Ensaio

The Forecast Quality of Portuguese GDP Expenditure Components[1]

Pedro M. Esperança[2], João C. Fonseca[3], Paulo Júlio[4]

1. Introduction

The accuracy of macroeconomic forecasts has been subject to a vast scrutiny. The majority of the studies compare the quality of macroeconomic forecasts across different organizations and/or time spans. This analysis is usually undertaken for real Gross Domestic Product (GDP) or other macroeconomic variables, but not for the major expenditure components of GDP – namely, private consumption (C), public consumption (G), investment[5] (I), exports (X) and imports (M)[6]. The few exceptions include Ash et al. (1998), who evaluate the quality of OECD's forecasts for GDP components using a directional analysis approach, and Timmermann (2007), who explores IMF's forecasts for the current account for several world regions, but does not address forecasts for other GDP components.

This article focuses simultaneously on two dimensions of forecast quality (bias and dispersion) through three different perspectives: across institutions, across time spans and, most importantly, across GDP components. We use forecast data issued for Portugal by five different national and international institutions – Organization for Economic Co-operation and Development (OECD), International Monetary Fund (IMF), European Commission (EC), *Banco de Portugal* (BdP) and Portuguese Government Budget Office (GBO) – for four different time spans – labeled 18-month, 12-month, 6-month and 0-month. Our analysis, covering the 2002-2010 period, uses a scaled statistic which takes into account the inherent levels of volatility of each GDP component, and explores the contributions of expenditure components to the GDP forecast error. The scaled statistic suggests that investment is the hardest component to predict at longer horizons (1-year ahead predictions), and public consumption at shorter horizons (same-year predictions). Optimistic GDP forecasts at longer horizons result from overly optimistic forecasts for investment and exports. At shorter horizons, GDP forecasts are closer to actual values, but this is achieved with large deviations in components' predictions, which tend to cancel out. We propose a new statistic – termed Mean of Total Weighted Absolute Error (MTWAE) – to summarize the quality of forecasts across components for each institution and time span, thus evaluating whether accurate GDP predictions are obtained through more or less accurate components' predictions. This statistic suggests that, at 18-, 6- and 0-month spans, OECD issues the least reliable forecasts, even though its GDP forecasts are, on average, very accurate at shorter horizons.

The second half of the twentieth century witnessed a major revolution on economic forecasting with the appearance of formal economy-wide models and sophisticated econometric techniques (Wallis, 1989). Equivalent advances in evaluation methods followed and a number of important contributions to the topic were made during the 50s and 60s (Theil, 1958, 1966; Zarnowitz, 1967; Mincer and Zarnowitz, 1969). By the end of this period researchers stressed the importance of evaluating the accuracy of the forecasts being issued (Cairncross, 1969; Moore, 1969) and in the subsequent two decades the accuracy of macroeconomic forecasts originating from both public and private institutions was subject to a close

1 The opinions expressed in this article represent the views of the authors and do not necessarily correspond to those of the Portuguese Ministry of Economy and Employment.

2 NOVA School of Business and Economics. This article was written while the author was visiting at the *Gabinete de Estratégia e Estudos*, Portuguese Ministry of Economy and Employment.

3 *Gabinete de Estratégia e Estudos*, Portuguese Ministry of Economy and Employment. Corresponding author: joao.fonseca@gee.min-economia.pt

4 *Gabinete de Estratégia e Estudos*, Portuguese Ministry of Economy and Employment.

5 More specifically, gross fixed capital formation.

6 In this article, we always refer to real growth rates, even if not explicitly stated.

inspection – see for instance Stekler (1972, 1987), McNees (1976, 1978, 1986, 1988), Zarnowitz (1979, 1984), Holden and Peel (1985, 1990), Clemen and Winkler (1986), Nordhaus (1987) and Joutz (1988).

The literature has kept growing in recent years. For instance, Fildes and Stekler (2002) have conducted a survey on the state of macroeconomic forecasting focusing their analysis on studies made for the United States and the United Kingdom. Öller and Barot (2000) analyze OECD and national institutions' forecasts for GDP growth and inflation for 13 European countries (Portugal not included) and conclude that: (i) OECD and national institutions' forecasts are not significantly different in predictive quality; (ii) both produce efficient forecasts, although they tend to overestimate at longer horizons; (iii) there is an inverse relationship between accuracy and the forecast horizon; (iv) at 1-year horizon, growth forecasts perform better than a same-change alternative; and (v) in general, GDP forecasts have not improved consistently over time. Pons (2000) compares OECD and IMF's GDP growth forecasts for G7 countries and finds OECD's forecasts to be superior to those issued by the IMF. However, the author does not detect a consistent pattern of over or underestimation. Loungani (2001) compares *Consensus* to OECD, IMF and World Bank's forecasts for GDP growth for 63 countries, including Portugal, and concludes that these display very similar degrees of accuracy. Similar results are also found by Melander et al. (2007), for *Consensus*, EC, IMF and OECD's forecasts. Vuchelen and Gutierrez (2005) analyze OECD's GDP growth forecasts for 21 European countries, including Portugal, and show that, although evaluation statistics may suggest valueless forecasts, they occasionally contain some information and perform better than the same-change extrapolation at 1-year horizon.

The aforementioned literature focuses predominantly on GDP growth forecasts, while neglecting how these forecasts are assembled. In general, GDP forecasts issued by institutions result from adding up the contributions from the corresponding expenditure components, for which analyzing the forecast accuracy through this perspective may enable one to identify the major flaws in forecast models, and shed some additional light on the quality of GDP forecasts. *Ceteris paribus*, GDP forecasts which are obtained with smaller average errors in the corresponding expenditure components should be more reliable than those presenting larger average errors.

This article is organized as follows. The next section introduces our statistical methodology. Section 3 describes the data. Section 4 displays the results and conducts the respective analysis. Section 5 concludes.

2. Methodology

2.1. Notation

Let $F_t(s)$ represent the *s*-period ahead forecast for the target variable A_t that is, $F_t(s)$ is the forecast for year t produced s months in advance, where t is the *forecast period* (the period for which we are producing a forecast) and s is the *forecast horizon* or *time span* (the number of months between the production of the forecast $F_t(s)$ and the actual realization of A_t). Let $e_t(s)$ be the corresponding forecast error, *i.e.*, the difference between actual and forecasted values

$$e_t\,(s) = A_t - F_t(s) \qquad (1)$$

for t, $s \in \mathbb{N}_0$. From (1) it is clear that a positive forecast error implies an underestimation, whereas a negative error implies an overestimation, of A_t. Henceforth the forecast horizon s will be suppressed for notational convenience, if not strictly needed.

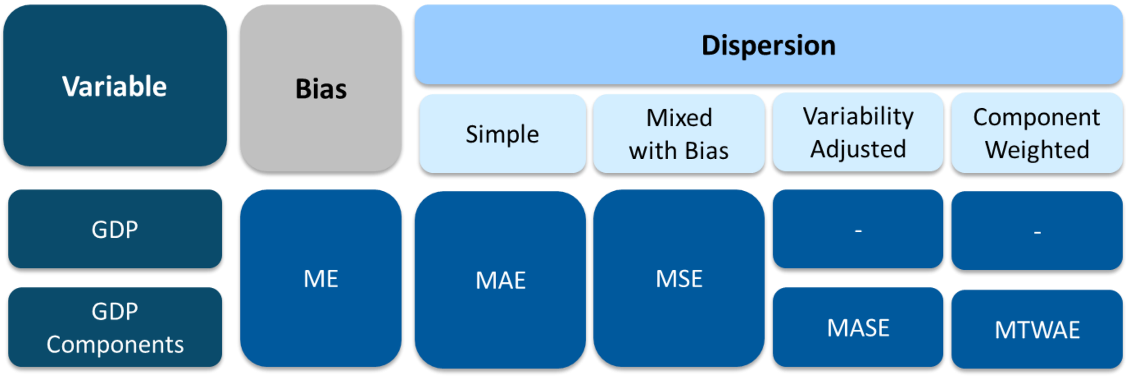
2.2. Quality

We evaluate the forecast quality of GDP and of the corresponding expenditure components across two dimensions: bias and dispersion (Figure 1).

Figure 1 – Dimensions of Quality

Bias measures the average deviation of forecasts, F_t, from actual values, A_t, whereas dispersion measures how off-target forecasts are (*i.e.,* how distant forecasts are from actual values), on average. The following figure identifies the statistics used herein to evaluate these two dimensions.

Figure 2 – Statistics by Quality Dimension



Bias is measured through the so-called Mean Error (ME)

$$ME := \frac{1}{T}\sum_{t=1}^{T} e_t$$

where T is the sample size. The ME is the average forecast error across the sample period, thus providing a simple measure of central tendency. A negative value suggests that forecasts tend to be overly optimistic, whereas a positive value points towards pessimistic forecasts.

Dispersion is evaluated according to four distinct statistics, each calibrated to capture a specific feature. The Mean Absolute Error (MAE)

$$MAE := \frac{1}{T}\sum_{t=1}^{T} |e_t|$$

provides a measure of the average total forecast error, regardless of the direction of the error. Hence, a lower MAE reflects more accurate forecasts. The Mean Squared Error (MSE)

$$MSE := \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

also provides a measure of average total forecast error, but attributes disproportionally higher contributions to larger deviations from target. The MSE can be decomposed into a bias-component and a variance-component

$$MSE := ME^2 + Var$$

where Var denotes the variance of the forecast errors. The decomposition of MSE shows in squared percentage points (p.p.) the source of a forecast's poor performance: bias or dispersion. For the same MSE one can have low bias and high dispersion or vice-versa. For this reason, MSE is a composite measure of forecast quality.

The previous dispersion statistics are only valid when comparing a given variable's forecasts coming from different institutions or forecasting methods – one of the dimensions of our analysis. If one aims to compare the quality of an institution's forecasts across a group of variables – another dimension that we investigate here – these statistics are inadequate, as they do not take into account the intrinsic level of volatility of each series. A more volatile series is naturally harder to predict and thus forecast errors tend to be larger; however, this does not necessarily mean that the forecast model is performing worse. A comparative statistic which addresses this issue can be obtained by scaling each series' errors with the inverse of the corresponding in-sample average absolute difference between the actuals of consecutive periods (a measure of volatility). This yields the Mean Absolute Scaled Error (MASE)

$$MASE := \frac{1}{T}\sum_{t=1}^{T}\left|\frac{e_t}{\frac{1}{J}\sum_{j=1}^{J}|A_j - A_{j-1}|}\right| = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{e_t}{V}\right|$$

where J is the sample size. Table 1, which presents the volatility of each series as measured by V, shows that investment, exports and imports are much more volatile than the remaining series. For these variables one should naturally expect larger errors in the corresponding forecasts. Thus, scaled statistics should be used to evaluate a model's quality in predicting the different components of GDP.

Table 1 – Volatility as measured by V.

	Volatility	Volatility relative to GDP
Gross Domestic Product	1.84	-
Private Consumption	1.49	0.81
Public Consumption	1.96	1.06
Investment	4.48	2.44
Exports	6.78	3.69
Imports	5.91	3.22

It is also possible to decompose the GDP forecast error into the individual contributions of the corresponding expenditure components. This exercise enables one to identify which components contribute the most to the GDP forecast error, and whether the errors in forecasted expenditure components tend to add up or to cancel out, to determine the GDP forecast. Let z_t denote the effective real growth rate of variable Z at year t, and z_t^f the corresponding forecasted real growth rate, $Z = GDP, C, G, I, X, M$; and define $w_t^Z = Z_t/GDP_t$ – the variable Z's share on GDP at t. The effective real GDP growth rate can therefore be decomposed into the corresponding contributions from expenditure components

$$gdp_t \equiv c_t w_{t-1}^C + g_t w_{t-1}^G + i_t w_{t-1}^I + x_t w_{t-1}^X - m_t w_{t-1}^M \qquad (2)$$

With forecasted values, a similar version of equation (2) does not hold, since neither the weights used by institutions nor the base year for those weights are known. Instead, we use effective weights, and consequently an additional discrepancy term, ϵ_t, has to be included

$$gdp_t^f \equiv c_t^f w_{t-1}^C + g_t^f w_{t-1}^G + i_t^f w_{t-1}^I + x_t^f w_{t-1}^X - m_t^f w_{t-1}^M + \epsilon_t \qquad (3)$$

Since effective weights are generally close to those used by institutions, ϵ_t takes small values. Let e_t^Z denote the forecast error of variable Z's growth rate, *i.e.* $e_t^Z = z_t - z_t^f$. Subtracting (3) from (2) and taking the average yields

$$\frac{1}{T}\sum_{t=1}^{T}(e_t^{gdp} + \epsilon_t) \equiv \frac{1}{T}\sum_{t=1}^{T}(e_t^c w_{t-1}^C + e_t^g w_{t-1}^G + e_t^i w_{t-1}^I + e_t^x w_{t-1}^X - e_t^m w_{t-1}^M) \qquad (4)$$

In equation (4), $T^{-1}\sum_{t=1}^{T} e_t^z w_{t-1}^Z$ represents the average contribution of the forecast error arising from variable Z, in p.p., to the GDP growth forecast error. Hence, a negative value means that the component is, on average, overestimated, and systematically contributes to overly optimistic GDP forecasts, whereas a positive value has the opposite interpretation. As it is clear from (4), even if GDP forecast errors are small, this can be achieved with large forecast errors in the respective GDP components, due to a "cancel-out effect."

For this reason, we propose an additional measure of forecast quality, which evaluates the sum across components of the absolute distance between forecasted and actual contributions. We term this new statistic *Mean of Total Weighted Absolute Error* (MTWAE), since it reflects the mean of the sum across components of absolute errors, weighted by the corresponding shares on GDP. Letting

$$\boldsymbol{e}_t = (e_t^c, e_t^g, e_t^i, e_t^x, e_t^m)' \qquad \text{and} \qquad \boldsymbol{w}_t = (w_t^C, w_t^G, w_t^I, w_t^X, w_t^M)'$$

denote the vector of forecast errors and the vector with the corresponding component shares on GDP, MTWAE can be defined as

$$MTWAE \coloneqq \frac{1}{T}\sum_{t=1}^{T}|\boldsymbol{e}_t|'\boldsymbol{w}_{t-1}$$

where $|\boldsymbol{e}_t|$ is a vector whose entries are the absolute values of the entries in $\boldsymbol{e}_t$. This statistic is computed for every institution and forecast horizon. Those institutions with higher values in MTWAE achieve a given GDP forecast with higher absolute forecast errors across components, even if these errors cancel out. Thus, the lower is the MTWAE, the higher is the quality of institutions' predictions in general, *ceteris paribus*. Naturally, the MTWAE statistic can be decomposed in the respective components' contributions, $T^{-1}\sum_{t=1}^{T}|e_t^z|w_{t-1}^Z$.

3. Data

Our dataset contains information on forecasts for Gross Domestic Product, Private Consumption, Public Consumption, Investment (namely gross fixed capital formation), Exports and Imports (all in volume percent change), issued for the 2002-2010 period. Forecasts from five institutions are analyzed: OECD, IMF, EC, BdP and the Portuguese GBO. Actual values and GDP expenditure component shares were taken from Statistics Portugal. Forecasts were aggregated into four categories, according to the issue date, as summarized in Table 2. Notice that, although institutions' forecasts are not issued exactly in the same month, comparing the forecast accuracy across institutions requires them to be classified according to the semester in which they are issued. For convenience, these forecasts are labeled 18-, 12-, 6- and 0-month ahead forecasts. Hence, 18-month (6-month) forecasts are those made in the first semester of the previous (same) year, and 12-month (0-month) forecasts are those made on the second semester of the previous (same) year. As such, some caution is required when comparing forecasts across institutions, since one institution may have used updated information that was not available to other institutions at the time they issued their forecasts. This is particularly relevant for the forecasts issued by BdP: since these are issued later, they use one additional quarter of information relative to other institutions. The forecasts from GBO are only available at 12-month and IMF does not publish forecasts for Portugal's GDP expenditure components.

Table 2 – Forecast horizon and issue date

Forecast period	Forecast horizon	Issue date
t	0	2nd semester t
	6	1nd semester t
	12	2nd semester t-1
	18	1nd semester t-1

4. Results

4.1. Bias

Table 3 presents the ME for every institution and forecast horizon. Several facts are readily uncovered. First, all institutions tend to overestimate GDP growth at 18- and 12-month spans and underestimate it at 6- (with the exception of EC) and 0-month spans, as given by the change in the sign of ME. Moreover, biases for 18- and 12-month spans are quite significant, varying between -0.82 (BdP) and -1.37 (OECD) p.p. in the former case, and between -0.47 (BdP) and -0.98 (GBO) p.p. in the latter. From international institutions, EC forecasts display the lowest biases for 6-month spans and over. As expected, bias for all institutions is significantly reduced as the time span falls from 12- to 6-month, suggesting that the accuracy of forecasts for year t significantly improves as the information for $t-1$ becomes available. Hence, forecasts with time spans of over one year should be interpreted with more caution, as they are generally associated with large errors.

Table 3 – Bias Statistics: Mean Error

	GDP				Private Consumption (C)				Public Consumption (G)				Investment (I)				Exports (X)				Imports (M)			
	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0
IMF	-1,26	-0,81	0,17	0,17																				
OECD	-1,37	-0,82	0,07	0,06	-0,40	0,25	0,42	0,42	0,98	1,39	1,21	0,99	-6,67	-4,63	-0,82	0,16	-2,76	-1,26	1,71	0,39	-3,13	-0,86	1,69	1,44
EC	-1,09	-0,65	-0,04	0,14	0,03	0,36	0,61	0,40	0,95	1,12	1,00	0,71	-4,53	-3,06	-0,86	0,86	-2,08	-1,00	0,51	0,07	-1,58	-0,35	1,12	1,18
BdP	-0,82	-0,47	0,14	0,22	0,20	0,28	0,58	0,44	1,30	1,23	1,39	0,31	-3,33	-2,02	0,04	0,78	-2,32	-1,00	0,70	0,05	-1,34	-0,06	1,68	0,71
GBO		-0,98				0,16				1,89				-5,23				-2,00				-1,30		

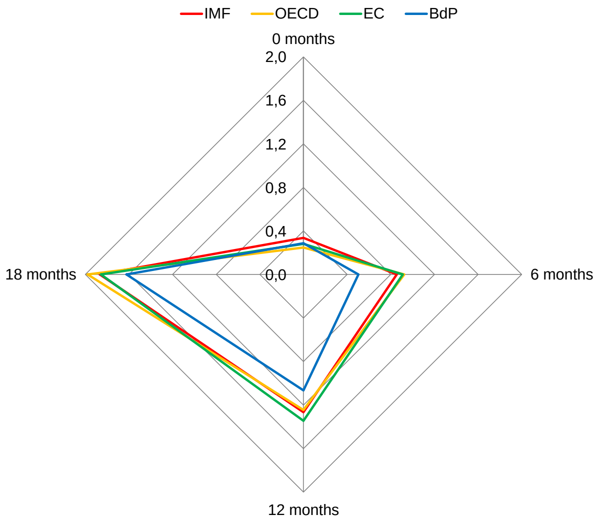
Second, all institutions tend to underestimate private consumption and public consumption at all time spans (except OECD for private consumption at the 18-month horizon), with public consumption having a higher ME (except BdP at 0-month forecasts). That is, forecasts for public consumption are more biased, on average, than those for private consumption. Inversely, all institutions tend to overestimate investment at 6-months and over (except BdP at 6-months), but underestimate it at 0-month spans. This component displays the largest ME at 18- and 12-month horizons for all institutions, with values comprised between -2,02 (BdP, 12-month predictions) and -6,67 (OECD, 18-month predictions) p.p.. Exports and imports have a similar ME pattern: all institutions overestimate at 18- and 12-months and underestimate at 6- and 0-months. These results suggest that negative GDP forecast errors at longer horizons (18- and 12-month spans) may be driven by overly optimistic investment and exports forecasts. At shorter horizons (6- and 0-month spans), institutions' forecasts underestimate the effective values on average, except for investment in some cases.

4.2. Dispersion

Table 4 shows the MAE and the MASE statistics for every institution and forecast horizon, for GDP and the corresponding expenditure components' forecasts.

MAE points towards average absolute errors that are monotonically decreasing in the forecast horizon for all institutions. The exceptions are forecasts for public consumption issued by BdP (at 6- and 12-month horizons) and by EC (at 0- and 6-months horizons). This fact is fully expected, since more information is available as the forecast horizon shortens. This statistic indicates that, at the 18-month span, OECD's forecasts display the highest average absolute errors (1.98 p.p.), whereas BdP's forecasts have the smallest average absolute errors (1.62 p.p.). At 12- and 6-month spans BdP's forecasts are still those with the smallest average absolute errors (1.06 and 0.50 p.p., respectively). At the 0-month span the forecast accuracy of different institutions is quite similar, varying between 0.25 (OECD) and 0.34 (IMF) p.p.. Figure 3 provides a graphical perspective of the MAE for all institutions and all time spans.

Figure 3 – Mean Absolute Error of GDP growth forecasts



The MAE statistic for GDP components suggests that forecasts for investment, exports and imports have the lowest accuracy. However, this does not imply that forecast models perform worse in predicting these components – since they are more volatile, they are also naturally harder to predict. We take this issue into account by using MASE to compare the accuracy of forecasts across GDP components. The radar plots in Figure 4 (in the appendix) illustrate the differences between MAE and MASE for different time spans. As the volatility measure is above 1 for all variables, the values displayed by the MAE are systematically larger than those from the MASE. More importantly, the conclusions yielded by each of these measures are substantially different.

From the analysis of MASE we conclude that, after correcting for volatility, forecast models perform comparatively worse when predicting investment at longer horizons (18- and 12-months) and public consumption at shorter horizons (6- and 0-months). This is not surprising, since investment decisions in the long-run are crucially affected by expectations, while in the short-run those decisions were already made and investment projects that have gone underway will hardly be canceled. Public consumption, on the other hand, is a political variable, often used by policy-makers to manipulate the economic cycle and to boost GDP growth, particularly in electoral periods. Hence, it is natural that, even in the short-run, public consumption cannot be accurately predicted by forecast models, when compared with other components, and adjusting for volatility. Institutions' forecast models seem to perform relatively well when predicting private consumption at longer horizons, but the volatility adjusted forecast accuracy does not increase as much as those of other components as the horizon shortens. In fact, at the 0-month span, when volatility is taken into consideration, exports and imports are the best predicted GDP components.

Although there is not one single institution providing the most accurate forecasts for all variables at all time spans, BdP ranks first in most cases, according to the MASE. A clear exception is public consumption, where the forecasts from EC seem to outperform those from BdP at 18-, 12- and 6-months spans. As we pointed out earlier, the good performance of BdP forecasts may be associated with the issue date of those forecasts. Among international institutions, EC delivers more accurate forecasts than OECD, on average, at most time spans, for all variables except private consumption at longer horizons.

The MSE should not be used to compare the quality of institutions' forecasts across GDP components, since it is not scaled. However, its disaggregation into a bias component and a dispersion component provides a new insight on the source of a forecast's poor performance. This disaggregation is shown in Table 5. Naturally, the dispersion (given by variance) for GDP, private consumption and public consumption is much lower than for investment, exports and imports, confirming the conclusions from MASE. The share of the MSE that is explained by the bias component relative to dispersion is also greater for these latter components, confirming that the poor quality embodied in the predictions for those components is mostly induced by the higher volatility levels. Table 5 also shows that investment forecast errors have the largest bias for all institutions at longer horizons (18- and 12-months), but this bias is significantly reduced at shorter horizons (6- and 0-months).

The contributions of expenditure components to the average GDP forecast error are detailed in Table 6. Notice that average discrepancies, ϵ, originating from the difference between actual GDP expenditure components shares and the shares used by institutions in forecast models, are small.[7] At longer horizons (18- and 12-month), the overly optimistic forecasts for investment explain most of the large deviations of forecasted GDP growth from actual values. In fact, for all institutions, investment forecast errors contribute in more than 100% to the forecast error of GDP growth, even though investment represents a smaller share of GDP as compared to other components. This result suggests that institutions should direct their efforts into improving the accuracy of investment forecasts. The external sector is also overestimated and the contributions of the forecast errors of exports and imports are significant in magnitude, especially at the 18-month span. However, since imports contribute negatively to GDP, the corresponding forecast errors partially offset those from other components. The contributions of private consumption and public consumption to the GDP forecast error are smaller, even at longer time spans, as these components are easier to predict. In fact, those components which are harder to predict also display the largest contributions to the GDP forecast error.

At 6- and 0-month spans, GDP forecast errors are small in average, although this is achieved through large errors in components' predictions. These errors tend to cancel out: except for investment, all components are generally underestimated, but the forecast error originating from imports, which enters with a negative sign in the GDP identity, mostly offsets those arising from other components. In fact, imports present the largest contribution to GDP forecast error at shorter horizons in almost all institutions. This is generally confirmed by MTWAE in Table 7: the average forecast errors across all components are comprised between 2.73 and 3.85 p.p. for the 6-month span, and between 1.54 and 2.05 p.p. for the 0-month span, with imports consistently presenting the largest contribution to the statistic.

[7] In practice, ϵ may also accommodate any statistical discrepancy shown by the data, and the marginal contribution of the change in inventories, which are often unreported by institutions and thus ignored in the analysis below.

Table 4 – Standard Dispersion Statistics: MAE and MASE (in p.p.)

		GDP				Private Consumption (C)				Public Consumption (G)				Investment (I)				Exports (X)				Imports (M)			
		18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0
MAE	**IMF**	1,87	1,27	0,85	0,34																				
	OECD	1,98	1,24	0,92	0,25	1,29	0,82	0,74	0,59	2,04	1,51	1,51	1,28	6,67	4,70	2,99	1,28	6,70	4,58	3,33	1,87	5,19	3,79	3,43	1,56
	EC	1,86	1,34	0,91	0,28	1,30	0,91	0,73	0,56	1,24	1,23	1,48	1,27	5,52	3,56	2,25	1,66	5,87	4,98	2,83	1,63	4,84	3,83	2,32	1,59
	BdP	1,62	1,06	0,50	0,29	1,15	0,88	0,66	0,53	1,51	1,53	1,69	0,87	4,02	3,21	1,50	1,28	6,17	4,56	2,61	1,21	4,60	3,78	2,13	0,88
	GBO		1,45				1,00				1,89				5,39				4,57				3,93		
MASE	**IMF**	0,98	0,66	0,45	0,18																				
	OECD	1,03	0,65	0,48	0,13	0,77	0,49	0,44	0,35	1,06	0,78	0,79	0,66	1,46	1,03	0,66	0,28	0,89	0,61	0,44	0,25	0,80	0,59	0,53	0,24
	EC	0,97	0,70	0,48	0,15	0,77	0,55	0,44	0,34	0,64	0,64	0,77	0,66	1,21	0,78	0,49	0,36	0,78	0,66	0,38	0,22	0,75	0,59	0,36	0,25
	BdP	0,85	0,56	0,26	0,15	0,69	0,53	0,39	0,32	0,78	0,79	0,88	0,45	0,88	0,70	0,33	0,28	0,82	0,61	0,35	0,16	0,71	0,58	0,33	0,14
	GBO		0,76				0,60				0,98				1,18				0,61				0,61		

Table 5 – MSE Components (in squared p.p.)

		GDP				Private Consumption (C)				Public Consumption (G)				Investment (I)				Exports (X)				Imports (M)			
		18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0	18	12	6	0
MSE	**IMF**	4,60	2,31	0,91	0,14																				
ME^2+Var	**OECD**	5,33	2,20	1,18	0,09	2,36	1,00	0,70	0,45	5,24	3,88	3,56	2,19	61,57	30,64	14,73	2,27	62,19	31,10	21,12	4,66	44,28	19,70	18,97	4,70
	EC	4,76	2,30	1,01	0,11	2,58	1,26	0,91	0,44	2,88	3,10	3,56	2,11	42,02	18,81	6,67	3,89	51,30	39,85	14,25	4,48	36,27	21,00	8,07	4,41
	BdP	3,73	1,42	0,37	0,10	2,02	1,13	0,69	0,43	4,84	3,67	3,59	1,40	28,16	16,89	2,92	2,75	53,84	28,80	10,08	2,08	31,06	20,80	8,17	1,62
	GBO		3,03				1,47				4,74				41,18				35,06				24,86		
Bias	**IMF**	1,59	0,65	0,03	0,03																				
ME^2	**OECD**	1,88	0,67	0,01	0,00	0,16	0,06	0,18	0,18	0,95	1,92	1,46	0,97	44,52	21,42	0,67	0,03	7,61	1,59	2,92	0,15	9,78	0,74	2,87	2,07
	EC	1,20	0,42	0,00	0,02	0,00	0,13	0,37	0,16	0,91	1,26	1,00	0,50	20,50	9,37	0,74	0,74	4,33	1,01	0,26	0,01	2,51	0,12	1,25	1,40
	BdP	0,67	0,22	0,02	0,05	0,04	0,08	0,34	0,19	1,69	1,52	1,92	0,10	11,11	4,09	0,00	0,61	5,38	1,00	0,48	0,00	1,79	0,00	2,83	0,50
	GBO		0,96				0,02				3,58				27,33				4,01				1,68		
Dispersion	**IMF**	3,01	1,66	0,88	0,11																				
Var	**OECD**	3,45	1,53	1,18	0,09	2,20	0,94	0,52	0,28	4,29	1,96	2,10	1,22	17,05	9,22	14,06	2,24	54,57	29,52	18,20	4,51	34,49	18,96	16,10	2,63
	EC	3,56	1,87	1,00	0,09	2,58	1,13	0,54	0,28	1,97	1,84	2,56	1,61	21,52	9,44	5,93	3,15	46,97	38,84	13,99	4,48	33,76	20,88	6,83	3,01
	BdP	3,06	1,20	0,35	0,06	1,98	1,05	0,35	0,24	3,16	2,16	1,67	1,31	17,05	12,80	2,92	2,13	48,46	27,80	9,60	2,08	29,26	20,79	5,33	1,12
	GBO		2,07				1,45				1,15				13,85				31,05				23,18		

An examination across institutions shows that the lower GDP forecast errors displayed by BdP are associated with better predictions for investment: in fact, this component seems to display the highest gain from the additional information available to BdP when issuing their forecasts. At shorter horizons, BdP also issues the most reliable forecasts, with the lowest average forecast errors across all components, as shown by MTWAE. The forecasts issued by EC are associated with more accurate contributions vis-à-vis OECD's, particularly for investment and the external sector. This is reflected into a lower MTWAE for all forecast horizons.

Table 6 - Contributions of expenditure components (in p.p.) to average GDP forecast error

	OECD				BoP				EC				GBO
	18	12	6	0	18	12	6	0	18	12	6	0	12
Private Consumption	-0,26	0,17	0,28	0,27	0,14	0,19	0,38	0,28	0,03	0,24	0,40	0,26	0,10
Public Consumption	0,20	0,28	0,24	0,20	0,26	0,25	0,28	0,06	0,19	0,22	0,20	0,14	0,38
Investment	-1,67	-1,16	-0,29	-0,02	-0,86	-0,54	-0,05	0,13	-1,18	-0,81	-0,29	0,15	-1,31
Exports	-0,85	-0,41	0,52	0,12	-0,72	-0,32	0,21	0,01	-0,66	-0,35	0,13	0,02	-0,65
Imports	-1,21	-0,36	0,70	0,56	-0,54	-0,06	0,66	0,26	-0,63	-0,17	0,41	0,46	-0,53
GDP by components	-1,36	-0,77	0,05	0,01	-0,65	-0,37	0,16	0,22	-0,98	-0,53	0,02	0,11	-0,94
discrepancy (ϵ)	0,01	0,04	-0,02	-0,05	0,16	0,10	0,02	0,01	0,11	0,12	0,07	-0,03	0,04
GDP reported	**-1,37**	**-0,82**	**0,07**	**0,06**	**-0,82**	**-0,47**	**0,14**	**0,22**	**-1,10**	**-0,65**	**-0,04**	**0,14**	**-0,98**

Table 7 - The MTWAE statistic and its decomposition (in p.p.)

	OECD				BoP				EC				GBO
	18	12	6	0	18	12	6	0	18	12	6	0	12
MTWAE	**6,91**	**4,93**	**3,85**	**2,05**	**5,87**	**4,70**	**2,73**	**1,54**	**6,10**	**4,86**	**3,12**	**2,02**	**5,32**
Private Consumption	0,84	0,53	0,48	0,38	0,75	0,57	0,43	0,35	0,84	0,60	0,48	0,37	0,65
Public Consumption	0,41	0,30	0,31	0,26	0,30	0,31	0,34	0,18	0,25	0,25	0,30	0,26	0,38
Investment	1,67	1,26	0,72	0,27	1,19	1,01	0,34	0,35	1,38	1,03	0,65	0,31	1,38
Exports	2,00	1,39	1,01	0,55	1,85	1,36	0,79	0,35	1,77	1,51	0,83	0,48	1,39
Imports	1,99	1,45	1,33	0,60	1,77	1,45	0,83	0,33	1,86	1,48	0,87	0,61	1,52

4. Conclusions

This article analyzes the quality of forecasts for real GDP growth and for the corresponding expenditure components on two dimensions: bias and dispersion. We use forecast data issued for Portugal by five national and international institutions, covering the 2002-2010 period. Our conclusions indicate that, along the bias dimension, forecasts for real GDP growth are on average optimistic at longer forecast horizons. This is mostly explained by the negative contributions of investment and exports to the GDP forecast error. At shorter horizons, forecasts for GDP growth are in general accurate; however, this is achieved with large errors in GDP expenditure components' predictions, whose effects tend to cancel out. To measure this, we propose a new statistic – termed Mean of Total Weighted Absolute Error, which evaluates the average absolute forecast error across all GDP expenditure components, in percentage points. This statistic shows that, even though the average absolute errors of GDP forecasts are below 1 percentage point for all institutions and for same-year predictions, the total forecast error across all components is comprised between 1.5 and 4 percentage points. The forecasts issued by *Banco de Portugal* are generally better than those from other institutions, particularly for larger horizons, an outcome which is possibly related with the larger set of information available at that time – in general, their forecasts are issued a couple of months after those from other institutions. The forecast accuracy of international institutions is very similar; however, those issued by the European Commission have the upper edge when predicting GDP components.

Standard statistical measures indicate that investment, exports and imports are the hardest components to predict; however, these components are also the most volatile, for which these measures are inappropriate to compare the forecast accuracy across different components. We address this issue by using a scaling factor, which corrects the mean absolute error for the inverse of the volatility of each series. This statistic suggests that forecast models perform comparatively worse when predicting investment at longer horizons (1-year head predictions) and public consumption at shorter horizons (same-year predictions), and perform comparatively better when predicting private consumption and imports at longer horizons and exports and imports at shorter horizons.

5. Bibliography

Ash, J., D. Smyth, and S. Heravi (1998). *Are OECD forecasts rational and useful? A directional analysis,* International Journal of Forecasting, 14, 381-391.

Cairncross, A. K. (1969). *Economic forecasting,* Economic Journal, 79, 797-812.

Clemen, R. T. and R. L. Winkler (1986). *Combining economic forecasts*, Journal of Business and Economic Statistics, 4, 39-46.

Fildes, R. and H. O. Stekler (2002). *The state of macroeconomic forecasting*, Journal of Macroeconomics, 24, 435-468.

Holden, K. and D. Peel (1985). *An evaluation of quarterly National Institute forecasts,* Journal of Forecasting, 4, 227-234.

Holden, K. and D. Peel (1990). *On testing for unbiasedness and efficiency of forecasts*, The Manchester School, 58, 120-127.

Hyndman, R. J. and A. B. Koehler (2006). *Another look at measures of forecast accuracy,* International Journal of Forecasting, 22, 679-688.

Joutz, F. (1988). *Informational efficiency tests of quarterly macroeconometric GNP forecasts from 1976 to 1985,* Managerial and Decision Economics, 9, 311-330.

Loungani, P. (2001). *How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth,* International Journal of Forecasting, 17, 419-432.

McNees, S. K. (1976). *An evaluation of economic forecasts: extension and update*, New England Economic Review, 20-44.

McNees, S. K. (1978). *The "rationality" of economic forecasts,* American Economic Review, 68, 301-305.

McNees, S. K. (1986). *Forecasting accuracy of alternative techniques: a comparison of US macroeconomic forecasts,* Journal of Business and Economic Statistics, 4, 5-15.

McNees, S. K. (1988). *How accurate are macroeconomic forecasts?* New England Economic Review,

15-36.

Melander, A., G. Sismanidis, and D. Grenouilleau (2007). *The track record of the Commission's forecasts – an update,* European Economy – Economic Papers 291, Directorate General Economic and Monetary Affairs, European Commission.

Mincer, J. A. and V. Zarnowitz (1969) *The evaluation of economic forecasts*, in Economic forecasts and expectations: analysis of forecasting behavior and performance, National Bureau of Economic Research, NBER Chapters, 1-46.

Moore, G. H. (1969). *Forecasting short-term economic change,* Journal of the American Statistical Association, 64, 1-22.

Nordhaus, W. D. (1987). *Forecasting efficiency: concepts and applications,* Review of Economics and Statistics, 69, 667-674.

Öller, L. E. and B. Barot (2000). *The accuracy of European growth and inflation forecasts,* International Journal of Forecasting, 16, 293-315.

Pons, J. (2000). *The accuracy of IMF and OECD forecasts for G7 countries,* Journal of Forecasting, 19, 53-63.

Stekler, H. O. (1972). *An analysis of turning point forecasts,* American Economic Review, 62, 724-729.

Stekler, H. O. (1987). *Who forecasts better?* Journal of Business and Economic Statistics, 5, 155-158.

Theil, H. (1958). Economic Forecasts and Policy, Amsterdam: North-Holland Pub. Co.

Theil, H. (1966). Applied economic forecasting, Amsterdam: North-Holland Pub. Co.

Timmermann, A. (2007). *An evaluation of the World Economic Outlook forecasts,* IMF Staff Papers, 54, 1-33.

Vuchelen, J. and M. I. Gutierrez (2005). *A direct test of the information content of the OECD growth forecasts,* International Journal of Forecasting, 21, 103-117.

Wallis, K. F. (1989). *Macroeconomic forecasting: a survey,* Economic Journal, 99, 28-61.

Zarnowitz, V. (1967). An appraisal of short-term economic forecasts, New York: National Bureau of Economic Research.

Zarnowitz, V. (1979). *An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level,* Journal of Business, 52, 1-33.

Zarnowitz, V. (1984). *The accuracy of individual and group forecasts from business outlook surveys,* Journal of Forecasting, 3, 11-26.

Appendix

Figure 4 – Mean Absolute Error and Mean Absolute Scaled Error (in p.p.)