GEE Papers
Número 41
Outubro de 2011

Evaluating the forecast quality of GDP components

Pedro M. Esperança
João C. Fonseca
Paulo Júlio

Evaluating the forecast quality of GDP components*

Abstract:

We assess and compare the quality of forecasts issued for Portugal, at several time spans. Our analysis, covering the 2002-2010 period, focuses on real GDP growth and the corresponding expenditure components. We use a scaled statistic to compare the forecast accuracy of GDP components with different volatility levels, and explore the contributions of expenditure components to the GDP forecast error. Moreover, we propose two new statistics – termed Mean of Total Weighted Absolute Error and Mean of Total Weighted Squared Error – to evaluate the overall accuracy of components' predictions. The results suggest that GDP forecasts are generally optimistic at longer horizons (1-year ahead predictions), mainly due to overly optimistic forecasts in investment and exports. At shorter horizons (same-year predictions), GDP forecasts are more accurate, but this is achieved with relatively large errors in components' predictions, whose effects tend to cancel out.

Acknowledgments

We are grateful to Miguel de Carvalho and Ricardo Pinheiro-Alves for precious comments.

Pedro M. Esperança
Currently M.Sc. student at the University of Oxford. This article was written while the author was visiting at the *Gabinete de Estratégia e Estudos*, Portuguese Ministry of Economy and Employment.
Email: pedro.esp88@gmail.com

João C. Fonseca
Gabinete de Estratégia e Estudos, Portuguese Ministry of Economy and Employment.
Email: joao.fonseca@gee.min-economia.pt

Paulo Júlio
Gabinete de Estratégia e Estudos, Portuguese Ministry of Economy and Employment.
Email: paulo.julio@gee.min-economia.pt

JEL Classification: C52, C53, E37

Keywords: Forecast evaluation, GDP expenditure components, Contributions to GDP growth, Mean of total weighted absolute error, Mean of total weighted squared error, Portugal

[*] 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.

Index

Evaluating the forecast quality of GDP components*

Pedro M. Esperança† João C. Fonseca‡ Paulo Júlio§

October 5, 2011

Abstract

We assess and compare the quality of forecasts issued for Portugal, at several time spans. Our analysis, covering the 2002–2010 period, focuses on real GDP growth and the corresponding expenditure components. We use a scaled statistic to compare the forecast accuracy of GDP components with different volatility levels, and explore the contributions of expenditure components to the GDP forecast error. Moreover, we propose two new statistics—termed Mean of Total Weighted Absolute Error and Mean of Total Weighted Squared Error—to evaluate the overall accuracy of components' predictions. The results suggest that GDP forecasts are generally optimistic at longer horizons (1-year ahead predictions), mainly due to overly optimistic forecasts in investment and exports. At shorter horizons (same-year predictions), GDP forecasts are more accurate, but this is achieved with relatively large errors in components' predictions, whose effects tend to cancel out.

JEL Classification: C52, C53, E37
Keywords: Forecast evaluation, GDP expenditure components, Contributions to GDP growth, Mean of total weighted absolute error, Mean of total weighted squared error, Portugal

*We are grateful to Miguel de Carvalho and Ricardo Pinheiro-Alves for precious comments. 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.

†Currently M.Sc. student at the University of Oxford. This article was written while the author was visiting at the *Gabinete de Estratégia e Estudos,* Portuguese Ministry of Economy and Employment.

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

§*Gabinete de Estratégia e Estudos,* Portuguese Ministry of Economy and Employment. Corresponding author: paulo.julio@gee.min-economia.pt

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), government consumption (G), investment (I), exports (X) and imports (M).[1] 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 contributes to the literature by focusing simultaneously on these three different perspectives of forecast quality: 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 Cooperation and Development (OECD), International Monetary Fund (IMF), European Commission (EC), Central Bank of Portugal (*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 prediction models perform comparatively worse when predicting investment at longer horizons (1-year ahead predictions), and government 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 two new statistics—termed Mean of Total Weighted Absolute Error (MTWAE) and Mean of Total Weighted Squared Error (MTWSE)—to summarize the reliability 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. These statistics suggest that GBO's forecasts—only available at the 12-month horizon—are the least reliable. At the remaining time 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

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

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 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 across this dimension 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

$$e_t\left(s\right) = A_t - F_t\left(s\right) \tag{1}$$

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 Standard Evaluation Statistics

To evaluate the quality of forecasts we start with the simplest and standard measures of forecast evaluation: Mean Error (ME), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), respectively given by

$$\mathrm{ME} := \frac{1}{T}\sum_{t=1}^{T} e_t \tag{2}$$

$$\mathrm{MAE} := \frac{1}{T}\sum_{t=1}^{T} |e_t| \tag{3}$$

$$\mathrm{RMSE} := \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2} \tag{4}$$

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 (how biased forecasts are, on average). A negative value suggests that forecasts tend to be overly optimistic, whereas a positive value points towards pessimistic forecasts. The MAE provides a measure of the average total forecast error, regardless of the direction of the error (how off-target forecasts are, on average). Hence, a lower MAE reflects more accurate forecasts. The RMSE also provides a measure of total forecast error, but attributes disproportionally higher contributions to larger deviations from target.

2.3 Comparative Evaluation Statistics

Even though these standard statistics are already very informative, alternative measures are also possible to construct, allowing a direct comparison among any two forecasting methods. Take as a benchmark the often-called *naïve forecast*: a 'model' predicting the same change as in the last observed period (a random walk-like behavior)

$$N_t(s) = \begin{cases} A_{t-1} & , s \in \{0, 6\} \\ A_{t-2} & , s \in \{12, 18\} \end{cases} \tag{5}$$

for all t, where N_t stands for the naïve forecast. At 12- and 18-month horizons the actual values of $t-1$ are not yet known, and so the previous ones (those of $t-2$) must be used, something that is not often emphasized in the literature. Only historical data is needed to provide these forecasts, for which it constitutes the simplest forecasting method. Given such easiness in computation and the nature of the underlying assumptions, it serves naturally as a minimum benchmark that any formal forecasting model should outperform. Letting η_t denote the error of the naïve forecast

$$\eta_t = A_t - N_t \tag{6}$$

we can compute the Theil's (1966) U_2 statistic

$$\mathrm{U}_2 := \sqrt{\frac{\sum_{t=1}^{T}(A_t - F_t)^2}{\sum_{t=1}^{T}(A_t - N_t)^2}} = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} \eta_t^2}} \tag{7}$$

which provides a parsimonious comparative statistic *vis-à-vis* the naïve forecasting method. A value of 1 means that naïve and institutions' forecasts have similar forecasting ability; a value smaller than 1 implies that institutions' forecasts outperform the naïve forecasts; and a value larger than 1 attests to a better forecasting accuracy of the naïve method. In this latter case, institutions' forecasts have no valuable content, on average, when compared to the naïve method. This statistic can be used to compare *any* two forecasting methods by replacing N_t in (7) with the forecasts from any other alternative model; for instance, from a Vector Autoregressive (VAR) model.

2.4 Scaled Statistics

The previous statistics are only valid when comparing a variable's forecasts coming from different institutions or forecasting methods—one of the dimensions of our analysis. If one aims to compare the accuracy 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

Table 1: *Volatility as measured by V.*

	Volatility	*Volatility relative to GDP*
Gross Domestic Product	1.84	1.00
Private Consumption	1.49	0.81
Government Consumption	1.96	1.06
Investment	4.48	2.44
Exports	6.78	3.69
Imports	5.91	3.22

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). Let J be the sample size. The scaled errors

$$v_t = e_t \left(\frac{1}{J} \sum_{j=1}^{J} |A_j - A_{j-1}| \right)^{-1} = e_t V^{-1} \tag{8}$$

can thus be used in (3), with v_t replacing e_t, to obtain the Mean Absolute Scaled Error (MASE).[2] 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.

2.5 Contributions Analysis

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

$$gdp_t \equiv c_t w_{t-1}^C + g_t w_{t-1}^G + i_t w_{t-1}^I + x_t w_{t-1}^X - m_t w_{t-1}^M \tag{9}$$

[2] As pointed out by Hyndman and Koehler (2006), who proposed this statistic, this is equivalent to rescaling the errors with the MAE of the naïve forecasting method. However, in our case, this interpretation is only valid for the 0- and 6-month forecasts, as for 12- and 18-month forecasts the errors of the naïve method are given by $A_t - A_{t-2}$ instead of $A_t - A_{t-1}$.

With forecasted values, a similar version of equation (9) 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

$$gdp_t^f \equiv c_t^f w_{t-1}^C + g_t^f w_{t-1}^G + i_t^f w_{t-1}^I + x_t^f w_{t-1}^X - m_t^f w_{t-1}^M + \epsilon_t \tag{10}$$

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 (10) from (9) and taking the average yields

$$\frac{1}{T}\sum_{t=1}^{T}\left(e_t^{gdp} + \epsilon_t\right) \equiv \frac{1}{T}\sum_{t=1}^{T}\left(e_t^c w_{t-1}^C + e_t^g w_{t-1}^G + e_t^i w_{t-1}^I + e_t^x w_{t-1}^X - e_t^m w_{t-1}^M\right) \tag{11}$$

In equation (11), $T^{-1}\sum_{t=1}^{T} e_t^z w_{t-1}^Z$ represents the average contribution of the forecast error arising from variable Z, in percentage points (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 (11), 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 two additional measures of forecast quality. The first evaluates the sum across components of the absolute distance between forecasted and actual contributions. We term this new statistic the 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

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

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

$$\text{MTWAE} := \frac{1}{T}\sum_{t=1}^{T} |\mathbf{e}_t|' \, \mathbf{w}_{t-1} \tag{12}$$

where $|\mathbf{e}_t|$ is a vector whose entries are the absolute values of the entries in $\mathbf{e}_t$. This statistic is computed for every institution and forecast horizon. Those institutions with higher values in the 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 more reliable are institutions' predictions in general, *ceteris paribus*. Naturally, the MTWAE statistic can be decomposed in the respective components' contributions, $T^{-1}\sum_{t=1}^{T} |e_t^z| \, w_{t-1}^Z$.

The second statistic evaluates the sum across components of the squared errors, each weighted by the corresponding GDP share. It is thus similar to the MTWAE, expect that larger errors contribute disproportionately more to the statistic. Letting

$$\mathbf{\Omega}_t = \mathrm{diag}(w_t^C, w_t^G, w_t^I, w_t^X, w_t^M)$$

this statistic, named Mean of Total Weighted Squared Error (MTWSE), is defined as

$$\mathrm{MTWSE} := \frac{1}{T}\sum_{t=1}^{T} \mathbf{e}_t' \mathbf{\Omega}_{t-1} \mathbf{e}_t \tag{13}$$

Naturally, one can take the square root of the MTWSE to convert the measurement unit to the original scale. However, it may be advantageous to use instead the expression presented in (13), since this can be easily decomposed into the respective components' contributions, $T^{-1}\sum_{t=1}^{T}(e_t^z)^2 w_{t-1}^Z$.

3 Data

Our dataset contains information on forecasts for *Gross Domestic Product*, *Private Consumption*, *Government 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: Organization for Economic Cooperation and Development (OECD), International Monetary Fund (IMF), European Commission (EC), Bank of Portugal (*Banco de Portugal*—BdP) and Portuguese Government Budget Office (GBO). Actual values and GDP expenditure components shares were taken from the Portuguese National Institute of Statistics. 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 forecasts issued by BdP: since these are issued later, they use one additional quarter of information relative to other institutions. 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 Gross Domestic Product

We first focus on GDP growth forecasts. Table 3 presents the statistics for every institution and forecast horizon. Considering the three standard measures of forecast quality, 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. Among international institutions, EC's 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 extreme caution, as they are generally associated with large errors.

Second, MAE points towards average absolute errors that are monotonically decreasing in the forecast horizon for all institutions. 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.. Forecasts issued by GBO—only available at 12-month—have the lowest accuracy. Figure 1 provides a graphical perspective of MAE for all institutions and all time spans. The RMSE presents similar qualitative results, also pointing out towards significant forecast errors at longer horizons.

We now turn to some comparative statistics. Two Theil's U_2 statistics were computed: one which compares institutions' forecasts with the naïve method (U_2n), and another which takes as benchmark the forecasts of a VAR(1) model (U_2var). From the analysis of U_2n, we observe that institutions' forecasts perform better than the naïve method at all time spans, even though differences are small in some cases at the 18-month horizon (*i.e.*, the

Table 3: *Evaluation Statistics: GDP*

		GDP Growth			
		18	*12*	*6*	*0*
ME	*OECD*	-1.37	-0.82	0.07	0.06
	IMF	-1.26	-0.81	0.17	0.17
	BdP	-0.82	-0.47	0.14	0.22
	EC	-1.09	-0.65	-0.04	0.14
	GBO		-0.98		
MAE	*OECD*	1.98	1.24	0.92	0.25
	IMF	1.87	1.27	0.85	0.34
	BdP	1.62	1.06	0.50	0.29
	EC	1.86	1.34	0.91	0.28
	GBO		1.45		
RMSE	*OECD*	2.31	1.48	1.09	0.31
	IMF	2.15	1.52	0.95	0.38
	BdP	1.93	1.19	0.61	0.32
	EC	2.18	1.52	1.00	0.34
	GBO		1.74		
U2n	*OECD*	0.95	0.61	0.52	0.15
	IMF	0.89	0.63	0.46	0.18
	BdP	0.80	0.49	0.29	0.15
	EC	0.90	0.63	0.48	0.16
	GBO		0.72		
U2var	*OECD*	1.20	0.77	0.60	0.17
	IMF	1.12	0.79	0.52	0.21
	BdP	1.01	0.62	0.34	0.18
	EC	1.14	0.79	0.55	0.19
	GBO		0.91		
MASE	*OECD*	1.08	0.68	0.50	0.13
	IMF	1.02	0.69	0.46	0.18
	BdP	0.88	0.58	0.27	0.16
	EC	1.01	0.73	0.49	0.15
	GBO		0.79		

U_2n statistic is close to 1), mainly for OECD.[3] As expected, shortening the horizon reveals the superiority of institutions' forecasts relative to the naïve method.

Although conceptually more sophisticated than the naïve method, obtaining a forecast from a VAR model is still a simple task in terms of information and computations required—especially when compared with economy-wide models—so that it can serve as an appropriate benchmark against which we can evaluate institutions' forecasts. Besides GDP, our VAR model considered 5 additional variables, easily obtained from available macroeconomic databases: GDP deflator, exchange rate, oil price, interest rate and an index of the Portuguese stock market. Prior to running the VAR, all variables (except the

[3] If one had instead compared 18-month forecasts for year t with a naïve forecast using data from $t-1$ (and not $t-2$), one would erroneously conclude that naïve forecasts can outperform institutions' forecasts for this time span, at least in some cases.

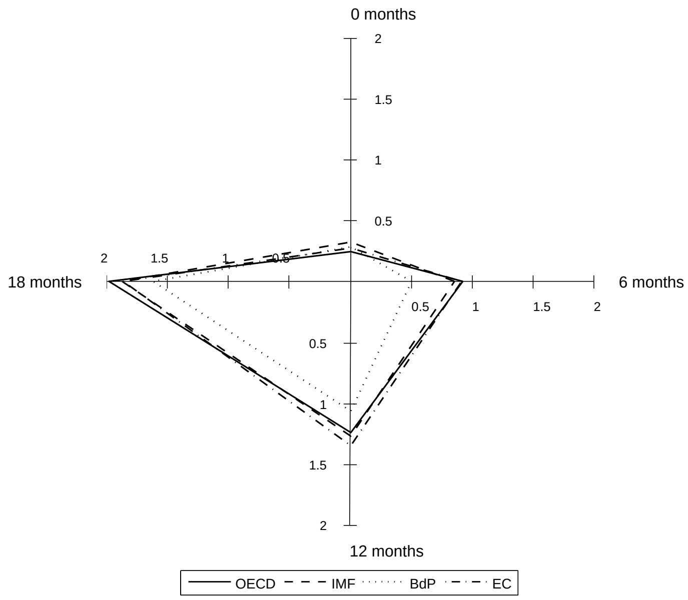

Figure 1: *Mean Absolute Error of GDP growth forecasts.*

interest rate) were converted to percentage changes, which assured stationarity. To obtain 2-year ahead forecasts (12- and 18-months), we used forecasts from the VAR model for $t-1$. This recursive method allowed us to obtain forecasts for t using only the information available at $t-2$. As shown by the U_2var statistic, institutions' forecasts at 18-months are outperformed by this simple VAR model. For BdP, the statistic is marginally identical to 1, suggesting a similar accuracy between both methods, even though the predictions from BdP use more updated data (namely, the first quarter of $t-1$). This confirms that models add little forecast accuracy relative to more simple and traditional methods at larger horizons, at least for GDP. However, institutions' prediction models clearly outperform the VAR model at shorter horizons.

This brief analysis shows that GBO's forecasts are the most biased (upwards) and the most inaccurate at 12-month, and that, except at the 0-month span, BdP's forecasts have the highest accuracy. This latter result may be related with the timing of issuance of forecasts: BdP issues its forecasts a couple of months after other institutions have done so, which allows them to have definite data from an additional quarter. The accuracy of forecasts issued by international institutions does not seem to differ substantially from one another, with EC taking a small lead in terms of unbiasedness. These results also suggest that prediction models used by institutions perform well at shorter horizons, but their accuracy at larger horizons—especially at the 18-month horizon—is clearly limited.

4.2 GDP Components

The statistics for GDP components are presented in Table 4. The standard statistics (MAE and RMSE) suggest 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 these 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 2 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 MAE are systematically larger than those from MASE. More importantly, the conclusions yielded by each of these statistics 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 government 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. Government 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, government 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 MASE. A clear exception is government consumption, where forecasts issued by EC seem to outperform those issued by BdP at 18-, 12- and 6-month spans. As we pointed out earlier, the good performance of BdP's forecasts may be associated with the issue date of those forecasts. Except for exports, GBO provides the least accurate forecasts at 12-months, even for government consumption, although it has more information on the budget than other institutions. 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.

Another conclusion that emerges from Table 4 is that the VAR model only outperforms institutions' forecasts at the 18-month span for investment and exports, and even in these cases only for some institutions.[4] Clearly, forecast models used by institutions are able to

[4]The VAR model was computed as explained for GDP, except that this variable was replaced by the

Table 4: *Evaluation Statistics: GDP Components*

		Private Consumption (C)				Gov. 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*
ME	*OECD*	-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
	BdP	0.20	0.29	0.59	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
	EC	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
	GBO		0.16				1.89				-5.23				-2.00				-1.30		
MAE	*OECD*	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
	BdP	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
	EC	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
	GBO		1.00				1.89				5.39				4.57				3.93		
RMSE	*OECD*	1.54	1.00	0.84	0.67	2.29	1.97	1.89	1.48	7.85	5.53	3.84	1.51	7.89	5.58	4.60	2.16	6.65	4.44	4.36	2.17
	BdP	1.42	1.06	0.83	0.65	2.20	1.92	1.90	1.18	5.31	4.11	1.71	1.66	7.34	5.37	3.18	1.44	5.57	4.56	2.86	1.27
	EC	1.61	1.12	0.95	0.66	1.70	1.76	1.89	1.45	6.48	4.34	2.58	1.97	7.16	6.31	3.78	2.12	6.02	4.58	2.84	2.10
	GBO		1.21				2.18				6.42				5.92				4.99		
U2n	*OECD*	0.85	0.55	0.45	0.36	0.91	0.78	0.83	0.65	1.19	0.84	0.71	0.28	0.87	0.61	0.50	0.23	0.98	0.66	0.56	0.28
	BdP	0.78	0.58	0.44	0.35	0.87	0.76	0.84	0.52	0.81	0.62	0.31	0.30	0.81	0.59	0.34	0.16	0.82	0.59	0.37	0.16
	EC	0.88	0.62	0.51	0.36	0.67	0.70	0.83	0.64	0.99	0.66	0.48	0.36	0.79	0.69	0.41	0.23	0.89	0.68	0.36	0.27
	GBO		0.67				0.86				0.98				0.65				0.74		
U2var	*OECD*	0.61	0.40	0.34	0.28	0.80	0.69	0.52	0.41	1.22	0.86	0.48	0.19	1.12	0.79	0.65	0.30	1.01	0.67	0.56	0.28
	BdP	0.57	0.42	0.34	0.27	0.77	0.67	0.53	0.33	0.83	0.64	0.21	0.21	1.04	0.76	0.45	0.20	0.84	0.69	0.37	0.16
	EC	0.64	0.45	0.39	0.27	0.59	0.62	0.52	0.40	1.01	0.68	0.32	0.24	1.02	0.90	0.53	0.30	0.91	0.69	0.36	0.27
	GBO		0.48				0.76				1.00				0.84				0.76		
MASE	*OECD*	0.86	0.55	0.50	0.39	1.04	0.77	0.77	0.65	1.49	1.05	0.67	0.29	0.99	0.67	0.49	0.28	0.88	0.64	0.58	0.26
	BdP	0.77	0.59	0.44	0.36	0.77	0.78	0.86	0.44	0.90	0.72	0.34	0.28	0.91	0.67	0.39	0.18	0.78	0.64	0.36	0.15
	EC	0.87	0.61	0.49	0.38	0.64	0.63	0.76	0.65	1.23	0.79	0.50	0.37	0.86	0.73	0.42	0.24	0.82	0.65	0.39	0.27
	GBO		0.67				0.97				1.20				0.67				0.66		

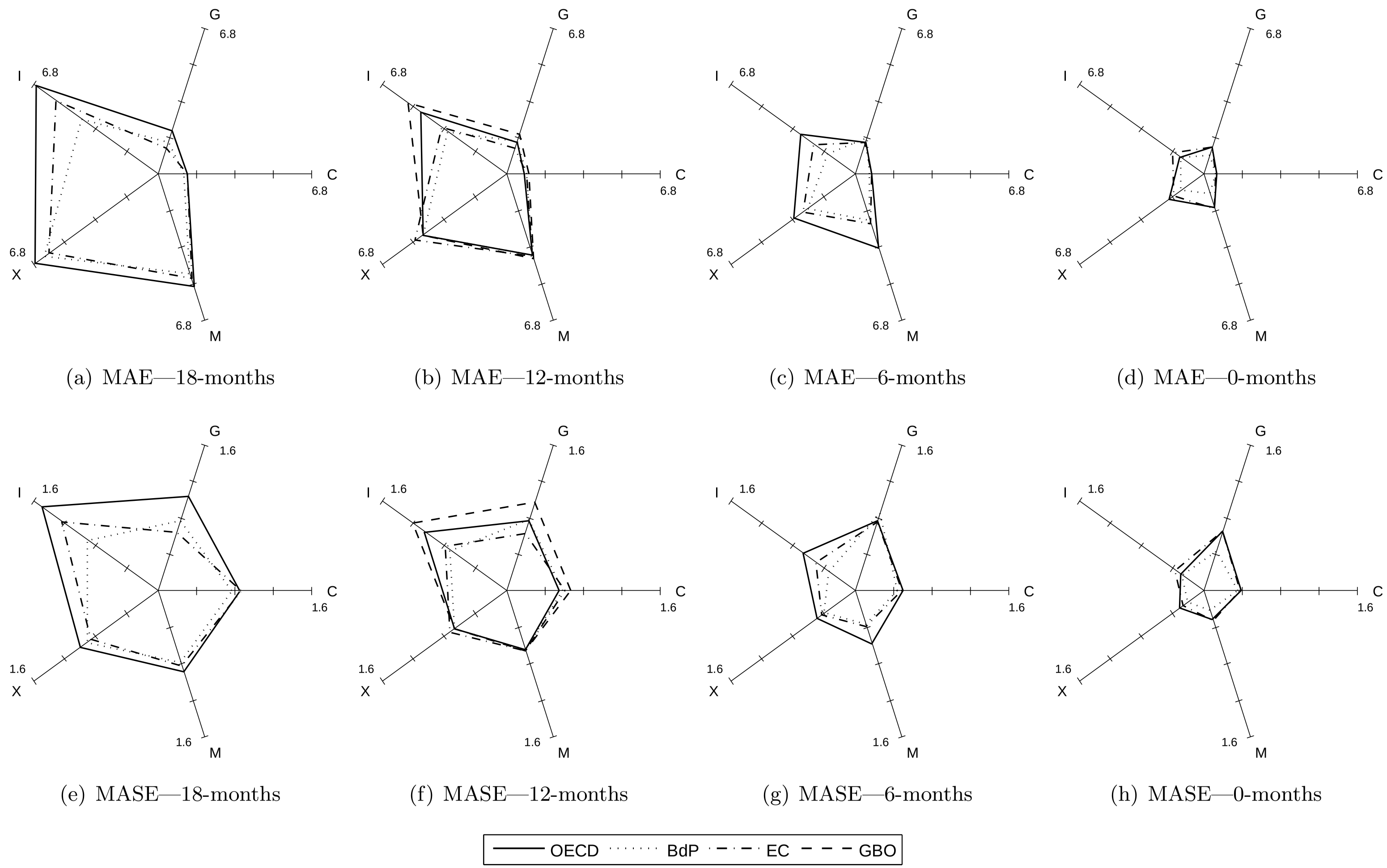

Figure 2: *Mean absolute error and mean absolute scaled error of GDP components.*

provide better forecasts for private consumption and government consumption relative to the naïve and VAR methods. This outcome probably results from two facts. Firstly, the variables selected in the VAR model were appropriate to capture movements in GDP, but do not accurately reflect movements in other variables; and secondly, institutions' models use microeconomic foundations to explain the behavior of private agents, something that is completely neglected in the VAR.

We find also that government consumption is systematically underestimated (contrary to other components), as shown by ME, and even at 0-month spans the average forecast errors are among the highest, with GBO providing the most biased forecasts for this variable. Economic theory provides a justification for this result: unexpected increases in government consumption present larger effects on aggregate demand *vis--vis* expected increases, since households cannot adjust their decisions timely. Thus, policy–makers seem to use government consumption deliberately to manipulate the economic cycle and to boost GDP—a conclusion that is also suggested by MASE. Moreover, GBO's forecasts are the most upwardly biased for private consumption, investment and the external sector. This hints that political objectives may be embodied in these predictions, namely the government's role in managing expectations: an optimistic forecast will be embodied by economic agents, possibly leading to a higher expenditure level, that would not have occurred under less optimistic forecasts. This may possibly avoid, smooth or delay a recession.

In general, forecasts for investment, exports and imports, are tendencially optimistic at longer horizons (18- and 12-month spans), whereas forecasts for private consumption and government consumption are tendencially pessimistic. Thus, negative GDP forecast errors 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.

4.3 Decomposing GDP Forecast Errors into Components' Contributions

The contributions of expenditure components to the average GDP forecast error are detailed in Table 5. Notice that average discrepancies, ϵ, originating from the difference between actual component shares on GDP and the shares used by institutions in forecast models, are small.[5] At longer horizons (18- and 12-month spans), 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

respective component.

[5]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.

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 government consumption to the GDP forecast error are smaller, even at longer time spans, as these components are easier to predict (*i.e.*, their volatility is comparatively smaller). 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 on 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. This is generally confirmed by MTWAE in Table 6: 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.

An examination of MTWAE and of MTWSE in Tables 6 and 7 across institutions shows that the lower GDP forecast errors displayed by BdP at longer horizons 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. Forecasts issued by EC are associated with more accurate contributions *viz-à-viz* OECD's, particularly for investment and the external sector. This is reflected into a lower MTWAE for all forecast horizons. The MTWSE is also lower for the EC except at the 12-month span, where OECD performs slightly better. This results from OECD issuing better forecasts for the external sector. The reliability of GBO's forecasts for exports and imports are also among the lowest, but the effects tend to cancel out, on average.

5 Concluding remarks

This article analyzes the quality of forecasts for real GDP growth and the corresponding expenditure components. We use forecast data issued for Portugal by five national and international institutions, covering the 2002–2010 period. Our conclusions indicate that forecasts for real GDP growth are on average optimistic at longer forecast horizons. This is mostly explained by optimistic forecasts for investment and exports. 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 two new statistics: the first, termed Mean of Total Weighted Absolute Error, evaluates the average absolute forecast errors across all GDP expenditure compo-

Table 5: *Contributions of expenditure components (in percentage points) to the average GDP forecast error.*

	OECD				BdP				EC				GBO
	18	*12*	*6*	*0*	*18*	*12*	*6*	*0*	*18*	*12*	*6*	*0*	*12*
C	-0.25	0.17	0.28	0.27	0.14	0.19	0.38	0.28	0.03	0.24	0.40	0.26	0.10
G	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
I	-1.67	-1.16	-0.29	-0.02	-0.86	-0.54	-0.05	0.13	-1.18	-0.81	-0.28	0.15	-1.31
X	-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
M	-1.21	-0.36	0.70	0.56	-0.54	-0.06	0.66	0.26	-0.63	-0.17	0.41	0.46	-0.54
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	-1.37	-0.82	0.07	0.06	-0.82	-0.47	0.14	0.22	-1.09	-0.65	-0.04	0.14	-0.98

Notes: (i) **GDP** in the table represents the mean of e_t^{gdp}, which is equal to the Mean Error statistic in Table 3, by definition, since it uses the GDP predictions reported by institutions to compute the forecast error; (ii) The sum of the contributions in the table differs from **GDP** by an error ϵ whose source is explained in Section 2.5.

Table 6: *The MTWAE statistic and its decomposition (in percentage points).*

	OECD				BdP				EC				GBO
	18	*12*	*6*	*0*	*18*	*12*	*6*	*0*	*18*	*12*	*6*	*0*	*12*
MTWAE	6.90	4.93	3.85	2.05	5.87	4.70	2.73	1.54	6.10	4.86	3.12	2.03	5.32
Components' contributions to the MTWAE													
C	0.84	0.53	0.48	0.38	0.75	0.57	0.43	0.35	0.85	0.60	0.48	0.37	0.65
G	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
I	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
X	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
M	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

Table 7: *The MTWSE statistic and its decomposition.*

	OECD				BdP				EC				GBO
	18	*12*	*6*	*0*	*18*	*12*	*6*	*0*	*18*	*12*	*6*	*0*	*12*
MTWSE	18.47	9.07	6.53	1.50	13.57	8.05	2.89	0.78	15.07	9.51	3.52	1.42	11.42
Components' contributions to the MTWSE													
C	1.01	0.43	0.30	0.19	0.88	0.48	0.29	0.18	1.11	0.54	0.38	0.19	0.62
G	0.21	0.16	0.14	0.09	0.19	0.15	0.15	0.06	0.12	0.12	0.14	0.09	0.19
I	4.40	2.44	0.97	0.10	2.35	1.60	0.16	0.15	3.15	1.70	0.67	0.11	3.12
X	5.94	2.98	2.03	0.41	5.19	2.60	0.98	0.17	4.98	3.81	1.22	0.38	3.43
M	6.90	3.07	3.09	0.71	4.95	3.23	1.31	0.22	5.71	3.34	1.11	0.65	4.05

nents in percentage points, whereas the second, named Mean of Total Weighted Squared Error, provides a measure of the average squared forecast errors across all GDP expenditure components. The Mean of Total Weighted Squared Error shows that, even though 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. Forecasts issued by *Banco de Portugal* are generally better than those from other institutions, particularly at 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. Forecasts issued by the Portuguese Government Budget Office—only available at the 12-month span—are generally the least accurate. 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 government 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.

To our knowledge, this article is the first to provide an integrated evaluation of macroeconomic forecasts along three dimensions—across institutions, across time spans, and across GDP components. An application for Portugal is considered here; however, natural extensions arise, such as extending this analysis to G7 countries. Since these countries have naturally longer data spans, a more robust analysis involving econometric techniques and hypothesis testing can be developed.

References

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.

——— (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.

——— (1978): "The "rationality" of economic forecasts," *American Economic Review*, 68, 301–305.

——— (1986): "Forecasting accuracy of alternative techniques: a comparison of US macroeconomic forecasts," *Journal of Business and Economic Statistics*, 4, 5–15.

——— (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.

——— (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.

——— (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.

——— (1979): "An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level," *Journal of Business*, 52, 1–33.

——— (1984): "The accuracy of individual and group forecasts from business outlook surveys," *Journal of Forecasting*, 3, 11–26.