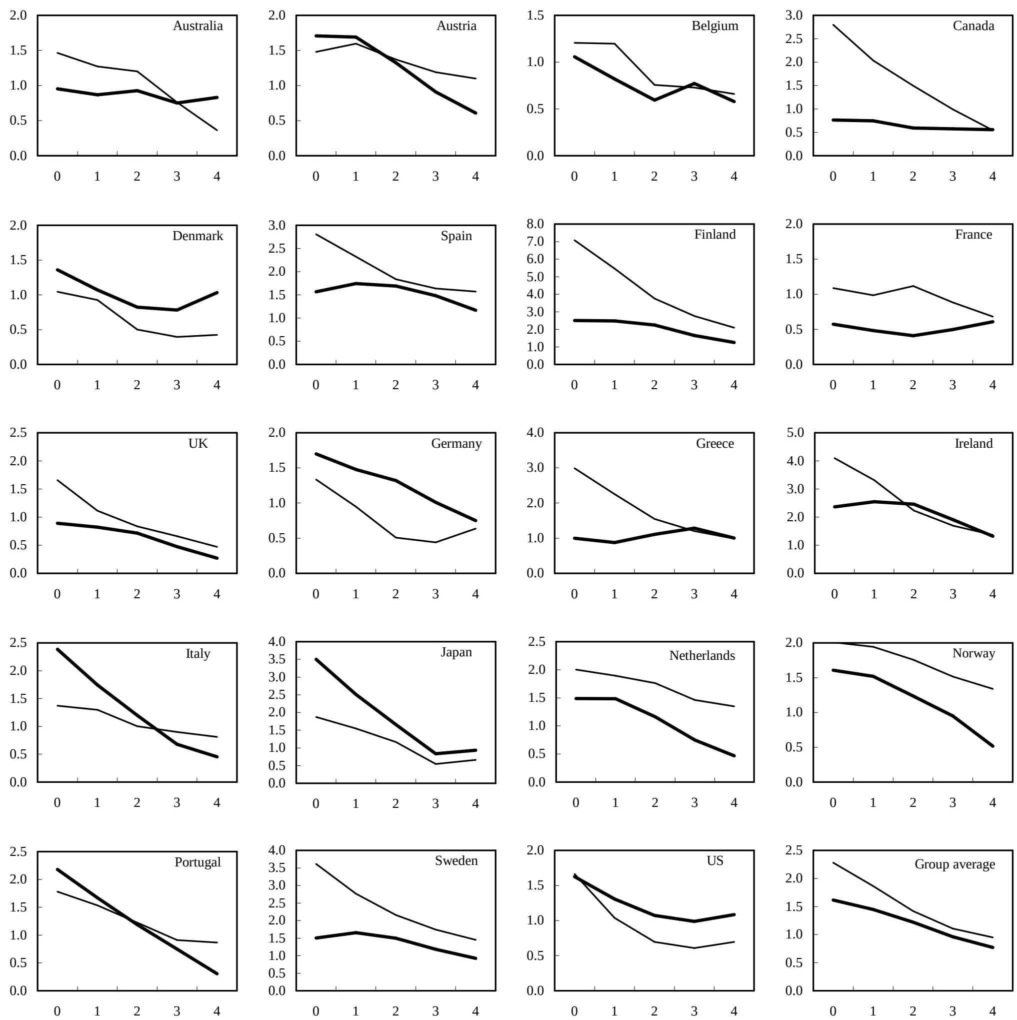
Source: OECD *Economic Outlook* 78, authors own calculations. Budget's cyclical component is given as a difference between actual budget balance and cyclically adjusted balance, represented in per cent of GDP. Therefore positive mean value implies that on average country's cyclically adjusted budget balance has been in surplus and *vice versa*, negative mean value refers to a structural deficit. Summary statistics are based on final data estimates 1985–2005 (*Economic Outlook* 2005).

Appendix C. Revisions in output gap across vintages I



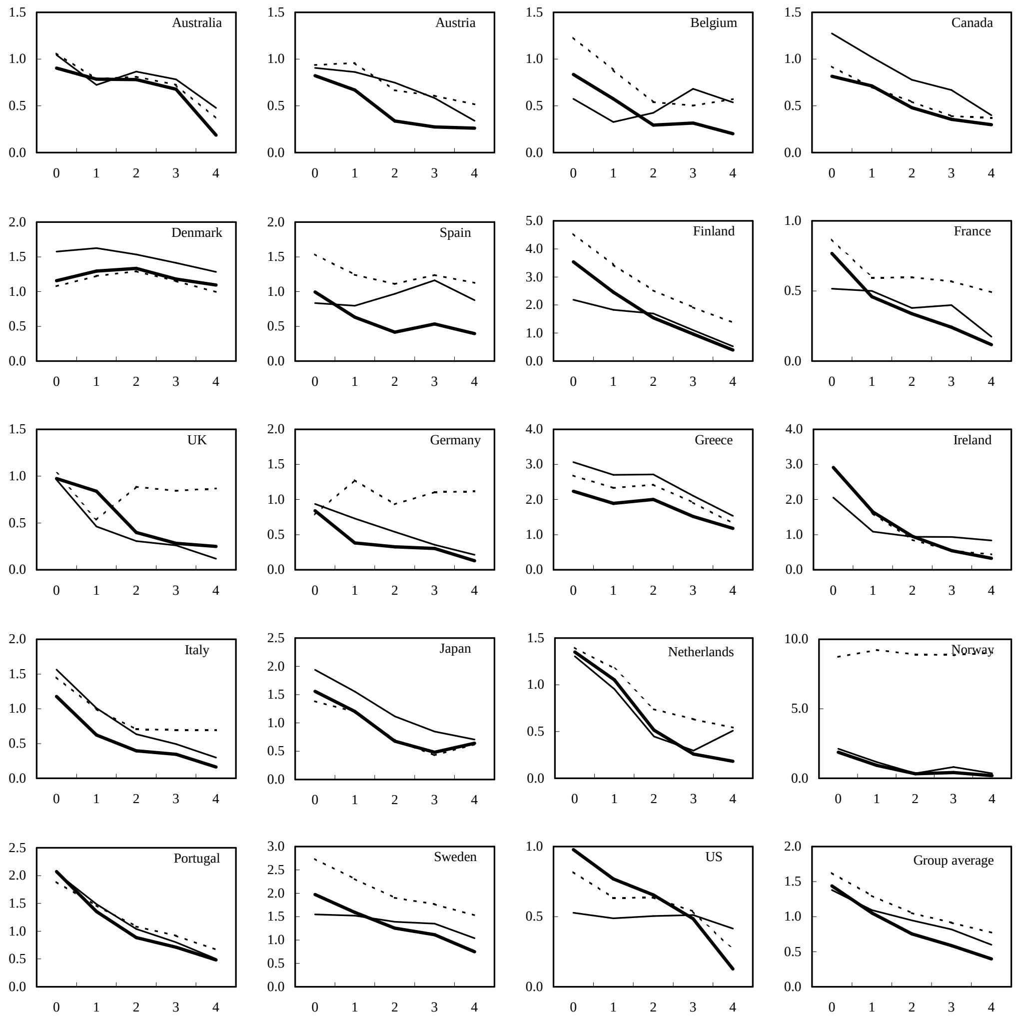
Legend: bold solid line — revisions (RMSE) in OECD data; thin solid line — revisions (RMSE) in own calculations based on final data (issued in OECD *Economic Outlook* 2005); thin dotted line — revisions (RMSE) in own calculations based on real time data. The difference between the thin solid line and the dotted line indicates the effect of year-to-year output revisions on the output gap estimate in addition to the impreciseness in estimates that is caused by the end-point problem.

Appendix D: Revisions in Output Gap Across Vintages II



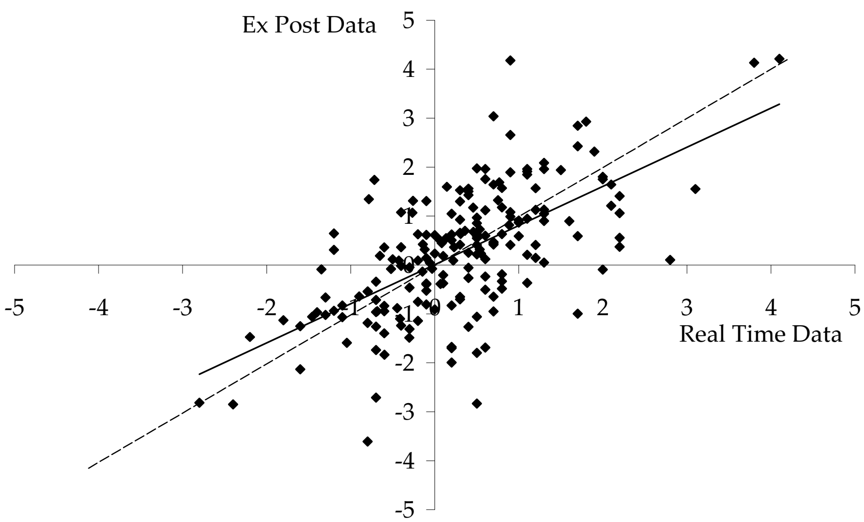
Legend: bold solid line — revisions (RMSE) in OECD data; thin solid line — revisions (RMSE) in own calculations based on real time data but final gap estimates are OECD's.

Appendix E. Revisions in cyclically adjusted balance across vintages



Legend: bold solid line — revisions (RMSE) in OECD data; thin solid line — revisions (RMSE) in own calculations; thin dotted line — revisions (RMSE) in own calculations but final CAB data is OECD's. Group average excludes Norway.

Appendix F. Real time CAB *vs ex post* CAB



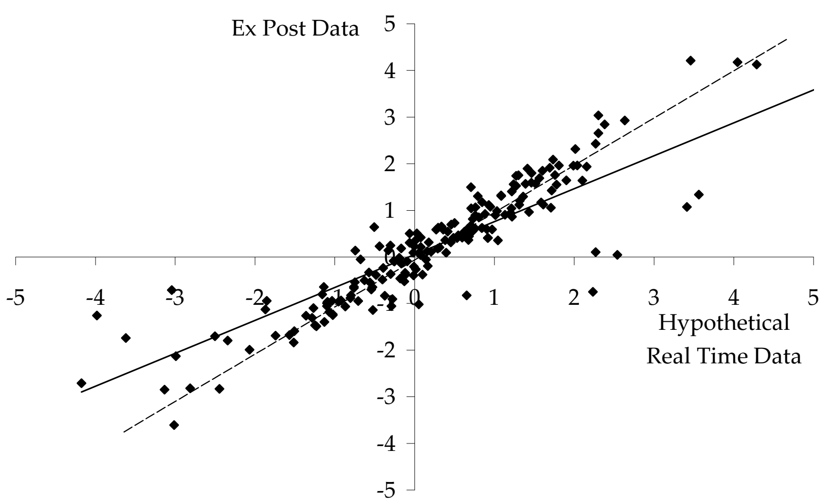


Figure 1: Yearly change in cyclically adjusted budget balance.
Source: OECD *Economic Outlook* 58-78, authors' own calculations.

Regression equation for real time data *(t-statistics in brackets)*:

$$\Delta CAB_{expost} = \underset{(0.06)}{0.07} + \underset{(11.27)}{0.80} * \Delta CAB_{realtime} \quad R^2=0.635$$

Regression equation for hypothetical data:

$$\Delta CAB_{expost} = \underset{(1.04)}{0.05} + \underset{(22.37)}{0.71} * \Delta CAB_{Hypothetical} \quad R^2=0.732$$

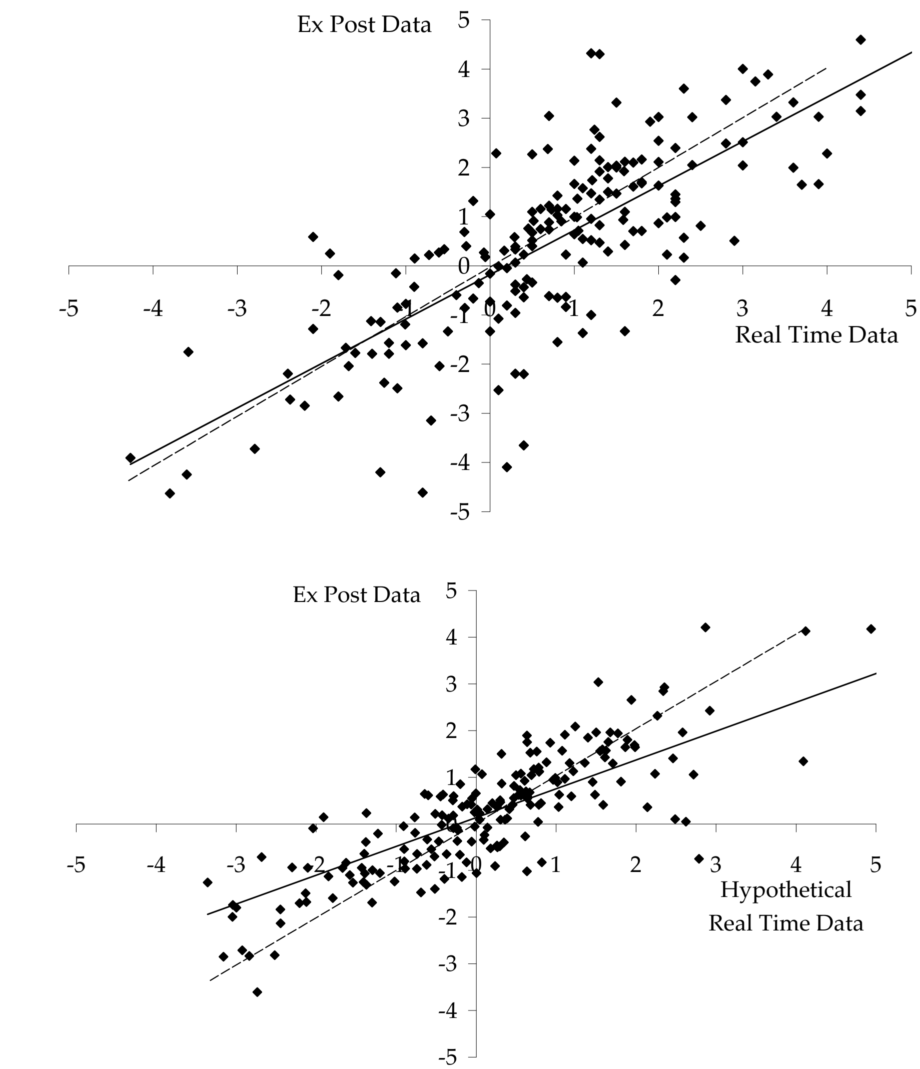


Figure 2: Two year change in cyclically adjusted budget balance.
Source: OECD *Economic Outlook* 58-78, authors' own calculations.

Regression equation for real time data *(t-statistics in brackets)*:

$\Delta CAB_t - CAB_{t-2} =$ -0.18 + 0.90*($CAB_{t|t} - CAB_{t-2|t}$ R^2=0.620

(-1.85) *(17.50)*

Regression equation for hypothetical data:

$CAB_{t|t} - CABt - 2|t =$ 0.13 + 0.617*($CAB_{t|Hypothetical} - CAB_{t-2|Hypothetical}$)

(2.207) *(17.187)*

R^2=0.611

Working Papers of Eesti Pank 2007

No 1
Rasmus Kattai
Constants do not stay constant because variables are varying