Further Results on MSFE Encompassing*

Massimiliano Marcellino
I.G.I.E.R., Università Bocconi
and
European University Institute

This Version: October 1998

Abstract

We show that the standard condition for MSFE encompassing is no longer valid when the forecasts to be compared are biased. We propose a simple modification of such a condition and of tests for its validity. The relationship between these tests, pooling regressions and tests for non-nested hypotheses is also analysed, together with their multivariate versions. The theoretical results are illustrated by an empirical example on inflation and deficit forecasts, key variables for the formulation of monetary and fiscal policy.

J.E.L. Classification: C52, C53

*I would like to thank Mike Artis, Clive Granger and, in particular, Anindya Banerjee for helpful comments on an earlier draft. The usual disclaimers apply.

1. Introduction

Several different forecasts of the same phenomenon are often available. The usual approach in this case is either to choose one of them on the basis of a certain criterion, such as minimum mean squared forecast error (MSFE) or mean absolute error (MAE), or to combine them in a pooled forecast, again with the aim of minimizing a certain loss function, see e.g. Clemen (1989).

An alternative approach is based on the so-called encompassing principle, see e.g. Mizon (1984), which was originally developed for in-sample model comparison, and later on extended to forecast comparison, starting with the seminal paper by Chong and Hendry (1986). Forecast encompassing requires one model to be able to correctly predict the forecasts of the competing models, i.e. the models that produce the alternative forecasts, or at least their implied MSFEs. In this case the competing models become redundant, and all the relevant information is contained in the encompassing model.

The evaluation of forecast encompassing requires knowledge of the forecasting models, or at least of the relationship among the competing forecasts, while for MSFE encompassing only the forecasts and the realizations are necessary. The larger information set is often not available when comparing, for example, forecasts from international organizations such as the IMF or the OECD, or forecasts from econometric models with consensus forecasts. These econometric models, even if available, can also be so complex that standard techniques for forecast encompassing, e.g., Ericsson (1992), become hardly applicable. Therefore, in this paper we will mainly focus on MSFE encompassing. When the forecasting models are available, the analysis of MSFE encompassing is anyway interesting, because this property is a necessary condition for the validity of stronger notions of forecast encompassing.

MSFE encompassing tests, namely tests for checking whether one model can correctly predict the MSFE arising from a competing model, were originally proposed by Chong and Hendry (1986), as mentioned, and were further analysed and extended by Lu and Mizon (1991), Ericsson (1992), Ericsson and Marquez (1993). They were applied, among others, by Andrews *et al.* (1996) for the comparison of unemployment, growth and inflation forecasts by three UK macroeconometric modelling groups, and by Artis and Marcellino (1998b) for IMF, EC, and OECD deficit forecasts. In this paper we review the available results on MSFE encompassing tests, and extend them.

In Section 2 we start by defining formally the condition for MSFE encompassing. We then show that such a condition is no longer valid when the forecasts under comparison are biased, and modify it properly. In this case, also the standard relationship between MSFE encompassing and MSFE dominance (the former implies the latter) no longer necessarily holds, as it is derived theoretically and illustrated with an example. We think that this is an important extension because forecast unbiasedness is often rejected in practice, which can be also justified theoretically if the loss function of the forecaster is asymmetric (e.g., Granger and Newbold (1986, Ch. 4)), if other goals rather than accuracy are important, e.g. publicity (Laster *et al.* (1997)), or if there are unaccounted structural changes over the forecast period (e.g. Clements and Hendry(1997c)).

In Section 3 we review standard MSFE encompassing tests, and indicate how they should be modified in the presence of biased forecasts. Next, we recall the role of pooling regressions for MSFE encompassing, and compare the relative merits of alternative specifications. Then, we show how a MSFE encompassing test can be derived as a Cox (1961) statistic, by exploiting the theory developed in Pesaran (1974).

In Section 4 several multivariate versions of the tests are described and compared, yielding a general framework for jointly testing for encompassing, unbiasedness, and efficiency.

In Section 5 we present an empirical example, in order to illustrate some of the theoretical results. We compare the IMF inflation and deficit forecasts for the G7 countries with random walk based forecasts. Notwithstanding their simplicity, the latter can be rather robust in the presence of structural changes (Clements and Hendry (1997b)), which are likely in the case of the deficit. Actually, for this variable the naive forecasts perform quite well, they even MSFE encompass the IMF forecasts for Japan.

Section 6 summarises and concludes.

2. MSFE Encompassing

In this section we define the notion of MSFE encompassing, and show how the conditions for its validity have to be modified when the forecasts under comparison are biased. We also analyze the relationship between MSFE dominance and encompassing, showing that it is also affected by biasedness of the forecasts.

We assume that only two univariate forecasting models are available, M^1 and M^2, their parameters are known and constant over the forecast period, the variable to be forecast, y, is stationary, and the forecast horizon is one period. These hypotheses are useful for focusing on the main topics, they will be relaxed later on. The forecasts for period t made at time $t-1$ by M^1 and M^2 are labelled $\widehat{y}_t^1$ and $\widehat{y}_t^2$, with $t = T+1, T+2, ..., T+N$.

For the reasons exposed in the Introduction, we will not require $\widehat{y}_t^1$ and $\widehat{y}_t^2$ to be unbiased, namely,

$$\begin{array}{lllll} M^1 \Rightarrow & y_t = a + b\widehat{y}_t^1 + u_{1t}, & u_{1t} \sim i.i.d.(0, \sigma_{1u}), & cov(u_{1t}, \widehat{y}_{t-i}^2) = 0 & i \geq 0, \\ M^2 \Rightarrow & y_t = c + d\widehat{y}_t^2 + u_{2t}, & u_{2t} \sim i.i.d.(0, \sigma_{2u}), & cov(u_{2t}, \widehat{y}_{t-i}^1) = 0 & i \geq 0. \end{array} \tag{2.1}$$

Unbiasedness requires no constant and a unit coefficient on the forecast, $(a = 0, b = 1)$ and/or $(c = 0, d = 1)$. Under unbiasedness u_{it} coincides with the forecast error $e_{it} = y_t - \widehat{y}_t^i$, $i = 1, 2$; otherwise they will be different. Notice also that each equation in (2.1) is a statement about the expectation of y conditional on $\widehat{y}_t^1$ and $\widehat{y}_t^2$, so that each model's proprietor assumes that the competing forecast is redundant.

From the decomposition

$$y_t - \widehat{y}_t^2 = (y_t - \widehat{y}_t^1) + (\widehat{y}_t^1 - \widehat{y}_t^2), \tag{2.2}$$

it follows that

$$MSFE^2 = MSFE^1 + E(\widehat{y}_t^1 - \widehat{y}_t^2)^2 + 2E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)]. \tag{2.3}$$

The prediction of M^1 for $MSFE^2$, $MSFE_1^2$, is

$$MSFE_1^2 = MSFE^1 + E(\widehat{y}_t^1 - \widehat{y}_t^2)^2 + 2E[(a + (b-1)\widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)]. \tag{2.4}$$

M^1 encompasses M^2 with respect to the MSFE if and only if $MSFE_1^2 = MSFE^2$. If $\widehat{y}_t^1$ is unbiased, the last term in the right hand side of (2.4) is equal to zero, and MSFE encompassing requires lack of correlation between the forecast error from one model and the difference of the two forecasts, i.e.,

$$E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] = 0. \tag{2.5}$$

If unbiasedness does not hold, it is still possible to have MSFE encompassing but the condition becomes

$$E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] = E[(a + (b-1)\widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)],$$

that can be rewritten as

$$E[u_{1t}(\widehat{y}_t^1 - \widehat{y}_t^2)] = 0, \tag{2.6}$$

namely, the error u_{1t} (no longer the forecast error $y_t - \widehat{y}_t^1$) must be uncorrelated with the difference of the forecasts.

Four outcomes are possible in a MSFE encompassing evaluation:

1. M^1 MSFE encompasses M^2.
2. M^2 MSFE encompasses M^1.
3. Neither 1 nor 2 hold.
4. Both 1 and 2 hold.

Cases 1 and 2 pose no problems. Case 3 indicates that both models are somewhat misspecified, and should be reformulated. This issue is further discussed later on in the context of pooling regressions. In case 4, M^1 and M^2 are said to be observationally equivalent with respect to the MSFE (see Mizon and Richard (1986)). It can be easily shown that when the forecasts are unbiased this happens if

$$E[\widehat{y}_t^2(\widehat{y}_t^2 - \widehat{y}_t^1)] = E[\widehat{y}_t^1(\widehat{y}_t^2 - \widehat{y}_t^1)], \tag{2.7}$$

i.e., the covariance of the two competing forecasts with the forecast difference is the same. It is quite unlikely that (2.7) holds in the population, but in finite samples the two covariances can be very close. If this is the case, the two models should be compared on the basis of other criteria, such as the mean absolute forecast error.

It is now worth analysing the relationship between MSFE dominance of M^1 ($\mathrm{MSFE}^1 < \mathrm{MSFE}^2$) and MSFE encompassing. When $\widehat{y}_t^1$ is unbiased, MSFE dominance is only a necessary condition for MSFE encompassing, and selecting a model according to this criterion does not ensure that the resulting forecast errors cannot be explained by the alternative forecasts. Instead, MSFE encompassing is a sufficient condition for MSFE dominance. Both propositions can be easily derived from a comparison of (2.3) and (2.4), see e.g. Ericsson (1992).

When $\widehat{y}_t^1$ is biased, MSFE dominance is no longer necessary for MSFE encompassing, and the latter is not sufficient for the former. Actually, from (2.4), it can be $MSFE_1^2 = MSFE^2$, i.e. M[1] MSFE encompasses M[2], but

$$E(\widehat{y}_t^1 - \widehat{y}_t^2)^2 + 2E[(a + (b-1)\widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] < 0,$$

which implies $MSFE^2 < MSFE^1$.

As an example, let us supplement (2.1) with a description of the relationship between the two forecasts,

$$\widehat{y}_t^2 = \alpha + \beta\widehat{y}_t^1 + \eta_t, \quad \eta_t \sim i.i.d.(0, \sigma_\eta), \tag{2.8}$$

with $cov(u_{1t}, \eta_t) = 0$. It is

$$E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] = E[(a + (b-1)\widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)],$$

so that $MSFE_1^2 = MSFE^2$, and M[1] MSFE encompasses M[2].[1] Notice that using the condition (2.5) to verify encompassing, we would reject it because it is $E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] \neq 0$. Then we have,

$$\begin{aligned} MSFE^2 - MSFE^1 &= E(\widehat{y}_t^1 - \widehat{y}_t^2)^2 + 2E[(a + (b-1)\widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] = \\ &= E[(1-\beta)\widehat{y}_t^1 - \alpha - \eta_t]^2 + 2E[(a + (b-1)\widehat{y}_t^1)((1-\beta)\widehat{y}_t^1 - \alpha - \eta_t)] = \\ &= E(\widehat{y}_t^1)^2(1-\beta)(2b-1-\beta) + 2E(\widehat{y}_t^1)[a(1-\beta) - \alpha(b-\beta)] + \\ &\quad +\alpha^2 + \sigma_\eta - 2a\alpha. \end{aligned}$$

This quantity can be larger or smaller than zero, e.g., it is $MSFE^2 < MSFE^1$ for $2b - 1 < \beta < 1$, $\alpha = a(1-\beta)/(b-\beta)$, $\sigma_\eta < \alpha(2a - \alpha)$.

To conclude, it can be worth stressing that MSFE encompassing provides a measure of the relative performance of a model, not of its overall goodness in forecasting, so that the latter should be separately assessed.

[1] Instead, $E[(y_t - \widehat{y}_t^2)(\widehat{y}_t^2 - \widehat{y}_t^1)] \neq E[(c + (d-1)\widehat{y}_t^2)(\widehat{y}_t^2 - \widehat{y}_t^1)]$, so that M[2] does not MSFE encompass M[1].

3. Univariate MSFE encompassing tests

In this section we review existing tests for MSFE encompassing, and suggest how they should be modified to take forecast biasedness into account. We then relate them to pooling regressions, and evaluate whether biasedness poses additional problems also in this framework. Next, we show how classical tests for non-nested hypotheses can be adapted to test for MSFE encompassing. Finally, we indicate how some of the simplifying assumptions that we maintain for clarity can be relaxed.

3.1. MSFE encompassing tests

One of the earliest attempts to provide a statistical tool for choosing a forecasting formula, Hoel (1947), is based on the significance of the regressor in the model

$$y_t - \widehat{y}_t^1 = \phi(\widehat{y}_t^2 - \widehat{y}_t^1) + u_t, \quad u_t \sim i.i.d.(0, \sigma_u). \tag{3.1}$$

The underlying idea is that when $\phi = 0$ the forecast error from M^1 cannot be explained by M^2, which is therefore redundant. From (2.5), a test for $\phi = 0$ is also a MSFE encompassing test, as noticed by Ericsson (1992) and Clements and Hendry (1993). Yet, from (2.6), this statement is correct only if $\widehat{y}_t^1$ is unbiased, i.e., $a = 0$, $b = 1$ in (2.1). When this hypothesis does not hold, in order to construct a MSFE encompassing test based on the hypothesis $\phi = 0$, the regression should be modified into

$$\widehat{u}_{1t} = \phi(\widehat{y}_t^2 - \widehat{y}_t^1) + u_t, \quad u_t \sim i.i.d.(0, \sigma_u), \tag{3.2}$$

where $\widehat{u}_{1t}$ is the estimated counterpart of u_{1t}. This regression is sometimes run in empirical analysis, see e.g. Artis (1988), and the discussion in the previous section provides a theoretical rationale for it.

Chong and Hendry (1986) instead considered an equation similar to (3.1), namely,

$$y_t - \widehat{y}_t^1 = \xi \widehat{y}_t^2 + u_t, \tag{3.3}$$

and proposed to test for $\xi = 0$ for MSFE encompassing. They showed that the t-statistic for $\xi = 0$ has a $N(0, 1)$ distribution for large T and N. This is also the suggestion by Ericsson (1992) in the case of stationary variables. Yet, again, this is a MSFE encompassing test only if $\widehat{y}_t^1$ is unbiased. Otherwise, the dependent variable should be substituted by $\widehat{u}_{1t}$.

Continuing the example in the previous section, when the forecasts are biased, even under MSFE encompassing of M^1 for M^2 both the OLS estimator of ϕ in (3.1) and that of ξ in (3.3) converge to non zero values, because $E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)]$ and $E[(y_t - \widehat{y}_t^1)\widehat{y}_t^2]$ are different from zero. Instead, using $\widehat{u}_{1t}$ as the dependent variable in either (3.1) or (3.3), the OLS estimator of the coefficient of the regressor converges to zero, correctly indicating that M^1 MSFE encompasses M^2.

3.2. Pooling regressions

Some of the aforementioned drawbacks of standard MSFE encompassing tests can be also avoided by adopting the model

$$y_t = e + f\widehat{y}_t^1 + g\widehat{y}_t^2 + u_t, \tag{3.4}$$

which nests both equations in (2.1). Equation (3.4) has been often employed in the literature on forecast pooling, see, e.g., Clemen (1989) and Wallis (1989), because the estimates of the parameters provide the optimal weights in the sense of minimising the MSFE for the pooled forecast. The hypothesis $g = 0$ corresponds to M^1 encompasses M^2 with respect to MSFE. When $e = 0$, $f = 1$ the forecasts from M^1 are also unbiased. This restriction leads to (3.3), while $e = 0$, $f + g = 1$ (which holds when at least one forecast is unbiased) leads to (3.1). It can be easily verified that in the case of the example in the previous section the OLS estimators of e, f, and g converge, respectively, to a, b, and zero.

The benefits from not imposing $f + g = 1$ are emphasised by Fair and Shiller (1990). However, they use a different version of (3.4), namely,

$$y_t - y_{t-1} = c + \xi(\widehat{y}_t^1 - y_{t-1}) + \eta(\widehat{y}_t^2 - y_{t-1}) + u_t. \tag{3.5}$$

Paradoxically, when $\xi + \eta \neq 1$ this formulation appears to be inappropriate (because of non cancellation of the y_{t-1} terms), and can lead to biased estimators of ξ and η. Let us consider an example just by Fair and Shiller (p.377). They assume, in our notation, that y is generated as the sum of two orthogonal variables plus noise and that each of M^1 and M^2 lets y depend on one of the two variables only. Hence, we should have $e = 0$ and $f = g = 1$ in equation (3.4). However, the estimators of ξ and η in (3.5) cannot both converge to one, otherwise there would be $-y_{t-1}$ on the left hand side of (3.5) and $-2y_{t-1}$ on the right hand side. Hence, the formulation in (3.4) seems still preferable.

Equation (3.4) is also advantageous when many models have to be compared, say n, because one regression instead of n type-(3.3) regressions is sufficient. Yet, the probable high correlation among the regressors can create problems of collinearity. In such a case a change of regressors as in

$$y_t = e + (f+g)\widehat{y}_t^1 + g(\widehat{y}_t^2 - \widehat{y}_t^1) + u_t \tag{3.6}$$

might be required.[2]

The encompassing-unbiasedness hypotheses have also strong implications in terms of MSFE dominance. In particular, ($e = 0$, $f = 1$, $g = 0$) implies that M^1 MSFE dominates M^i, where M^i is whatever model whose forecasts can be expressed as a linear combination of those by M^1 and M^2. Granger and Newbold (1973) have suggested that a forecasting model should have conditional efficiency equal to one, where conditional efficiency is defined as the ratio of the sampling variances of the forecast errors of the pooled and original models. This requirement is therefore satisfied when ($e = 0$, $f = 1$, $g = 0$). This hypothesis also implies the equality to zero of the first two terms in the Theil (1958) decomposition

$$MSFE = (\bar{\widehat{y}}^1 - \bar{y}) + (S_{\widehat{y}^1} - rS_y)^2 + (1-r^2)S_y^2,$$

where $\bar{\widehat{y}}^1$ and $\bar{y}$ are the sample means of predictor and predicted series, $S_{\widehat{y}^1}$ and S_y are their standard deviations, and r is the sample correlation between them.

To conclude, it can be worth commenting briefly on the role of forecast pooling and model rispecification when both regressors are significant in (3.4) or (3.6), see Chong and Hendry (1986), Diebold (1989), Ericsson (1992) for further discussion. The first option is current practice in applied forecasting, but this outcome is in general related to a misspecification of both models, which should therefore be changed somehow before forecasting. However, when M^1 and M^2 are large macroeconometric models it seems difficult to modify them quickly, so that forecast pooling can be a second best. Moreover, in certain cases, see e.g.

[2] An alternative approach is to extract the common infornation in the n competing forecasts using principal component techniques, and use it as a forecast. See Chang *et al.* (1998) for details.

the aforementioned example by Fair and Shiller (1990), pooled forecasts could be seen as the forecasts by a completing-type model which could alleviate the misspecification problems of the competing models.

3.3. Non-nested tests

The problem of comparing M^1 and M^2 with respect to their forecasting performance can be also cast in the non-nested hypothesis testing framework, the non nested models being those in (2.1). Under the additional hypothesis of normal errors, the numerator of the Cox test, as explicitly derived by Pesaran (1974) for the linear regression case[3], is proportional to the difference of the log of the (estimated) variance of u_2, $\widehat{\sigma}_{u2}$, and its prediction made by M^1, $\widehat{\sigma}_{u2\,1}$. Under the assumption $a = 0$, $b = 1$, $c = 0$, $d = 1$, the non-nested models to be compared become

$$\begin{array}{lll} M^1 \Rightarrow & y_t = \widehat{y}_t^1 + u_{1t}, & u_{1t} \sim i.i.d.N(0, \sigma_{1u}), \\ M^2 \Rightarrow & y_t = \widehat{y}_t^2 + u_{2t}, & u_{2t} \sim i.i.d.N(0, \sigma_{2u}). \end{array} \tag{3.7}$$

In this case, following Pesaran (1974), the Cox statistic for the first model versus the second one can be written as

$$CP = \frac{\frac{N}{2} log(MSFE^2/MSFE_1^2)}{\sqrt{\frac{MSFE^1}{(MSFE_1^2)^2}(\widehat{y}^{2\prime}(I - \widehat{y}^1(\widehat{y}^{1\prime}\widehat{y}^1)^{-1}\widehat{y}^{1\prime})\widehat{y}^2)}},$$

where $\widehat{y}^1$ and $\widehat{y}^2$ are the $N \times 1$ vectors of forecasts from M^1 and M^2. The statistic is distributed as $N(0, 1)$ for large T and N. Its interpretation as a MSFE encompassing test is straight forward. Given that $MSFE^2/MSFE_1^2 = 1$ if and only if $E[(y_t - \widehat{y}_t^1)(\widehat{y}_t^1 - \widehat{y}_t^2)] = 0$, the tests for $\phi = 0$ in (3.1) or $\xi = 0$ in (3.3) provide simpler alternatives.

As an alternative to Cox procedure, Atkinson (1970) suggested embedding the non-nested models into a nesting framework, by taking a weighted combination of them. The test is then that the weight associated to one model is equal to zero. In our case the nesting framework is provided by equation (3.4), and the test is the t-statistic for $g = 0$, which again admits an immediate MSFE encompassing interpretation. Under the hypothesis $e = 0$, $f + g = 1$ this statistic also corresponds to the out-of-sample version of the C test by Davidson and MacKinnon (1981).

[3] See also MacKinnon (1983) for some simplifications.

3.4. Some extensions

So far we have assumed that the error terms in the regressions of interest are i.i.d. variables, e.g., (2.1), (3.2), or (3.4), we have maintained that the variables to be forecast are stationary, and the parameters of the forecasting models remain stable over the forecast period. In practice, non linearity in the forecasting models and h-step ahead forecasting, $h > 1$, can induce heteroskedasticity and correlation in the error terms; economic variables are often nonstationary processes, in particular they can be well represented as processes integrated of order one ($I(1)$); and structural breaks do happen, and are one of the major sources of forecast failures (see e.g. Clements and Hendry (1997b)). These three problems have been already tackled in the literature, so that we only briefly address them here.

In the case of heteroskedastic correlated errors, OLS estimation with proper residual variance estimators and related robust tests can be adopted (see e.g. Fair and Shiller (1989), White (1980)), or a feasible GLS estimation method can be used (Ericsson and Marquez (1993)).

When the variables are integrated, a first requisite for the competing forecasts is to be of the same order of integration as the variables they are referred to. A second basic requirement is that they cointegrate with the variables. If any of these two basic properties is not satisfied, it is not worth going on with the comparison. As far as MSFE encompassing tests are concerned, $I(1)$-ness increases the rate of convergence of the estimators in level regressions, but complicates inference because standard distribution theory no longer holds (Sims *et al.* (1990)). Moreover, care as to be exerted to avoid running unbalanced or misspecifed regressions when some kind of stationarity transformation is adopted. In particular, as noticed by Ericsson (1992), equation (3.3) no longer provides a proper framework for a MSFE encompassing test because just when $\widehat{y}_t^1$ is an unbiased forecast, which we recall is the condition for (3.3) to be valid with stationary variables, the stationary variable $y - \widehat{y}_t^1$ is regressed on the $I(1)$ regressor $\widehat{y}_t^2$.

To avoid this problem, Ericsson (1992) suggested to adopt equation (3.1). Yet, this is a proper choice only if both $\widehat{y}_t^1$ *and* $\widehat{y}_t^2$ are unbiased, which guarantees stationarity both of the dependent and of the independent variables. Moreover, the gains from a faster rate of convergence of the estimators are lost. A preferable formulation seems again equation (3.4), combined with the distribution theory by

Phillips and Hansen (1990) to test for the hypotheses of MSFE encompassing and unbiasedness of $\widehat{y}_t^1$. As an alternative, we suggest to reparametrize (3.4) into the ECM formulation

$$\Delta y_t = e + f\Delta\widehat{y}_t^1 + g\Delta\widehat{y}_t^2 + h(y_{t-1} - b\widehat{y}_{t-1}^1) + k(y_{t-1} - d\widehat{y}_{t-1}^2) + u_t. \qquad (3.8)$$

M^1 MSFE encompasses M^2 for $f = b$, $g = 0$, $h = -1$,$k = 0$, and viceversa for $f = 0$, $g = d$, $h = 0$,$k = -1$. Tests for these hypotheses now have a standard distribution. Yet, the tests run in the regression in levels (equation (3.4)) can handle more general error terms, while the hypothesis of i.i.d. errors is important in the ECM (3.8).

Finally, structural changes are only a problem if they are not properly taken into account in the forecasting models, e.g. by means of some type of intercept corrections, and therefore affect the relationship between actual and fitted values. Whether this is the case can be checked by means of tests for parameter constancy in the regressions of subsections 3.1 and 3.2, e.g. Hansen (1992); but if non-constancy is detected it is difficult to prescribe a general remedy.

4. Multivariate MSFE encompassing tests

We now assume that x is a $n \times 1$ vector of variables whose forecasts by M^1 and M^2 are the $n \times 1$ vectors $\widehat{x}_t^1$ and $\widehat{x}_t^2$, with

$$\begin{array}{lllll} M^1 \Rightarrow & x_t = a + B\widehat{x}_t^1 + u_{1t}, & u_{1t} \sim i.i.d.(0, \Omega_{1u}), & cov(u_{1t}, \widehat{x}_{t-i}^2) = 0 & i \geq 0, \\ M^2 \Rightarrow & x_t = c + D\widehat{x}_t^2 + u_{2t}, & u_{2t} \sim i.i.d.(0, \Omega_{2u}), & cov(u_{2t}, \widehat{x}_{t-i}^1) = 0 & i \geq 0, \end{array} \qquad (4.1)$$

where a, c are $n \times 1$ vectors while B, D are $n \times n$ matrices. To start with, we wish to discuss alternative tests for trace MSFE encompassing, where the MSFE matrix implied by the two models is

$$\Phi^i = \begin{pmatrix} E(e_{1t}^i)^2 & E(e_{1t}^{i'} e_{2t}^i) & ... & E(e_{1t}^{i'} e_{nt}^i) \\ E(e_{1t}^{i'} e_{2t}^i) & E(e_{2t}^i)^2 & ... & E(e_{2t}^{i'} e_{nt}^i) \\ ... & & & \\ E(e_{1t}^{i'} e_{nt}^i) & E(e_{2t}^{i'} e_{nt}^i) & ... & E(e_{nt}^i)^2 \end{pmatrix},$$

$e_{jt}^i = x_{jt} - \widehat{x}_{jt}^i$, $j = 1, ..., n$, and $i = 1, 2$.

Trace MSFE encompassing of M^1 for M^2 requires $tr(\Phi^2) = tr(\Phi_1^2)$, where Φ_1^2, the prediction of M^1 for Φ^2, is

$$\Phi_1^2 = \Phi^1 + E(\widehat{x}_t^1 - \widehat{x}_t^2)(\widehat{x}_t^1 - \widehat{x}_t^2)' + 2E[(a + (B - I)\widehat{x}_t^1)(\widehat{x}_t^1 - \widehat{x}_t^2)'], \tag{4.2}$$

while Φ^2 can be written as

$$\Phi^2 = \Phi^1 + E(\widehat{x}_t^1 - \widehat{x}_t^2)(\widehat{x}_t^1 - \widehat{x}_t^2)' + 2E(x_t - \widehat{x}_t^1)(\widehat{x}_t^1 - \widehat{x}_t^2)'. \tag{4.3}$$

Hence, to have trace MSFE it must be

$$tr(E(x_t - \widehat{x}_t^1)(\widehat{x}_t^1 - \widehat{x}_t^2)') = tr(E[(a + (B - I)\widehat{x}_t^1)(\widehat{x}_t^1 - \widehat{x}_t^2)']). \tag{4.4}$$

It immediately follows that when the forecasts in $\widehat{x}_t^1$ are unbiased and B is diagonal, the condition simplifies to

$$tr(E(x_t - \widehat{x}_t^1)(\widehat{x}_t^1 - \widehat{x}_t^2)') = 0. \tag{4.5}$$

Trace MSFE dominance is often employed as a tool for selecting a multivariate forecasting model. However, as in the univariate case, when the forecasts are unbiased it is necessary but not sufficient for trace MSFE encompassing. The latter implies trace MSFE dominance, because $E(\widehat{x}_t^1 - \widehat{x}_t^2)(\widehat{x}_t^1 - \widehat{x}_t^2)'$ is a symmetric matrix so that its trace is non negative. These relationships no longer necessarily hold when the forecasts are biased. Notice also that while variable by variable MSFE dominance and encompassing imply trace MSFE dominance and encompassing, the converse is not necessarily true.

A first possible test for trace MSFE encompassing requires to run n type-(3.4) regressions, namely,

$$x_{jt} = e + f\widehat{x}_{jt}^1 + g\widehat{x}_{jt}^2 + u_{jt}, \quad j = 1, ..., n, \tag{4.6}$$

and test the hypothesis $g = 0$ in each of them. In order to have an overall size of α, the size of each test should be equal to α/n. (4.6) is a system of seemingly unrelated regression equations (SURE, see Zellner (1962)), and Nelson (1972) proposed to estimate it by means of GLS in order to obtain more efficient estimators than OLS. Trace MSFE encompassing is accepted when the hypothesis $g = 0$ is accepted in each equation. Yet, as mentioned, this is a sufficient but not necessary condition for trace MSFE encompassing, so that in particular cases the latter could be wrongly rejected.

In order to avoid this problem, the n variables and forecasts can be stacked into the $nN \times 1$ vectors X, $\widehat{X}^1$, $\widehat{X}^2$. These are then used in the regression

$$X_l = e + f\widehat{X}_l^1 + g\widehat{X}_t^2 + U_l, \quad l = 1, ..., nN, \tag{4.7}$$

and a t-test for $g = 0$ is performed. In general, the error term in (4.7) is heteroskedastic. The varying variance can be estimated from the system in (4.6), so that GLS estimation is feasible. Under the null hypothesis, using (4.7) instead of (4.6) can be also advantageous because of the larger number of available observations. Yet, we are imposing that the values of the constant and of the coefficient of the first forecast are equal for all variables, an hypothesis that can be relaxed by inserting proper dummies in the regression, at the cost of loosing degrees of freedom. Individual properties such as forecast unbiasedness and efficiency could then also be tested in this framework.

As far as efficiency is concerned, in (4.7) and in (4.6) we are assuming that the forecasts of other variables from the two models are not relevant explanatory variables. This hypothesis can be also relaxed, considering the system of equations

$$x_t = c + \Xi\widehat{x}_t^1 + H\widehat{x}_t^2 + v_t, \tag{4.8}$$

where Ξ and H are $n \times n$ matrices of parameters and c is an $n \times 1$ vector of constants. (4.8) boils down to (4.6) when Ξ and H are diagonal. The joint hypothesis of efficiency and encompassing requires therefore $\Xi = diag$, $H = 0$ while if $c = 0$, $\Xi = I$ the forecasts from M^1 are also unbiased. Notice that in this case encompassing is with respect to Φ^2, the MSFE matrix for M^2, which implies trace MSFE encompassing.

When $\Xi + H = I$, which holds for example when $\widehat{x}^1$ is unbiased, (4.8) can be re-written as

$$x_t - \widehat{x}_t^1 = c + H(\widehat{x}_t^2 - \widehat{x}_t^1) + v_t. \tag{4.9}$$

A test for $H = 0$ in (4.9) is invariant to isomorphic dynamic transformations of the underlying system for the x variables and corresponding forecasts, while this is not true in (4.8). This property follows from invariance of the 1-step ahead forecast errors to these transformations, and it also holds for the univariate case, see Clements and Hendry (1993). Under the null hypothesis, the test is also invariant to contemporaneous linear transformations (Clements and Hendry (1997a, Ch. 10.3)).

To conclude, Clements and Hendry (1993) suggest to use the determinant of the MSFE matrix for forecast comparisons instead of its trace, because the former is invariant to both dynamic and linear transformations.[4] Φ encompassing is sufficient but not necessary for determinant MSFE encompassing, so that it seems interesting to consider a specific test for the latter property. We propose to use the multivariate version of the Pesaran (1974) test, which was developed by Pesaran and Deaton (1978). We provide an explicit formula for the case where both sets of forecasts are unbiased, but this assumption can be easily relaxed at the cost of further notational complexity. The models to be compared are

$$\begin{array}{lll} M^1 \Rightarrow & x_t = \widehat{x}_t^1 + u_{1t}, & u_{1t} \sim i.i.d.N(0, \Phi^1), \\ M^2 \Rightarrow & x_t = \widehat{x}_t^2 + u_{2t}, & u_{2t} \sim i.i.d.N(0, \Phi^2), \end{array}$$

and the test statistic is

$$PD = \frac{\frac{N}{2} \log \frac{|\widehat{\Phi}^2|}{|\widehat{\Phi}_1^2|}}{\sqrt{V}} \overset{a}{\sim} N(0,1), \tag{4.10}$$

where

$$\begin{aligned} V &= (\widehat{X}^1 - \widehat{X}^2)' \{ ((\widehat{\Phi}_1^2)^{-1} \widehat{\Phi}^1 (\widehat{\Phi}_1^2)^{-1} \otimes I) + \\ & \quad -((\widehat{\Phi}_1^2)^{-1} \otimes I) \widehat{X}^1 \Psi^{-1} \widehat{X}^{1\prime} ((\widehat{\Phi}_1^2)^{-1} \otimes I) \} (\widehat{X}^1 - \widehat{X}^2), \\ \Psi &= \widehat{X}^{1\prime} ((\widehat{\Phi}_1^2)^{-1} \otimes I) \widehat{X}^1. \end{aligned}$$

5. An empirical example

In this section we present a simple empirical example in order to illustrate some of the previous theoretical results. We analyse the (yearly) deficit to gdp ratio and the inflation forecasts from the IMF for the G7 countries, over the period 1975-1994. These are two important variables for fiscal and monetary policy, and it seems therefore important to evaluate how accurate their forecasts are.

We consider current year forecasts, in the terminology e.g. of Artis (1988, 1997), namely those for period t published in the May issue of year t of the *World Economic Outlook*. The actual data are the first released values, which appear

[4]In the case of h-step ahead forecasting, they suggest to use the determinant of the second moment matrix of the stacked j-step ahead forecast errors, $j = 1, ..., h$.

in the May issue of year $t+1$ of the *Outlook*. A comparison with the OECD and EC deficit forecasts is presented in Artis and Marcellino (1998a, 1998b). Here we use as alternative forecasts those based on a random walk model for the variables, which therefore coincide with the actual values for year $t-1$ (which become available at the same time as the IMF forecasts). Hence, $\widehat{y}_t^1 = \widehat{y}_t^{IMF}$, $\widehat{y}_t^2 = y_{t-1}$. We can anticipate that, notwithstanding the likely misspecification of the random walk model (the persistence of most variables is rather low), and the larger information set embodied in the IMF forecasts, the alternative naive forecasts seem to perform rather well in some cases. We argue that this is due to the robustifying role of differencing in the presence of structural changes (see e.g. Clements and Hendry (1997b)).

In Table 5.1 we summarise the results for the deficit to gdp ratio forecasts.[5] The second column presents the ratio of the root MSFE of the IMF forecasts to that of the random walk (RW) forecasts. It is larger than one for Japan, Germany, and Italy, indicating that for these countries $\widehat{y}_t^2$ MSFE dominates $\widehat{y}_t^1$. The third and fourth columns report tests for unbiasedness of the forecasts. These are t-tests for $c=0$ and $d=0$ in the regression

$$y_t - \widehat{y}_t^i = c + d\widehat{y}_t^i + u_t, \qquad i = IMF, RW. \tag{5.1}$$

Unbiasedness is rejected in several cases for the IMF forecasts, with the exception of Canada and UK, while the RW seems to perform better in this respect with rejection only for Canada, Japan, and Germany.

These results suggest that the standard versions of the MSFE encompassing tests are not appropriate, and they should be modified as, e.g., in (3.2). The next four columns in Table 5.1 report the standard and modified test statistics for the IMF and RW forecasts, based respectively on the regression models (3.1) and (3.2). From the fifth column, the IMF forecast errors can be explained by the forecast difference $\widehat{y}_t^2 - \widehat{y}_t^1$ for Japan, Germany and Italy, the countries where the RW forecasts MSFE dominate. Yet, from the seventh column, when the residuals from equation (5.1) are used as dependent variables, the forecast

[5] All the calculations were performed with PcGive and PcFiml 9.01, see Hendry and Doornik (1997), Doornik and Hendry (1997). Detailed results are available upon request. In particular, the residuals of the regression models used when testing for unbiasedness and encompassing in general pass diagnostic checks for no autocorrelation, homoskedasticity, and normality, and the parameters are stable over time.

difference remains significant only in the case of Japan. Thus, the IMF MSFE encompasses the RW *also* for Germany and Italy. Instead, there are no changes in the significance of the statistics when testing whether the RW MSFE encompasses the IMF. This is the case for Japan, Germany, and Italy.

The last two columns of Table 5.1 report the t-tests for the significance of $\widehat{y}_t^1$ and $\widehat{y}_t^2$ as explanatory variables of y_t in equation (3.4). The results are in agreement with those from the modified tests. The two forecasts are never jointly significant, $\widehat{y}_t^2$ is relevant only for Japan, while none of them is significant for Germany and Italy. The outcome for Italy could be due to collinearity problems, the correlation among the two forecasts is 0.90 while it is 0.27 for Germany. This problem could be solved by adopting the regression in (3.6); the correlation between the regressors drops to 0.61, but the result of the test does not change.

The IMF performs much better in forecasting inflation. From Table 5.2, it always MSFE dominates the RW. Moreover, the forecasts are unbiased for all countries, while bias arises for Japan, Germany, and UK in the case of the RW. The values of the standard and modified MSFE encompassing tests are quite close for the IMF, which has to be the case because of unbiasedness. The IMF MSFE encompasses the RW for all countries, apart from US and Canada. The RW never encompasses the IMF. These results are confirmed by the joint tests, which also highlight the potential benefits in terms of MSFE reduction from pooling the IMF and RW forecasts for US and Canada.

We can now move to the joint analysis of the deficit to gdp ratio and inflation forecasts. The second and third columns of Table 5.3 report the ratio of the trace and determinant of the MSFE matrix, Φ, for the IMF to those for the RW. The IMF does better in all cases; the worse performance in some deficit forecasts is more than compensated by the good one in forecasting inflation. The fourth column reports the test for efficiency, i.e., for the irrelevance of the deficit forecasts in explaining inflation, and viceversa. Formally, it is a Wald test for Ξ and H diagonal in (4.8), an hypothesis that is accepted for all countries

The fifth and sixth columns present the joint test for efficiency and Φ encompassing. The IMF Φ encompasses the RW in all cases ($H = 0$ in (4.8)), apart from Japan and Germany, and therefore it also encompasses RW with respect to the trace and determinant of Φ. The RW never encompasses the IMF. The last column of Table 5.3 contains the test for the joint hypothesis of efficiency, encompassing, and unbiasedness of both IMF forecasts. It is accepted for France

Table 5.1: Deficit forecasts.

	RMSFE	*Bias*		*Encompassing*					
				Standard		Modified		Joint	
	IMF/RW	IMF	RW	IMF	RW	IMF	RW	IMF	RW
US	0.81	$t_c = \mathbf{-2.90}$ $t_f = \mathbf{-3.82}$	$t_c = 1.51$ $t_f = 1.70$	-1.31	**3.64**	0.86	**5.18**	**5.21**	-1.15
CAN	0.79	$t_c = -1.65$ $t_f = -1.10$	$t_c = \mathbf{2.60}$ $t_f = \mathbf{2.62}$	-0.15	**3.96**	-0.01	**2.39**	**2.84**	-0.04
JAP	1.88	$t_c = -1.19$ $t_f = \mathbf{-4.26}$	$t_c = 1.50$ $t_f = \mathbf{2.35}$	$\mathbf{-4.81}$	1.14	$\mathbf{-2.44}$	1.91	1.90	**2.53**
GER	1.11	$t_c = \mathbf{-4.31}$ $t_f = \mathbf{-4.57}$	$t_c = \mathbf{3.10}$ $t_f = \mathbf{3.37}$	$\mathbf{-3.30}$	0.38	-1.02	0.28	1.02	0.84
FR	0.81	$t_c = \mathbf{-2.15}$ $t_f = \mathbf{-2.01}$	$t_c = 1.49$ $t_f = 1.20$	-1.50	**3.40**	-0.65	**3.62**	**3.51**	0.63
IT	1.42	$t_c = \mathbf{-4.28}$ $t_f = \mathbf{-4.99}$	$t_c = 1.60$ $t_f = 1.66$	$\mathbf{-3.69}$	0.38	0.10	1.51	1.92	-0.35
UK	0.83	$t_c = 0.11$ $t_f = -1.68$	$t_c = 1.18$ $t_f = 1.54$	-0.60	**3.94**	-0.21	**3.36**	**3.50**	0.34

"Bias": t-tests for $c = 0$ and $d = 0$ (t_c and t_f) in the regression (5.1).
"Standard" IMF: t-test for $\phi = 0$ in equation (3.1) with IMF=1, RW=2.
"Standard" RW: t-test for $\phi = 0$ in equation (3.1) with IMF=2, RW=1.
"Modified" IMF: t-test for $\phi = 0$ in equation (3.2) with IMF=1, RW=2.
"Modified" RW: t-test for $\phi = 0$ in equation (3.2) with IMF=2, RW=1.
"Joint" IMF (RW): t-test for $f = 0$ ($g = 0$) in equation (3.4).
Significant values are reported in boldface.

Table 5.2: Inflation forecasts.

	RMSFE	*Bias*		*Encompassing*					
				Standard		Modified		Joint	
	IMF/RW	IMF	RW	IMF	RW	IMF	RW	IMF	R
US	0.31	$t_c = -0.16$ $t_f = 0.13$	$t_c = 1.27$ $t_f = -1.54$	**2.98**	**−14.7**	**2.92**	**−10.4**	**13.7**	**−3**
CAN	0.65	$t_c = -0.88$ $t_f = 1.23$	$t_c = 0.91$ $t_f = -1.20$	**2.78**	**−6.91**	**3.56**	**−4.64**	**7.83**	**−3**
JAP	0.53	$t_c = -0.44$ $t_f = -0.511$	$t_c = 1.14$ $t_f = \mathbf{-2.03}$	1.27	**−7.88**	1.42	**−5.77**	**6.97**	−1
GER	0.49	$t_c = -0.62$ $t_f = 0.56$	$t_c = \mathbf{2.02}$ $t_f = \mathbf{-2.57}$	0.11	**−7.98**	0.02	**−4.23**	**6.43**	−0
FR	0.72	$t_c = 0.79$ $t_f = -0.29$	$t_c = 0.19$ $t_f = -0.47$	−1.51	**−5.07**	−1.43	**−4.86**	**4.90**	1.
IT	0.59	$t_c = 1.67$ $t_f = -1.03$	$t_c = 0.80$ $t_f = -0.99$	0.21	**−6.76**	0.53	**−6.69**	**6.64**	−0
UK	0.35	$t_c = -0.13$ $t_f = 0.06$	$t_c = 1.51$ $t_f = \mathbf{-2.00}$	0.48	**−12.5**	0.19	**−6.35**	**11.3**	−0

"Bias": t-tests for $c = 0$ and $d = 0$ (t_c and t_f) in the regression (5.1).
"Standard" IMF: t-test for $\phi = 0$ in equation (3.1) with IMF=1, RW=2.
"Standard" RW: t-test for $\phi = 0$ in equation (3.1) with IMF=2, RW=1.
"Modified" IMF: t-test for $\phi = 0$ in equation (3.2) with IMF=1, RW=2.
"Modified" RW: t-test for $\phi = 0$ in equation (3.2) with IMF=2, RW=1.
"Joint" IMF (RW): t-test for $f = 0$ ($g = 0$) in equation (3.4).
Significant values are reported in boldface.

Table 5.3: Deficit and Inflation forecasts, joint analysis.

	$tr(\Phi)$	$\|\Phi\|$	*Eff.*	*Eff+Enc.*		*Eff.+Enc.+Unb.*
	IMF/RW	IMF/RW		IMF	RW	
US	0.24	0.08	7.46	10.1	**197**	**39.6**
CAN	0.47	0.61	7.95	11.0	**54.8**	**22.9**
JAP	0.49	0.99	1.85	**21.8**	**14.2**	–
GER	0.40	0.28	7.62	**13.9**	**103**	–
FR	0.54	0.36	4.64	9.04	**49.7**	13.4
IT	0.72	0.76	8.04	9.19	**176**	**49.8**
UK	0.17	0.09	4.99	5.11	**95.0**	12.3

Eff.: Wald test ($\chi^2(4)$) for Ξ and H diagonal in (4.8).
Eff.+Enc. IMF: Wald test ($\chi^2(6)$) for Eff. and $H = 0$ in (4.8).
Eff.+Enc. RW: Wald test ($\chi^2(6)$) for Eff. and $\Xi = 0$ in (4.8).
Eff.+Enc.+Unb.: Wald test ($\chi^2(10)$) for $c = 0$, $\Xi = I$, $H = 0$ in (4.8).
Significant values are reported in boldface.

and UK.

To conclude, notice that as a consequence of efficiency the system in (4.8) is made up of two equations like (3.4) for the deficit ratio and inflation. Yet, the results on encompassing and unbiasedness from the joint analysis differ from those from the univariate analysis. In particular, encompassing was rejected for the US and Canada in the case of inflation forecasts, as well as unbiasedness of the French IMF deficit ratio forecasts. Strictly speaking, the results of the univariate and multivariate tests cannot be compared because of the different null hypotheses and distributions of the statistics. Yet, in our case, the partly mismatching conclusions are likely due to the validity of only some of the components of the joint hypotheses under analysis. Overall, the univariate and multivariate approaches can be considered as complementary rather than substitute.

6. Conclusions

MSFE encompassing is an important and easily testable property. Hence, testing for its validity should become a first step in forecast comparison exercises. We have extended the standard definitions and testing procedures to let the forecasts under comparison be biased, that is quite common in practice. We have also clarified the relationship with non-nested tests, and introduced Cox-type MSFE encompassing tests. The empirical analysis supports the practical usefulness of these generalizations.

Similar extensions could be developed for other notions of forecast encompassing, which is relevant when the forecasting models are known and of manageable size. This represents an interesting subject for future research.

References

[1] Andrews, M.J., Minford, A.P.L., and Riley, J. (1996), "On Comparing Macroeconomic Models Using Forecast Encompassing Tests", *Oxford Bulletin of Economics and Statistics*, 58, 279-305.

[2] Artis, M. (1988), "How accurate is the World Economic Outlook? A post-mortem on short term forecasting at the International Monetary Fund", *IMF Staff Studies Behind the World Economic Outlook*, 1-49.

[3] Artis, M. (1997), "How accurate are the IMF's short term forecasts? Another examination of the World Economic Outlook", *IMF Staff Studies Behind the World Economic Outlook.*

[4] Artis, M. and Marcellino, M., (1998a), "Fiscal Solvency and Fiscal Forecasting in Europe", CEPR W.P. 1836.

[5] Artis, M. and Marcellino, M., (1998b), "Fiscal Forecasting: The Track Record of IMF, OECD and EC", mimeo, EUI.

[6] Atkinson, A.C. (1970), "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society*, Series B, 32, 211-243.

[7] Chan, Y., Stock, J.H. and Watson, M.W. (1998), "A Dynamic Factor Model Framework for Forecast Combination", mimeo, Harvard University.

[8] Chong, Y.Y. and Hendry, D.F. (1986), "Econometric evaluation of linear macro-economic models", *Review of Economic Studies*, 53, 671-690.

[9] Christoffersen, P.F. and Diebold, F.X. (1994), "Optimal prediction under asymmetric loss", N.B.E.R. Technical W.P. 167.

[10] Clemen, R.T. (1989), "Combining Forecasts: A Review and Annotated Bibliography", *Journal of Forecasting*, 5, 559-583.

[11] Clements, M.P. and Hendry, D.F. (1993), "On the Limitations of Comparing Mean Square Forecast Errors", *Journal of Forecasting*, 12, 617-637.

[12] Clements, M.P. and Hendry, D.F. (1997a), *The Marshall Lectures on Economic Forecasting*, Cambridge University Press, *forthcoming.*

[13] Clements, M.P. and Hendry, D.F. (1997b), *The Zeuthen Lectures on Forecasting*, MIT Press, *forthcoming.*

[14] Clements, M.P. and Hendry, D.F. (1997c), "Intercept corrections and structural change", *Journal of Applied Econometrics*, 11, 475-494.

[15] Cox, D.R. (1961), "Tests of Separate Families of Hypotheses", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, 105-123, Berkeley: University of California Press.

[16] Davidson, R., and MacKinnon, J.G. (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, 49, 781-793.

[17] Diebold, F.X. (1989), "Forecast Combination and Encompassing: Reconciling Two divergent Literatures", *International Journal of Forecasting*, 5, 589-592.

[18] Diebold, F.X. and Mariano, R.S. (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

[19] Doornik, J.A. and Hendry, D.F. (1997), *Modelling Dynamic Systems using PcFiml 9 for Windows*, London: International Thomson Business Press.

[20] Ericsson, N.R. (1992), "Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration", *Journal of Policy Modeling*, 14, 465-495.

[21] Ericsson, N.R. and Marquez, J. (1993), "Encompassing the Forecasts of US Trade-Balance Models", *Review of Economics and Statistics*, 75, 19-31.

[22] Fair, R.C., and Shiller, R.J. (1989), "The Informational Content of Ex-ante Forecasts", *Review of Economics and Statistics*, 71, 325-331.

[23] Fair, R.C., and Shiller, R.J. (1990), Comparing Information in Forecasts from Econometric Models", *American Economic Review*, 80, 375-389.

[24] Godfrey, L.G. (1988), *Misspecification Tests in Econometrics*, Cambridge: Cambridge University Press.

[25] Granger, C.W.J. (1997), "Outline of forecast theory using generalized cost functions", mimeo, UCSD.

[26] Granger, C.W.J. and Newbold, P. (1973), "Some Comments on the Evaluation of Economic Forecasts", *Applied Economics*, 5, 35-47.

[27] Granger, C.W.J. and Newbold, P. (1986), *Forecasting economic time series*, 2nd edn., New York: Academic Press.

[28] Hansen, B.E. (1992), "Tests for Parameter Instability in Linear Models", *Journal of Policy Modeling*, 14, 517-533.

[29] Hendry, D.F. and Doornik, J.A. (1997), *Empirical Econometric Modelling using PcGive 9 for Windows*, London: International Thomson Business Press.

[30] Hoel, P.G. (1947), "On the Choice of Forecasting Formulas", *Journal of the American Statistical Association*, 42, 605-611.

[31] Laster, D., Bennett, P., and Geoum, I.S. (1997), "Rational Bias in Macroeconomic Forecasts", Federal Reserve bank of New York, Staff Reports n. 21.

[32] Lu, M. and Mizon, G.E. (1991), "Forecast encompassing and model evaluation", in Hackl, P. and Westlund, A.H. (eds.), *Economic structural change, Analysis and Forecasting*, 123-138, Berlin: Springer-Verlag.

[33] MacKinnon, J.G. (1983), "Model Specification Tests against Non-nested Alternatives", *Econometric Reviews*, 2, 85-110.

[34] Mizon, G.E. (1984), "The Encompassing Approach in Econometrics", in Hendry, D.F. and Wallis, K.F. (eds.), *Econometrics and Quantitative Economics*, 135-172, Oxford: Basil Blackwell..

[35] Mizon, G,E. and Richard, J.-F. (1986), "The Encompassing Principle and its Application to Testing Non-nested Hypotheses", *Econometrica*, 54, 657-678.

[36] Nelson, C.R. (1972), "The Prediction Performance of the F.R.B.-MIT-PENN Model of the US Economy", *American Economic Review*, 62, 902-917.

[37] Pagan, A.R. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, 25, 221-247.

[38] Pesaran, M.H., (1974), "On the General Problem of Model Selection", *Review of Economic Studies*, 41, 153-171.

[39] Pesaran, M.H., and Deaton, A.S. (1978), "Testing Non-nested Nonlinear Regression Models", *Econometrica*, 46, 677-694.

[40] Phillips, P.C.B., and Hansen, B.E. (1990), "Statistical Inference in Instrumental Variables Regressions with I(1) Processes", *Review of Economic Studies*, 57, 99-125.

[41] Sims, C.A., Stock, J.H., and Watson M.W. (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, 113-144.

[42] Theil, H. (1958), *Economic Forecasts and Policy*, Amsterdam: North Holland.

[43] Wallis, K.F., (1989), "Macroeconomic Forecasting: A Survey", *Economic Journal*, 99, 28-61.

[44] White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", Econometrica, 48, 817-838.

[45] Zellner, (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions, and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, 348-368.