doi:10.1111/j.1467-9892.2007.00538.x

MEASURING THE ADVANTAGES OF MULTIVARIATE VS. UNIVARIATE FORECASTS

By Daniel Peña and Ismael Sánchez

Universidad Carlos III de Madrid

First Version received May 2006

Abstract. Suppose we are interested in forecasting a time series and, in addition to the time series data, we have data from many time series related to the one we want to forecast. Since building a dynamic multivariate model for the set of time series can be a complex task, it is important to measure in advance the increase in precision to be attained by using multivariate forecasts with respect to univariate ones. This article presents a simple procedure designed to obtain a consistent estimate of this measure. Its performance is illustrated with Monte Carlo simulations and examples.

Keywords. ARMA models; dynamic regression models; prediction; VARMA models.

1. INTRODUCTION

Building a vector autoregressive moving-average (VARMA) model for a large set of stationary time series is not an easy task, especially when we have common factors. For instance, it can be shown (Peña and Box, 1987) that, when the series are driven by a set of common factors plus noise and the factors follow a vector autoregressive (VAR) model, the vector of series will also follow an autoregressive moving-average (ARMA) model but the autoregressive (AR) and moving-average (MA) matrices of the VARMA representation may not be identified. Some useful procedures for building multivariate models have been proposed. We could, for instance, fit a VARMA model, an unrestricted VAR, a reduced-rank model, or even a conintegration-based error correction model (see, for instance, Tiao and Tsay, 1989; Reinsel, 1993; Johansen, 1995; Reinsel and Velu, 1998; Peña *et al.*, 2001). But before this modelling effort, it seems useful to know the advantages multivariate forecasts are expected to provide with respect to the univariate ones. In particular, it would be useful to have a measure of the expected decrease in the mean-squared forecast error of the multivariate model with respect to the univariate one. It is obvious that, if the series are very weakly related, we cannot expect a large improvement on the univariate forecast by using the joint dynamics of the series. On the other hand, when the present value of one of the time series depends strongly on the past values of the others, we expect clear advantages in the multivariate forecasts with respect to the univariate ones. This article

0143-9782/07/06 886–909 JOURNAL OF TIME SERIES ANALYSIS Vol. 28, No. 6

introduces a simple measure of the expected decrease in the mean-squared forecast error of the multivariate model with respect to the univariate one. The proposed measure is estimated by using a procedure that does not require the specification of any multivariate structure; it can be computed by ordinary least squares (OLS) and provides us with an R^2-like measure (e.g. Pierce, 1979; Granger and Newbold, 1986) without assuming any particular parametric model. This measure is both consistent and asymptotically unbiased, and can be used to foresee the advantages of building a more elaborated multivariate model. A MATLAB program to perform these analyses can be downloaded from halweb.uc3m.es/esp/Personal/personas/ismael/eng/public.html.

This article is organized as follows. Section 2 analyses the relationship between univariate and multivariate forecast for a given time series, and proposes a simple measure of the gain in the mean-squared forecast error of the multivariate forecasts with respect to the univariate ones. Section 3 illustrates the usefulness of the proposed measure with some Monte Carlo simulations and Section 4 with two real-data examples.

2. A POPULATION PREDICTABILITY MEASURE FOR MULTIVARIATE VS. UNIVARIATE FORECASTS

Let $\{y_{1t}\}$, $\{y_{2t}\}, \ldots, \{y_{mt}\}$ be a set of stationary processes with zero mean and $E(y_{it}^2) < \infty$ that jointly form a m-variate stationary process $Y_t = (y_{1t}, \ldots, y_{mt})'$ with bounded spectral density matrix. We are interested in comparing, under a quadratic loss function, the performance of the h-step-ahead prediction of y_{1t+h} obtained from a multivariate model with respect to the one from a univariate model. We will start by introducing the notation for both the univariate and multivariate predictors without assuming finite parameter models. We then build a measure of the relative performance of both predictors that is consistently estimated using OLS.

2.1. *The univariate predictor*

We assume that y_{1t} have the following AR(∞) representation:

$$y_{1t} = \sum_{i=1}^{\infty} \phi_i y_{1t-i} + \eta_t, \tag{1}$$

where η_t is a white-noise sequence with $E(\eta_t^2) = \sigma^2$, and ϕ_i, $i = 1, 2, \ldots,$ are absolutely summable real coefficients such that the polynomial

$$A(z) = 1 - \sum_{i=1}^{\infty} \phi_i z^i \neq 0, \quad |z| \leq 1.$$

The optimal one-step-ahead prediction with the mean-squared prediction error (MSPE) criterion is

$$y^{\mathrm{U}}_{1t+1|t} = \sum_{i=1}^{\infty} \phi_i y_{1t+1-i}.$$

We are interested in h-step-ahead forecasts. Considering that y_{1t+h} also admits a lead-h AR representation as a function of y_{1t}, $y_{1t-1}, \ldots$, given by

$$y_{1t+h} = \sum_{i=1}^{\infty} \alpha_i y_{1t+1-i} + u_{t+h},$$

where $\alpha_i = \alpha_i(h)$, the optimal (with the MSPE criterion) h-step-ahead forecast is

$$y^{\mathrm{U}}_{1t+h|t} = \sum_{i=1}^{\infty} \alpha_i y_{1t+1-i}, \tag{2}$$

and $u_{t+h} = u_{t+h}(h)$ is the lead-h prediction error of this univariate predictor, with MSPE given by

$$\sigma^2_{\mathrm{U}}(h) = E(u^2_{t+h}). \tag{3}$$

The predictability of the series for different horizons using only its own past can be measured by the function

$$P_{\mathrm{U}}(h) = 1 - \gamma_0^{-1}\sigma^2_{\mathrm{U}}(h),$$

where $\gamma_i = E(y_{1t}, y_{1t+i})$. This predictability $P_{\mathrm{U}}(h)$ measures the decrease in mean-squared forecast error of predicting y_{1t+h} by using the univariate model with respect to using the unconditional mean, and it has been used by many authors (Yaglom, 1963; Box and Tiao, 1977; Granger and Newbold, 1986).

The finite partial sum $\sum_{i=1}^{k_1} \alpha_i y_{1t-i+1}$ provides a suboptimal predictor of y_{1t+h} from origin t based on the last k_1 observations. The optimal univariate predictor based on k_1 lags is denoted by

$$y^{\mathrm{U}(k_1)}_{1t+h|t} = \sum_{i=1}^{k_1} \alpha^*_i y_{1t-i+1}, \tag{4}$$

with $\alpha^*_i = \alpha^*_i(k_1, h)$. If $u^{(k_1)}_{t+h} = y_{1t+h} - y^{\mathrm{U}(k_1)}_{1t+h|t}$ is the prediction error of this predictor, then its MSPE is

$$\sigma^2_{\mathrm{U}}(k_1, h) = E(u^{(k_1)2}_{t+h}). \tag{5}$$

2.2. *The multivariate predictor*

Let the m-variate stationary process $Y_t = (y_{1t}, \ldots, y_{mt})'$ have the following VAR(∞) representation

$$Y_t = \sum_{i=1}^{\infty} \Pi_i Y_{t-i} + \mathbf{a}_t \tag{6}$$

with $\mathbf{a}_t = (a_{1t}, \ldots, a_{mt})'$ being an m-dimensional white noise or innovation process, that is, $E(\mathbf{a}_t) = 0, E(\mathbf{a}_t\mathbf{a}_t') = \Omega_a$ and $E(\mathbf{a}_t\mathbf{a}_s') = 0$ for $s \neq t$. The covariance matrix Ω_a is assumed to be non-singular. Moreover, Π_i, $i = 1, 2, \ldots$ are absolutely summable coefficient matrices satisfying

$$\Pi(z) = I_m - \sum_{i=1}^{\infty} \Pi_i z^i, \quad |z| \leq 1,$$

where I_m is the identity matrix of dimension m. We also assume that $\mathbf{a}_t$ have finite moments of order $s \geq 8$. This assumption on the order of the moments is needed for the proofs of the following theorems. For convenience, the representation in eqn (6) can also be written as $Y_t = J_m A Z_{t-1} + \mathbf{a}_t$, with $J_m = [I_m 0 \ldots]$, $Z_t = (Y_t, Y_{t-1}, \ldots)'$, and A is a companion matrix with the coefficient matrices Π_i, i =1, 2, …, in the first row, identity matrices I_m in the first subdiagonal, and matrices of zeroes elsewhere, as follows:

$$A = \begin{bmatrix} \Pi_1 & \Pi_2 & \cdots \\ I_m & 0 & \cdots \\ 0 & I_m & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix}. \tag{7}$$

Then

$$y_{1t+h} = c_1' J_m A^h Z_t + v_{t+h}, \tag{8}$$

where $c_1 = (1, 0, \ldots, 0)'$ is a vector of dimension $m \times 1$ and $v_{t+h} = v_{t+h}(h)$ is the lead-h prediction error, which is given by

$$v_{t+h} = \sum_{l=0}^{h-1} c_1' \Upsilon_l \mathbf{a}_{t+h-l}, \tag{9}$$

with $\Upsilon_l = J_m A^l$ and $\Upsilon_0 = I_m$. From eqn (8), y_{1t+h} can be written as

$$\begin{aligned} y_{1t+h} &= \beta_1' Y_{1t} + \beta_2' Y_{2t} + \cdots + \beta_m' Y_{mt} + v_{t+h} \\ &= \sum_{i=1}^{\infty} \beta_{1i} y_{1t-i+1} + \sum_{j=2}^{m} \sum_{b=1}^{\infty} \beta_{jb} y_{jt-b+1} + v_{t+h} \end{aligned} \tag{10}$$

with $\beta_{1i} = \beta_{1i}(h)$, $\beta_{jb} = \beta_{jb}(h)$, $\beta_j = (\beta_{j1}, \beta_{j2}, \ldots)'$ and $Y_{jt} = (y_{jt}, y_{jt-1}, \ldots)'$. For instance, in the VAR(1) case

$$Y_{t+1} = \begin{bmatrix} y_{1t+1} \\ y_{2t+1} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} + \begin{bmatrix} a_{1t+1} \\ a_{2t+1} \end{bmatrix},$$

and for $h = 1$ we have $\beta_{11} = \phi_{11}$, $\beta_{21} = \phi_{12}$. For $h = 2$ we then have $\beta_{11} = \phi_{11}^2 + \phi_{21}\phi_{12}$, $\beta_{21} = \phi_{12}(\phi_{11} + \phi_{22})$. Note that the term ϕ_{21}, which produces a feedback relationship between the variables, is present in eqn (10) for $h > 1$. Then, as h varies, the parameter vector β_j makes use of the appropriate elements of matrix A in eqn (7). This is an interesting feature, since it reveals that a single-equation model as in eqn (10), built for horizon h, contains all the necessary information without the need to manage a multivariate predictor. The optimal multivariate predictor is

$$y^{\mathrm{M}}_{1t+h|t} = c_1' J_m A^h Z_t = \sum_{j=1}^{m} \beta_j' Y_{jt}$$
$$= \sum_{i=1}^{\infty} \beta_{1i} y_{1t-i+1} + \sum_{j=2}^{m} \sum_{b=1}^{\infty} \beta_{jb} y_{jt-b+1},$$

and the MSPE of this predictor is $\sigma^2_{\mathrm{M}}(h) = E(v^2_{t+h})$. Then, the predictability of the series using the multivariate information is

$$P_{\mathrm{M}}(h) = 1 - \gamma_0^{-1} \sigma^2_{\mathrm{M}}(h).$$

This predictability $P_{\mathrm{M}}(h)$ measures the decrease in mean-squared forecast error of predicting y_{1t+h} using the multivariate model with respect to using the unconditional mean. The multivariate predictor based on finite partial sums is

$$y^{\mathrm{M}(L)}_{1t+h|t} = \sum_{i=1}^{k_1} \beta^*_{1i} y_{1t-i+1} + \sum_{j=2}^{m} \sum_{b=1}^{k_j} \beta^*_{jb} y_{jt-b+1}, \tag{11}$$

where $L \equiv L(h) = (k_1, k_2, \ldots, k_m)$ denotes the vector of orders of the partial sum. Note that these orders might depend on h. Then, the corresponding multivariate prediction error is

$$v^{(L)}_{t+h} = y_{1t+h} - y^{\mathrm{M}(L)}_{1t+h|t}.$$

The MSPE of this predictor will be denoted by $\sigma^2_{\mathrm{M}}(L, h) = E(v^{(L)2}_{t+h})$.

2.3. *Relation between the univariate and the multivariate predictors*

The univariate forecast $y^{\mathrm{U}}_{1t+h|t}$ can be expressed as $y^{\mathrm{U}}_{1t+h|t} = E(y_{1t+h}|Y_{1t})$. Then, taking expectations conditional to Y_{1t} in eqn (10), we can obtain an alternative expression for the univariate forecast $y^{\mathrm{U}}_{1t+h|t}$ as

$$y^{\mathrm{U}}_{1t+h|t} = \beta_1' Y_{1t} + \beta_2' E(Y_{2t}|Y_{1t}) + \cdots + \beta_m' E(Y_{mt}|Y_{1t}), \tag{12}$$

which implies that the univariate forecasts of a given variable are the result of substituting in the general form in eqn (10) the future values of the remaining variables by their conditional expectations, given the past values of the forecasted variable. Define

$$\varepsilon_{jt} = Y_{jt} - E(Y_{jt}|Y_{1t}), \quad j = 2, \ldots, m, \tag{13}$$

as the vector of forecast errors of the other variables when using the past values of the first variable. The components of ε_{jt}, say $\varepsilon_{jt-b+1} = y_{jt-b+1} - E(y_{jt-b+1}|y_{1t}, y_{1t-1}, \ldots)$, $b = 1, 2, \ldots$, are obtained from the regression

$$y_{jt-b+1} = \sum_{i=1}^{\infty} \theta_{ji}^{(b)} y_{1t-i+1} + \varepsilon_{jt-b+1}.$$

Then, we can write

$$y_{1t+h|t}^{\mathrm{U}} = y_{1t+h|t}^{\mathrm{M}} - \sum_{j=2}^{m} \sum_{b=1}^{\infty} \beta_{jb} \varepsilon_{jt-b+1}, \tag{14}$$

which relates the univariate and the multivariate predictor. From eqns (10), (14) and (13), we also obtain

$$u_{1t+h} = v_{1t+h} + \beta_h' \epsilon_t, \tag{15}$$

where $\beta_h' = (\beta_2', \ldots, \beta_m')$ and $\epsilon_t = (\varepsilon_{2t}', \ldots, \varepsilon_{mt}')'$. Then,

$$\sigma_{\mathrm{U}}^2(h) = \sigma_{\mathrm{M}}^2(h) + \beta_h' \mathbf{V} \beta_h, \tag{16}$$

where $\mathbf{V} = E(\epsilon_t \epsilon_t')$. Note that, when the parameters are known, the covariance matrix between the forecast errors ε_{jt} and the multivariate forecast error a_{1t+h} is zero. Equation (16) shows that, when the model is known, multivariate forecasts cannot be less precise than univariate forecasts. Note, however, that when the parameters are estimated, this may not be the case. The theoretical increase in precision with the multivariate predictor depends on the coefficients β_j, $j = 2, \ldots, m$, and on the independent variability of the other variables that is not explained by the history of the first component, which is contained in matrix $\mathbf{V}$. This equation can be interpreted as a dynamic ANOVA decomposition of the variability. Note that, if we have a zero-mean stationary process, β_h will tend to zero for large h and $\sigma_{\mathrm{U}}^2(\infty) = \sigma_{\mathrm{M}}^2(\infty)$. The predictability of the series using the multivariate predictor, with respect to the univariate one, will be defined as

$$P_{\mathrm{M|U}}(h) = 1 - \frac{\sigma_{\mathrm{M}}^2(h)}{\sigma_{\mathrm{U}}^2(h)}. \tag{17}$$

This predictability $P_{\mathrm{M|U}}(h)$ measures the decrease in mean-squared forecast error of predicting y_{1t+h} when using the multivariate model with respect to using the univariate model. In Section 2.4 we study the estimation of this measure without the need for building such a multivariate model.

2.4. *Estimation of the prediction variances*

Given a set of observations, $t = 1, \ldots, T$, we can estimate the prediction errors of both the univariate and the multivariate predictors and use them to estimate the predictability $P_{\mathrm{M|U}}(h)$. The univariate residual variance can be estimated in a direct way by fitting by OLS the univariate AR(k) model for horizon h, as in eqn (4), and by using the univariate residuals $\hat{u}_{1t+h}$ to provide an estimate of $\sigma^2_{\mathrm{U}}(h)$. Following Bhansali's (1997, 1999) notation, these predictors estimated for a given lead time h will be denoted as direct predictors. Conversely, the more usual approach consisting of estimating a model for $h = 1$ and then obtaining the lead-h predictions from the same model, replacing the unknown future values by their own forecast, will be denoted as 'plug-in' predictors. The methodology proposed in this article is based on the use of direct predictors. In order to estimate the model, we first make Assumption 1.

ASSUMPTION 1. *The autoregressive order k_1 fitted for obtaining $y^{\mathrm{U(k_1)}}_{1t+h|t}$ is such that, when $T \to \infty$,*

(i) *k_1 is chosen as a function of T such that $k_1^3/T \to 0$ as $k_1,\ \ T \to \infty$*
(ii) *k_1 is chosen as a function of T such that $T^{1/2}(\sum_{i=k_1+1}^{\infty} |\alpha_i|) \to 0$*

Condition (i) implies that, although the univariate AR order grows with T, it is small compared with the sample size. Condition (ii) implies that the neglected coefficients in the AR(k_1) approximation will have a small effect. Both conditions (i) and (ii) are technical. They allow OLS estimates to have similar asymptotic properties as in models with a fixed number of parameters. In practice, it is not possible to check this assumption. It will, however, be useful to use a model selection criterion for the choice of k_1 with a penalty in the number of parameters related to the sample size. Let Y_1^{U} be the matrix of regressors:

$$Y_1^{\mathrm{U}} = \begin{bmatrix} y_{1k_1} & y_{1k_1-1} & \cdots & y_{11} \\ y_{1k_1+1} & y_{1k_1} & \cdots & y_{12} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1T-h} & y_{1T-h-1} & \cdots & y_{1T-h-k_1+1} \end{bmatrix},$$

and let $Y_h^{\mathrm{U}} = [y_{1k_1+h}, y_{1k_1+h+1}, \ldots, y_{1T}]'$ be the dependent variable of the regression. Then, the estimated univariate predictor is

$$\hat{y}^{U(k_1)}_{1t+h|t} = \sum_{i=1}^{k_1} \hat{\alpha}^*_i y_{1t-i+1}, \tag{18}$$

where

$$\hat{\alpha}^* \equiv \hat{\alpha}^*(k_1, h) = [\hat{\alpha}^*_1, \ldots, \hat{\alpha}^*_{k_1}]' = \left(Y_1^{\mathrm{U}\prime} Y_1^{\mathrm{U}}\right)^{-1} Y_1^{\mathrm{U}\prime} Y_h^{\mathrm{U}}.$$

The univariate h-step-ahead prediction error of our direct predictor is

$$\hat{u}^{(k_1)}_{t+h} = y_{1t+h} - \hat{y}^{\mathrm{U}(k_1)}_{1t+h|t}, \quad t = k_1, \ldots, T - h. \tag{19}$$

The estimator of $\sigma^2_U(h)$ will be

$$\hat{\sigma}^2_{\mathrm{U}}(h) = \frac{\sum_{t=k_1}^{T-h}\left(\hat{u}^{(k_1)}_{t+h}\right)^2}{T-h-2k_1+1}, \tag{20}$$

where the denominator is the number of terms of the numerator, corrected by the number of estimated parameters, k_1. Consequently, if a mean is estimated, the denominator would be $T - h - 2k_1$ instead. Then, following Bhansali (1993), it holds that

$$\hat{\sigma}^2_{\mathrm{U}}(h) \xrightarrow{p} \sigma^2_U(h), \tag{21}$$

$$E\left[\hat{\sigma}^2_{\mathrm{U}}(h)\right] = \sigma^2_{\mathrm{U}}(h) + o\left(T^{-1}\right), \tag{22}$$

where '$\xrightarrow{p}$' denotes convergence in probability. Result (22) is an interesting one, since it leads to an easy procedure to obtain an asymptotically unbiased estimator of $\sigma^2_{\mathrm{U}}(h)$. Note that $\hat{\sigma}^2_{\mathrm{U}}(h)$ is the residual variance of a linear regression with autocorrelated errors. It is well known that, when OLS is applied to a linear model with non-spherical disturbances, the residual variance is a biased estimator of the variance of perturbations, with the bias being a complex function of the true parameters. However, by eqn (22), in a direct predictor the use of a correction by degrees of freedom allows us to build an asymptotically unbiased estimator of the variance of perturbations.

The estimation of the variance of the multivariate predictor, $\sigma^2_{\mathrm{M}}(h)$, can also be carried out by fitting the multivariate dynamic regression model, as in eqn (11). We again use a direct predictor. We make a similar assumption as before:

Assumption 2. *The autoregressive orders k_j, $j = 2, \ldots, m$, fitted for obtaining $y^{M(\mathrm{L})}_{1t+h|t}$ in eqn* (11) *are such that, when $T \to \infty$,*

(i) *k_j is chosen as a function of T such that $k_j^3/T \to 0$ as $k_j, T \to \infty$*
(ii) *k_j is chosen as a function of T such that $T^{1/2}(\sum_{b=k_j+1}^{\infty} |\beta_{jb}|) \to 0$*

The interpretation and usefulness of this assumption is similar to Assumption 1. Let us define $k_{\mathrm{M}} = \max(k_1, \ldots, k_m)$. We can then define the matrix of the multivariate regressors as

$$Y_1^{\mathrm{M}} = \begin{bmatrix} y_{1k_{\mathrm{M}}} & y_{1k_{\mathrm{M}}-1} & \cdots & y_{1k_{\mathrm{M}}-k_1+1} & \cdots & y_{mk_{\mathrm{M}}} & \cdots & y_{mk_{\mathrm{M}}-k_m+1} \\ y_{1k_{\mathrm{M}}+1} & y_{1k_{\mathrm{M}}} & \cdots & y_{1k_{\mathrm{M}}-k_1} & \cdots & y_{mk_{\mathrm{M}}-1} & \cdots & y_{mk_{\mathrm{M}}-k_m+2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ y_{1T-h} & y_{1T-h-1} & \cdots & y_{1T-h-k_1+1} & \cdots & y_{mT-h} & \cdots & y_{mT-h-k_m+1} \end{bmatrix}, \tag{23}$$

which includes the present and past values of all the other time series, and let $Y_h^{\mathrm{M}} = [y_{1k_{\mathrm{M}}+h}, y_{1k_{\mathrm{M}+h+1}}, \ldots, y_{1T}]'$ be the dependent variable of the regression. Then

$$\hat{y}^{\mathrm{M}(L)}_{1t+h|t} = \sum_{i=1}^{k_1} \hat{\beta}^*_{1i} y_{1t-i+1} + \sum_{j=2}^{m} \sum_{b=1}^{k_j} \hat{\beta}^*_{jb} y_{jt-b+1}, \tag{24}$$

where

$$\hat{\beta}^* \equiv \hat{\beta}^*(L,h) = \left[\hat{\beta}^*_{11},\ldots,\hat{\beta}^*_{mk_m}\right]' = \left(Y_1^{\mathrm{M}\prime} Y_1^{\mathrm{M}}\right)^{-1} Y_1^{\mathrm{M}\prime} Y_h^{\mathrm{M}}.$$

Finally, the multivariate h-step-ahead prediction error of this estimated predictor is

$$\hat{v}^{(L)}_{t+h} = y_{1t+h} - \hat{y}^{\mathrm{M}(L)}_{1t+h|t}; \quad t = k_{\mathrm{M}},\ldots,T-h. \tag{25}$$

The estimator of $\sigma^2_{\mathrm{M}}(h)$ is given by

$$\hat{\sigma}^2_{\mathrm{M}}(h) = \frac{\sum_{t=k_{\mathrm{M}}}^{T-h} \left(\hat{v}^{(L)}_{t+h}\right)^2}{T-h-k_{\mathrm{M}}+1-S} \quad \text{where } S \equiv S(h) = \sum_{j=1}^{m} k_j. \tag{26}$$

Note that denominator of eqn (26) is the number of terms in the numerator corrected by the number of estimated parameters, S. If a constant term is included in the model, the denominator of eqn (26) would be $T - h - k_{\mathrm{M}} - S$. Theorem 1 shows that this estimator is consistent and asymptotically unbiased. This theorem extends the results of Banshali (1993) to multivariate models.

THEOREM 1. *Under Assumptions* 1 *and* 2, *and where* '$\xrightarrow{p}$' *denotes convergence in probability*,

(i) $\hat{\sigma}^2_M(h) \xrightarrow{p} \sigma^2_M(h)$
(ii) $E\left[\hat{\sigma}^2_M(h)\right] = \sigma^2_M(h) + o\left(T^{-1}\right)$

2.5. *Estimation of the predictability $P_{\mathrm{M|U}}(h)$*

The estimation of the predictability of the multivariate predictor with respect to the univariate predictor can be made by using the following estimator:

$$\hat{P}_{\mathrm{M|U}}(h) = 1 - \frac{\hat{\sigma}^2_{\mathrm{M}}(h)}{\hat{\sigma}^2_{\mathrm{U}}(h)}, \tag{27}$$

where $\hat{\sigma}^2_{\mathrm{U}}(h)$ is defined by eqn (20) and $\hat{\sigma}^2_{\mathrm{M}}(h)$ is defined by eqn (26). Theorem 2 shows some useful properties of $\hat{P}_{\mathrm{M|U}}(h)$.

THEOREM 2. *Under Assumptions* 1 *and* 2, *and where* '$\xrightarrow{p}$' *denotes convergence in probability*,

(i) $\hat{P}_{\mathrm{M|U}}(h) \xrightarrow{p} P_{\mathrm{M|U}}(h)$
(ii) $E[\hat{P}_{\mathrm{M|U}}(h)] = P_{\mathrm{M|U}}(h) + o(T^{-1})$

The proof of this theorem is a straightforward application of eqns (21) and (22) and Theorem 1. The predictability $P_{\mathrm{M|U}}(h)$ is based on the comparison of the variance of the h-step-ahead prediction errors of the univariate predictor, $\sigma^2_{\mathrm{U}}(h)$, and of the multivariate model, $\sigma^2_{\mathrm{M}}(h)$, assuming known parameters. In order to take into account the effect of parameter estimation on the predictions, the asymptotically unbiased estimators presented in eqns (20) and (26) are used, in an attempt to compensate the downward bias of estimating $\sigma^2_{\mathrm{U}}(h)$ and $\sigma^2_{\mathrm{M}}(h)$ respectively. This downward bias, however, is not the only effect of parameter estimation. It is well known that the use of estimated parameters inflates the variance of the out-of-sample prediction errors, $\sigma^2_{\mathrm{U}}(h)$ and $\sigma^2_{\mathrm{M}}(h)$ with respect to the case of known parameters (Yamamoto, 1976, 1981; Baillie, 1980; Fuller and Hasza, 1981). For instance, in the univariate AR(p) case, the asymptotic out-of-sample MSPE of the one-step-ahead prediction error can be approximated as $\sigma^2_{\mathrm{U}}(1)(1 + p/T)$. It is therefore makes sense to compare the performance of the univariate and the multivariate predictors based on this out-of-sample MSPE, instead of only comparing $\sigma^2_{\mathrm{U}}(h)$ and $\sigma^2_{\mathrm{M}}(h)$. This comparison is the basis of the final prediction error (FPE) criterion (Akaike, 1970). The FPE criterion consists of the asymptotic approximation, up to terms of magnitude $O(T^{-1})$, of the one-step-ahead MSPE. A derivation of the h-step FPE criterion can, however, be made and then a measure of predictability can be built by comparing the h-step FPE criterion of the univariate predictor, denoted as $\mathrm{FPE}_{\mathrm{U}}(h)$, and the multivariate one, denoted as $\mathrm{FPE}_{\mathrm{M}}(h)$.

For the lead-h univariate predictor, Bhansali (1999) has derived the h-step FPE criterion by generalizing the approach of Akaike (1970). This h-step FPE criterion is

$$\mathrm{FPE}_{\mathrm{U}}(h) = \hat{\sigma}^2_{\mathrm{U}}(h)\left(1 + \frac{k_1}{T}\right), \tag{28}$$

where k_1 is the number of estimated parameters. The intuition behind the result is eqn (28) is that for a direct predictor, the h-step-ahead prediction error is in fact a one-step-ahead prediction error, because the length of the step is already h. In a similar fashion (Lewis and Reinsel, 1985; Reinsel, 1993, p. 142), the FPE criterion for the direct multivariate predictor is

$$\mathrm{FPE}_{\mathrm{M}}(h) = \hat{\sigma}^2_{\mathrm{M}}(h)\left(1 + \frac{S}{T}\right). \tag{29}$$

We can now define a measure of predictability, denoted as $\hat{F}_{\mathrm{M|U}}(h)$, based on the estimation of the reduction of lead-h MSPE of the multivariate predictor with respect to the univariate one as follows:

$$\hat{F}_{\mathrm{M|U}}(h) = 1 - \frac{\mathrm{FPE}_{\mathrm{M}}(h)}{\mathrm{FPE}_{\mathrm{U}}(h)}. \tag{30}$$

The interpretation of $\hat{F}_{\mathrm{M|U}}(h)$ is complementary to $\hat{P}_{\mathrm{M|U}}(h)$. The estimated predictability $\hat{P}_{\mathrm{M|U}}(h)$ only compares the estimated variance of the respective innovation processes. Its correction by the number of estimated parameters makes $\hat{P}_{\mathrm{M|U}}(h)$ asymptotically unbiased. However, as the multivariate predictor will usually require more parameters than the univariate one; $\hat{P}_{\mathrm{M|U}}(h)$ would provide an optimistic view of the relative out-of-sample performance of the multivariate predictor. $\hat{F}_{\mathrm{M|U}}(h)$ compares such out-of-sample performance and is explicitly linked to the specific predictor used. Then, since the proposed estimator of $\hat{\sigma}^2_{\mathrm{M}}(h)$ is based on some AR approximation, S will be, in general, larger than what an experienced analyst can get. As a consequence, $\hat{F}_{\mathrm{M|U}}(h)$ can give a pessimistic view of the relative performance of a multivariate predictor. An analyst can think of $\hat{P}_{\mathrm{M|U}}(h)$ as a potential benchmark, and that an inefficient modelling strategy can reduce such a benchmark to a value as low as $\hat{F}_{\mathrm{M|U}}(h)$. Then, the joint interpretation of $\hat{P}_{\mathrm{M|U}}(h)$ and $\hat{F}_{\mathrm{M|U}}(h)$ can help the analyst to decide better about the convenience of fitting a multivariate predictor.

2.6. *Practical considerations*

Suppose that we have a set of m time series of sample size T, and we want to estimate the predictability of the first component, given the others for horizons $1, \ldots, H$. In practice, this can be done easily by selecting a lag order L, where L is some fixed lag value, to be discussed later, and fitting the following two equations:

1. The AR(L) of y_{1t} given by eqn (18) and then estimate $\hat{\sigma}^2_{\mathrm{U}}(h)$ for $h = 1, \ldots, H$, by using eqn (20).
2. The multivariate dynamic regression with dependent variable y_{1t+h} and independent variables, the vector of past values of this series $Y'_{1L} = (y_{1t}, y_{1t-1}, \ldots, y_{1t-L+1})$ and the $1 \times L(m-1)$ vector of past values of the other series

$$Y'_{RL} = (y_{2t}, y_{2t-1}, \ldots, y_{2t-L+1}, \ldots, y_{mt}, \ldots, y_{mt-L+1}) = (Y'_{2L}, \ldots, Y'_{mL}),$$

as in eqn (24). This is equivalent to using the matrix in eqn (23) with $k_i = L$, $i = 1, 2, \ldots, m$. Then we use the residual variance of this dynamic regression to estimate $\hat{\sigma}^2_{\mathrm{M}}(h)$ by eqn (26).

So as to apply this method in an automatic way, we need to choose the values of L. The value of L must be such that $E(y_{t+h}|Y_{1t}, \ldots, Y_{mt}) = E(y_{t+h}|Y_{1L}, \ldots, Y_{mL})$. We know (see, for instance, Zellner and Palm, 1974) that, if the vector of time series follows a VAR(p) model, the univariate time series models have maximum order ARMA(pm, $(m-1)p$). Thus, the order of the univariate AR model fitted must be larger than the order of the multivariate. Suppose by using a model selection criteria we obtain that the univariate model can be approximated by an AR(k_1) model. Then we can select $L = k_1$ as the lag. Moreover, we can perform a sensitivity analysis and repeat the computation for $L = k \pm g$ for $g = 1, 2, \ldots$ and check that the residual variance of the dynamic regression does not change. Alternatively, different orders $k_2,\ldots,k_m$ can be used. A simple procedure to obtain them is to fit individual regressions of Y_{1t} with each of the remaining regressors $Y_{2t},\ldots,Y_{mt}$ and choose each lag order by an information criteria such as Bayesian information criterion (BIC) or Akaike information criterion (AIC).

A practical problem can arise when the number of series is larger than the sample size. Fitting the proposed dynamic regression requires that $T > Lm$. Thus, if the sample size is not large and we have many time series and a large value of L, the dynamic regression cannot be fitted. However, we can still estimate the residual variance of this regression as follows:

2* Compute the r principal components of the variables Y_{RL}. Let $Y_{RL}^{(r)}$ be the $1 \times r$ vector of the r largest principal components of the vector of variables Y_{RL}. Then, we will regress y_{1t+h} on its past values and on the variables $Y_{Rt}^{(r)}$. This regression can be estimated if $T > L + r$, and this requires $r < n - h$. Note that this procedure is equivalent to a singular value decomposition of the rectangular matrix Y_{RL} followed by a reparameterization of the regression coefficients. The value of r must be smaller than $n - h$. Thus, we can select r so as to include a large proportion of the explained variability while keeping this restriction. A simple solution is to take $r = \min(n - h, g_{99})$, where g_{99} is the number of principal components required to include 99% of the variability of the vector of time series.

3. SOME MONTE CARLO RESULTS

There are several alternatives to compute $\hat{P}_{\mathrm{M|U}}(h)$, which basically differ in the order selection procedure for the autoregressions involved. Since our goal is to propose a simple method to quickly evaluate the convenience of building a more sophisticated VARMA model, we will base our empirical experiments on simple and well-known order selection procedures. In this experiment we have used both AIC and BIC. The performance of both procedures was similar, with BIC having a somewhat better performance. Therefore, for brevity we will only report the results based on the BIC.

We considered three different VAR(1) for generating the data. The first model, M1, is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.8 & \phi \\ 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}; \quad \mathrm{var}(\mathbf{a}_t) = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}, \tag{31}$$

with $\mathbf{a}_t = (a_{1t}, a_{2t})'$, and we generated bivariate series from this model with $\phi = 0, 0.25, 0.5$ (if $\phi \geq 0.6$ the process is non-stationary). For the second model, M2, we decreased the dependency of the first series from its own past and also increased the interdependency of the two series. The model is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 & \phi \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}; \quad \mathrm{var}(\mathbf{a}_t) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \tag{32}$$

with $\phi = 0,\ 0.75,\ 1.50$ (if $\phi \geq 1.6$ the process is non-stationary). The third stationary model, M3, includes four series with a more complex dynamic structure:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \end{bmatrix} = \begin{bmatrix} 0.7 & & \phi' & \\ 0.6 & 0.1 & 0.0 & 0.2 \\ 0.5 & 0.4 & -0.8 & -0.3 \\ 0.0 & -0.4 & 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \\ y_{4t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \\ a_{3t} \\ a_{4t} \end{bmatrix}; \tag{33}$$

$$\mathrm{var}(\mathbf{a}_t) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 4 \end{bmatrix}, \tag{34}$$

with $\phi' = [0 \quad 0 \quad 0]$ and $\phi' = [0.5 \quad -0.4 \quad 0.1]$. In the three models, two sample sizes, $T = 100$ and 200 were included, and the number of replications was 20,000. In each replication, a sample of $100 + T + 5$ data was generated by one of the models. The first 100 data were discarded in order to ensure stationary initial conditions, and in the following T data, and for $h = 1, 3, 5$, the models estimated using OLS. The last five observations were considered as future observations to calculate out-of-sample prediction errors. For the uni-variate model in eqn (4), k_1 was selected by BIC in the range [1,6]. For the estimation of the multivariate predictors as in eqn (11), we first estimated the orders k_j for each series $j = 2, \ldots, m$. These orders were estimated by each series and each horizon by regressing y_{1t+h} on $y_{jt}, \ldots, y_{jt-k_j}+1,\ j = 2, \ldots, m$ at a time. The orders $k_j \equiv k_j(h)$ were selected by BIC also in the range [1,6].

In order to assess how well the proposed measure $\hat{P}_{\mathrm{M|U}}(h)$ corresponds to the out-of-sample relative forecast performance of univariate and multivariate models, we also computed out-of-sample prediction errors using the last five observations not used in the estimation of the models. The predictability measure obtained from these out-of-sample prediction errors will be denoted as $P_{\mathrm{M|U}}(h)$, which can be interpreted as the population value that we are estimating with $\hat{P}_{\mathrm{M|U}}(h)$. To compute $P_{\mathrm{M|U}}(h)$, the estimated multivariate model was a VAR(k) with k selected by using BIC in the range [0,6], and the univariate models were ARMA(p, q) models again with order selected by using BIC. Since the vector of time series follows a VAR(1) model, univariate time-series models will have maximum-order ARMA (m, $m - 1$). Consequently, in models M1 and M2, the range of p used in the BIC selection is [0,3] and the range of q is [0,2]. For model M3, the range of p is [0,5] and the range of q is [0,4]. By averaging the 20,000 squared prediction errors of each model at each horizon we obtain the MSIE at each horizon for the multivariate and univariate models. The predictability measure $P_{\mathrm{M|U}}(h)$ is then computed using these MSIE.

TABLE I

MEAN AND VARIANCE OF $P_{M|U}(h)$ USING MODEL M1 ALONG WITH 20,000 REPLICATIONS

	$\phi = 0.0$			$\phi = 0.25$			$\phi = 0.50$		
		$\hat{P}_{M\|U}$			$\hat{P}_{M\|U}$			$\hat{P}_{M\|U}$	
h	$P_{M\|U}$	Mean	Var	$P_{M\|U}$	Mean	Var	$P_{M\|U}$	Mean	Var
$T = 50$									
1	−0.001	0.013	0.0037	0.104	0.102	0.0118	0.304	0.286	0.0173
3	0.006	0.044	0.0094	0.056	0.072	0.0128	0.100	0.118	0.0160
5	0.011	0.054	0.0127	0.035	0.053	0.0130	0.036	0.066	0.0146
$T = 100$									
1	−0.006	−0.004	0.0007	0.099	0.102	0.0046	0.313	0.299	0.0071
3	−0.001	0.016	0.0017	0.045	0.059	0.0044	0.119	0.126	0.0061
5	0.002	0.019	0.0023	0.024	0.036	0.0034	0.053	0.069	0.0050
$T = 200$									
1	−0.002	0.001	0.0002	0.104	0.105	0.0021	0.314	0.306	0.0032
3	−0.003	0.007	0.0004	0.054	0.056	0.0018	0.138	0.131	0.0026
5	0.000	0.008	0.0005	0.025	0.030	0.0013	0.069	0.072	0.0020

Tables I to III summarize the results. It can be seen that conclusions are very similar for the three models. The tables show that, overall, $\hat{P}_{M|U}(h)$ is quite accurate, tending to yield average values that are closer to $P_{M|U}$ as the sample size increases. Even at $T = 50$ the variance of $\hat{P}_{M|U}$ is reasonably low, allowing to detect those situations in which the multivariate predictor is profitable. As expected, the lowest performance is obtained at $T = 50$ and $h = 5$. The variance of $\hat{P}_{M|U}$ is also large in the third model and $T = 50$. At samples sizes $T = 100$ and 200, the variance is very low. It is interesting to note in these experiments that the relative advantage of the multivariate predictor depends on the horizon. For

TABLE II

MEAN AND VARIANCE OF $P_{M|U}(h)$ USING MODEL M2 ALONG WITH 20,000 REPLICATIONS

	$\phi = 0.0$			$\phi = 0.75$			$\phi = 1.50$		
		$\hat{P}_{M\|U}$			$\hat{P}_{M\|U}$			$\hat{P}_{M\|U}$	
h	$P_{M\|U}$	Mean	Var	$P_{M\|U}$	Mean	Var	$P_{M\|U}$	Mean	Var
$T = 50$									
1	0.005	0.017	0.0034	0.361	0.358	0.0141	0.714	0.729	0.0047
3	−0.003	0.022	0.0042	0.074	0.056	0.0068	0.114	0.118	0.0075
5	0.002	0.022	0.0046	0.038	0.020	0.0066	0.037	0.049	0.0079
$T = 100$									
1	0.013	0.005	0.0006	0.366	0.361	0.0066	0.732	0.736	0.0019
3	−0.001	0.007	0.0008	0.062	0.063	0.0022	0.135	0.143	0.0025
5	−0.001	0.007	0.0008	0.036	0.022	0.0012	0.060	0.072	0.0018
$T = 200$									
1	0.013	0.002	0.0001	0.363	0.363	0.0030	0.735	0.739	0.0009
3	−0.002	0.003	0.0001	0.073	0.067	0.0010	0.160	0.154	0.0011
5	0.002	0.003	0.0001	0.028	0.024	0.0005	0.083	0.084	0.0007

TABLE III

MEAN AND VARIANCE OF $P_{M|U}(h)$ USING MODEL M3 ALONG WITH 20,000 REPLICATIONS

	$\phi' = [0 \quad 0 \quad 0]$			$\phi' = [0.5 \quad -0.4 \quad 0.1]$		
		$\hat{P}_{M\vert U}$			$\hat{P}_{M\vert U}$	
h	$P_{M\vert U}$	Mean	Var	$P_{M\vert U}$	Mean	Var
$T = 50$						
1	−0.048	0.025	0.0106	0.580	0.533	0.0146
3	−0.026	0.063	0.0187	0.221	0.194	0.0261
5	−0.008	0.075	0.0241	0.115	0.111	0.0336
$T = 100$						
1	−0.030	0.008	0.0019	0.560	0.542	0.0052
3	−0.013	0.020	0.0032	0.198	0.197	0.0073
5	−0.005	0.025	0.0040	0.102	0.100	0.0075
$T = 200$						
1	−0.016	0.002	0.0004	0.554	0.546	0.0024
3	−0.007	0.008	0.0007	0.206	0.196	0.0030
5	−0.003	0.010	0.0008	0.098	0.100	0.0028

instance, in the third experiment (Table III) with $\phi' = [0.5 \quad -0.4 \quad 0.1]$, the multivariate predictor is very competitive at $h = 1$. This advantage is clearly detected by $\hat{P}_{M|U}$. However, if our interest is to make predictions at $h = 5$, the multivariate predictor will not help out, as $\hat{P}_{M|U}$ also reveals. As a result, $\hat{P}_{M|U}(h)$ can provide a simple and accurate measure of what can be expected from a more elaborate multivariate model.

4. SOME EXAMPLES

In this section we illustrate the use of the proposed measure $\hat{P}_{M|U}(h)$ with two examples. In both examples we compute the predictability by running the two regressions in eqns (18) and (24) with a maximum lag equal to 6 and selecting the order by the BIC criterion. We also made the order selection by using the AIC criterion, but as the results were very similar, only the results obtained with BIC are reported here. Note that in both examples the time series are non-stationary but have been transformed to stationarity.

4.1. *Example 1: gas furnace data*

Box and Jenkins (1976, p. 381) built a transfer function model for the proportion of output CO_2 (y_t) as a function of the non-stochastic feed rate of methane (x_t) in a gas furnace. The data correspond to 296 readings at 9-second interval. The predictability measure considered in the previous section applied to these data are shown in Table IV. The analyses have been made with the first

TABLE IV

UNI-VARIATE AND MULTI-VARIATE PREDICTABILITY OF THE PROPORTION OF OUTPUT WITH THE GASFURNACE DATA

h	$\hat{P}_{\mathrm{U}}(h)$	$\hat{P}_{\mathrm{M}}(h)$	$\hat{P}_{\mathrm{M\|U}}(h)$	$\hat{F}_{\mathrm{M\|U}}(h)$
1	0.791	0.889	0.468	0.456
3	0.234	0.834	0.783	0.779
5	0.100	0.736	0.707	0.705

differences of both y_t and x_t. The column $\hat{P}_{\mathrm{U}}(h)$ shows that the univariate predictability of y_t decreases fast with the horizon. However, the column $\hat{P}_{\mathrm{M}}(h)$ indicates that this loss of predictability for higher horizons does not appear in the multivariate predictor. The column $\hat{P}_{\mathrm{M|U}}(h)$ indicates that the transfer function model is expected to lead to a reduction of MSPE, with respect to the univariate model, of 47.7% for $h = 1$ and as large as 78.3% for $h = 3$. In this example as the number of parameters is not large we only have an explanatory variable. For this reason, the values of $\hat{F}_{\mathrm{M|U}}(h)$ are only slightly smaller than those of $\hat{P}_{\mathrm{M|U}}(h)$. As a result, we conclude that the feed rate of methane is an excellent control variable for the output CO_2, even in the long term.

4.2. *Example 2: gross domestic product in europe*

In this second example we are interested in forecasting the quarterly rate of growth of the seasonally adjusted gross domestic product in Spain. In addition to this time series we also have the times series of this variable for eight other European countries: Belgium, Denmark, France, Italy, the Netherlands, Finland, UK and Norway. The series of $\nabla \log x_t$, where x_t is the seasonally adjusted product at market prices from Eurostat, are shown in Figure 1 in the period January 1980 to March 2002, with a total of 90 observations. The first series corresponds to Spain and then we show the other European countries in the same order presented before so that the second one is Belgium and the last one Norway. Table V shows the results of forecasting the series of Spain. According to $\hat{P}_{\mathrm{M|U}}(h)$, there is not much advantage in using the information on the other European countries at $h = 1$, since the expected reduction in MSPE is about 4%. This reduction is very small and, as suggested by $\hat{F}_{\mathrm{M|U}}(1)$, can vanish because of the estimation variability. The conclusion at $h = 1$ is that a multivariate predictor might have some advantage with a larger data set and careful modelling. With smaller data sets, an univariate predictor is preferred. We can be more optimistic at $h = 3$. According to $\hat{P}_{\mathrm{M|U}}(3)$, the estimated reduction in MSPE is about 20%. We also need careful modelling otherwise, as indicated by $\hat{F}_{\mathrm{M|U}}(3)$, we can lose an important portion of that advantage. As the horizon grows, the relative advantage of the multivariate predictor is reduced to an expected gain of 14.5%. Again, as suggested by $\hat{F}_{\mathrm{M|U}}(5)$, the small sample size can make that an excess of sampling variability significantly reduces

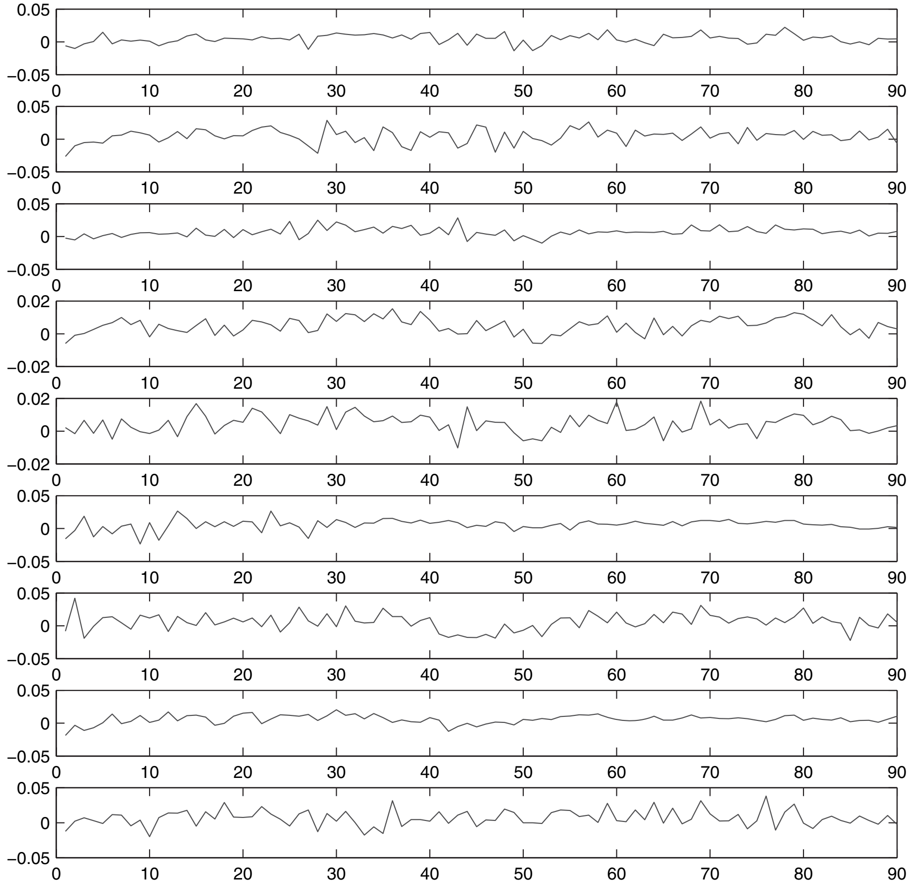

FIGURE 1. Rate of growth, $\nabla \log x_t$, of seasonally adjusted domestic product in the nine European countries. The first time series correspond to Spain and the last one to Norway.

TABLE V
UNIVARIATE AND MULTIVARIATE PREDICTABILITY OF RATE OF GROWTH OF GROSS DOMESTIC PRODUCT IN SPAIN

h	$\hat{P}_{\mathrm{U}}(h)$	$\hat{P}_{\mathrm{M}}(h)$	$\hat{P}_{\mathrm{M}\mid\mathrm{U}}(h)$	$\hat{F}_{\mathrm{M}\mid\mathrm{U}}(h)$
1	0.240	0.273	0.044	≤ 0
3	0.211	0.368	0.199	0.111
5	0.169	0.289	0.145	0.017

such advantage. Then, because of the small sample size, a multivariate predictor can be advantageous only to forecast in the medium term. Rate of growth, $\nabla \log x_t$, of seasonally adjusted domestic product in the nine

TABLE VI

VALUE OF $P_{\mathrm{M|U}}(h)$ FOR THE NINE COUNTRIES

h	Spain	Belgium	Denmark	France	Italy	The Netherlands	Finland	UK	Norway
1	0.044	0.020	0.159	0.309	0.308	0.276	0.183	0.223	0.088
3	0.199	0.000	0.105	−0.034	0.034	0.319	0.064	0.460	−0.016
5	0.145	0.159	0.040	0.258	0.208	−0.086	0.160	0.299	0.236

European countries. The first time series corresponds to Spain and the last one to Norway.

It is interesting to note that the relative advantage of the multivariate predictor could be very different in each country. For instance, the countries with the largest value of $\hat{P}_{\mathrm{M|U}}(1)$ are Italy (31%) and France (30%), whereas the smallest gain in the multivariate model with regard to the univariate one corresponds to Belgium (2%). Overall, the country that would have greater benefit from a multivariate model at all horizons is the UK. Table VI summarizes the value of $\hat{P}_{\mathrm{M|U}}(h)$ for all the countries.

APPENDIX

PROOF OF THEOREM 1 (i). From eqn (25) we have

$$\hat{v}_{t+h}^{(L)} = y_{1t+h} - \hat{y}_{1t+h|t}^{\mathrm{M}(L)} = y_{1t+h} - y_{1t+h|t}^{\mathrm{M}(L)} + \left(y_{1t+h|t}^{\mathrm{M}(L)} - \hat{y}_{1t+h|t}^{\mathrm{M}(L)}\right). \tag{35}$$

From eqns (10) and (11) we obtain

$$y_{1t+h} - y_{1t+h|t}^{\mathrm{M}(L)} = \sum_{j=1}^{m}\sum_{b=1}^{\infty} \beta_{jb} y_{jt-b+1} + v_{t+h} - \sum_{j=1}^{m}\sum_{b=1}^{k_j} \beta_{jb}^{*} y_{jt-b+1} \tag{36}$$

$$= v_{t+h} + \sum_{j=1}^{m}\sum_{b=1}^{k_j} \left(\beta_{jb} - \beta_{jb}^{*}\right) y_{jt-b+1} \tag{37}$$

$$+ \sum_{j=1}^{m}\sum_{b=k_j+1}^{\infty} \beta_{jb} y_{jt-b+1}. \tag{38}$$

If process y_{jt} is stationary, $y_{jt} = O_p(1)$. Then by Assumption 2(ii), we have

$$\sum_{b=k_{j+1}}^{\infty} \beta_{jb} y_{jt-b+1} = o_p(T^{-1/2}). \tag{39}$$

From Baxter's inequality (Baxter, 1963; Cheng and Pourahmadi, 1993) we have

$$\sum_{b=1}^{k_j}\left|\beta_{jb}-\beta^*_{jb}\right| \le c\sum_{i=k_j+1}^{\infty}\left|\beta_{jb}\right|,$$

with c a bounded constant, and in the sequel not always the same constant. Then, applying stationarity and Assumption 2(ii), we have

$$\sum_{b=1}^{k_j}\left(\beta_{jb}-\beta^*_{jb}\right)y_{jt-b+1} = o_p\left(T^{-1/2}\right).$$

On the other hand, we have

$$y^{\mathrm{M}(L)}_{1t+h|t} - \hat{y}^{\mathrm{M}(L)}_{1t+h|t} = \sum_{j=1}^{m}\sum_{b=1}^{k_j}\left(\beta^*_{jb}-\hat{\beta}^*_{jb}\right)y_{jt-b+1},$$

and by Assumption 2(i) and using the stationarity of the processes, we have

$$(\beta^*_{jb}-\hat{\beta}^*_{jb}) = O_p(T^{-1/2})$$

and hence

$$y^{\mathrm{M}(L)}_{1t+h|t} - \hat{y}^{\mathrm{M}(L)}_{1t+h|t} = O_p(k_{\mathrm{M}}T^{-1/2}),$$

with $k_{\mathrm{M}} = \max(k_j)$. Then,

$$\hat{v}^{(L)}_{t+h} = v_{t+h} + O_p(T^{-1/2}). \qquad \square$$

Proof of Theorem 1 (ii). Let us denote by

$$\hat{v}^{(L)} = Y_h - Y_1^{\mathrm{M}}\hat{\beta}^*$$

to the residuals of model (24) estimated by OLS, and $v^{(L)} = Y_h - Y_1^{\mathrm{M}}\beta^*$. Then,

$$E\left[\hat{\sigma}^2_{\mathrm{M}}(h)\right] = E(\hat{v}^{(L)\prime}\hat{v}^{(L)})T_{\mathrm{M}}^{-1} = E(v^{(L)\prime}M^*v^{(L)})T_{\mathrm{M}}^{-1},$$

where

$$M^* = I - Y_1^{\mathrm{M}}(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}})^{-1}Y_1^{\mathrm{M}\prime} \quad \text{and} \quad T_{\mathrm{M}} = T-h-k_{\mathrm{M}}+1-S.$$

Then,

$$E\left[\hat{\sigma}^2_{\mathrm{M}}(h)\right] = T_{\mathrm{M}}^{-1}E\left(v^{(L)\prime}v^{(L)}\right) - T_{\mathrm{M}}^{-1}E\left[v^{(L)\prime}Y_1^{\mathrm{M}}(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}})^{-1}Y_1^{\mathrm{M}\prime}v^{(L)}\right]. \tag{40}$$

For an arbitrary vector x and an $r \times r$ matrix A, let $||x|| = (x'x)^{1/2}$ be the Euclidean norm of x, and $||A|| = \sup_{||x||\le 1}(x'A'Ax)^{1/2}$ be the matrix norm of A. From eqns (35) and (36) we can write $v^{(L)}_{t+h} = v_{t+h} + w_t$, where

$$w_t = \sum_{j=1}^{m}\sum_{b=1}^{k_j}\left(\beta_{jb}-\beta^*_{jb}\right)y_{jt-b+1} + \sum_{j=1}^{m}\sum_{b=k_j+1}^{\infty}\beta_{jb}y_{jt-b+1}. \tag{41}$$

Let us denote $v = (v_{k_{\mathrm{M}}} + h,\ v_{k_{\mathrm{M}}} + h + 1, \ldots, v_T)'$ and $w = (w_{k_{\mathrm{M}}}, \ldots, w_{T-h})'$. Then $v^{(L)}v^{(L)'} = vv' + 2v\mathrm{w}' + ww'$. Therefore, since v is independent of w, $E(v^{(L)'}v^{(L)}) = E(v'v) + E(w'w)$, where

$$E(v'v) = E\left(\sum_{t=T-h}^{k_{\mathrm{M}k}} v_{t+h}^2\right) = (T - h - k_{\mathrm{M}} + 1)\sigma_{\mathrm{M}}^2(h).$$

From eqn (41) we obtain

$$\begin{aligned} E(w_t^2) = {} & \sum_{j=1}^{m}\sum_{i=1}^{m}\sum_{b=1}^{k_j}\left(\beta_{jb} - \beta_{jb}^*\right)\sum_{d=1}^{k_j}\left(\beta_{id} - \beta_{id}^*\right)E\left(y_{jt-b+1}y_{it-d+1}\right) \\ & + \sum_{j=1}^{m}\sum_{b=k_j+1}^{\infty}\beta_{jb}\sum_{d=k_j+1}^{\infty}\beta_{id}E\left(y_{jt-b+1}y_{it-d+1}\right) \\ & + 2\sum_{j=1}^{m}\sum_{i=1}^{m}\sum_{b=1}^{k_j}\left(\beta_{jb} - \beta_{jb}^*\right)\sum_{d=k_j+1}^{\infty}\beta_{id}E\left(y_{jt-b+1}y_{it-d+1}\right). \end{aligned}$$

By the stationarity of the process and applying Baxter's inequality and Assumption 2, we have

$$\begin{aligned} E(w_t^2) &= \sum_{j=1}^{m}\sum_{i=1}^{m}\sum_{b=1}^{k_j}\left(\beta_{jb} - \beta_{jb}^*\right)o\left(T^{-1/2}\right) + \sum_{j=1}^{m}\sum_{b=k_j+1}^{\infty}\beta_{jb}o\left(T^{-1/2}\right) \\ &\quad + 2\sum_{j=1}^{m}\sum_{i=1}^{m}\sum_{b=1}^{k_j}\left(\beta_{jb} - \beta_{jb}^*\right)o\left(T^{-1/2}\right) \\ &= \sum_{j=1}^{m}\sum_{i=1}^{m}o\left(T^{-1}\right) + \sum_{j=1}^{m}o\left(T^{-1}\right) + 2\sum_{j=1}^{m}\sum_{i=1}^{m}o\left(T^{-1}\right) = o(T^{-1}), \end{aligned}$$

and then $E(w'w) = o(1)$. As a result, we obtain

$$T_{\mathrm{M}}^{-1}E\left(v^{(L)\prime}v^{(L)}\right) = T_{\mathrm{M}}^{-1}(T_{\mathrm{M}} + S)\sigma_{\mathrm{M}}^2(h) + o(T_{\mathrm{M}}^{-1}). \tag{42}$$

We can write

$$E\left[v^{(L)\prime}Y_1^{\mathrm{M}}(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}})^{-1}Y_1^{\mathrm{M}\prime}v^{(L)}\right] = \mathrm{trace}\left\{E\left[\left(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}}\right)^{-1}Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\}.$$

Let us denote $\Gamma_y = E\left(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}}\right)$ and $\hat{\Gamma}_y = T_{\mathrm{M}}^{-1}\left(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}}\right)$ Then,

$$\mathrm{trace}\left\{E\left[\left(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}}\right)^{-1}Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\} \tag{43}$$

$$= \mathrm{trace}\left\{T_{\mathrm{M}}^{-1}\Gamma_y^{-1}E\left[Y_1^{M\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\} \tag{44}$$

$$+ \mathrm{trace}\left\{T_{\mathrm{M}}^{-1}E\left[\left(\hat{\Gamma}_y^{-1} - \Gamma_y^{-1}\right)Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\}. \tag{45}$$

In order to solve eqn (43) we will first analyse the term

$$\mathrm{trace}\left\{T_{\mathrm{M}}^{-1}\Gamma_y^{-1}E\left[Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\}.$$

Using similar arguments as in the preceding text, we have

$$E\left(Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right) = E\left(Y_1^{\mathrm{M}\prime}vv'Y_1^{\mathrm{M}}\right) + 2E\left(Y_1^{\mathrm{M}\prime}vw'Y_1^{\mathrm{M}}\right) + E\left(Y_1^{\mathrm{M}\prime}ww'Y_1^{\mathrm{M}}\right). \tag{46}$$

Applying Hölders' inequality, we obtain

$$E\left(\left\|Y_1^{\mathrm{M}\prime}vw'Y_1^{\mathrm{M}}\right\|\right) \leq E\left(\left\|Y_1^{\mathrm{M}}Y_1^{\mathrm{M}\prime}\right\|^2\right)^{1/2}E\left(\left\|vw'\right\|^2\right)^{1/2},$$

and since v and w are independent, $E(\|Y_1^{\mathrm{M}\prime}vw'Y_1^{\mathrm{M}}\|) = 0$. Analogously, by Assumption 2 and the stationarity of the series,

$$\begin{aligned} E\left(\left\|Y_1^{\mathrm{M}\prime}ww'Y_1^{\mathrm{M}}\right\|\right) &\leq E\left(\left\|Y_1^{\mathrm{M}}Y_1^{\mathrm{M}\prime}\right\|^2\right)^{1/2}E\left(\left\|ww'\right\|^2\right)^{1/2} \\ &= O(S^{1/2})o(1) = o\left(T_{\mathrm{M}}^{1/3}\right). \end{aligned}$$

Then

$$\begin{aligned} \mathrm{trace}\left\{T_{\mathrm{M}}^{-1}\Gamma_y^{-1}E\left[Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\} &= \mathrm{trace}\left\{T_{\mathrm{M}}^{-1}\Gamma_y^{-1}E\left(Y_1^{\mathrm{M}\prime}vv'Y_1^{\mathrm{M}}\right)\right\} \\ &\quad + o\left(T_{\mathrm{M}}^{-1/3}\right). \end{aligned}$$

Using eqn (9), we have

$$v_{t+h} = \sum_{l=0}^{h-1}\eta_{t+h-l}^{(l)},$$

where $\eta_{t+h-l}^{(l)} = c_1'\Upsilon_l\mathbf{a}_{t+h-l}$. If we denote $g_{i,j}^{(l)}$ to the (i,j) element of the matrix Υ_l we have

$$\eta_{t+h-l}^{(l)} = \sum_{j=1}^{m}g_{1j}^{(l)}a_{jt+h-l},$$

and by the properties of $\mathbf{a}_t$, we have

$$\begin{aligned} E\left[\eta_{t+h-l}^{(l)2}\right] &= c_1'\Upsilon_l\Omega_a\Upsilon_l'c_1 \equiv \chi\,\kappa_l^2; \quad l = 0, 1, \ldots, h-1; \\ E\left[\eta_{t+j}^{(l)}\eta_{t+j}^{(m)}\right] &= c_1'\Upsilon_l\Omega_a\Upsilon_m'c_1 \equiv \chi\,\kappa_{l,m}; \quad j = 1, \ldots, h \\ E\left[\eta_{t+h-m}^{(m)}\eta_{t+h-l}^{(l)}\right] &= 0; \quad l, m = 0, 1, \ldots, h-1; \\ E\left(v_{t+h}^2\right) &= \sum_{l=0}^{h-1}\chi\,\kappa_l^2 \equiv \sigma_{\mathrm{M}}^2(h); \quad l = 0, 1, \ldots, h-1. \end{aligned}$$

Then

$$Y_1^{\mathrm{M}\prime} v = \begin{bmatrix} \sum_{i=0}^{T_{\mathrm{M}}-1} \sum_{l=0}^{h-1} y_{1k_{\mathrm{M}}+i} \eta_{k_{\mathrm{M}}+h+i-l} \\ \sum_{i=0}^{T_{\mathrm{M}}-1} \sum_{l=0}^{h-1} y_{1k_{\mathrm{M}}-1+i} \eta_{k_{\mathrm{M}}+h+i-l} \\ \vdots \\ \sum_{i=0}^{T_{\mathrm{M}}-1} \sum_{l=0}^{h-1} y_{1k_{\mathrm{M}}-k_1+i} \eta_{k_{\mathrm{M}}+h+i-l} \\ \sum_{i=0}^{T_{\mathrm{M}}-1} \sum_{l=0}^{h-1} y_{2k_{\mathrm{M}}+i} \eta_{k_{\mathrm{M}}+h+i-l} \\ \vdots \\ \sum_{i=0}^{T_{\mathrm{M}}-1} \sum_{l=0}^{h-1} y_{mk_{\mathrm{M}}-k_m+i} \eta_{k_{\mathrm{M}}+h+i-l} \end{bmatrix}. \tag{47}$$

Multiplying the vector in eqn (47) by its transpose, we obtain a symmetric $S \times S$ matrix whose elements can easily be analysed using standard, but tedious, algebra. It can then be seen that $E\left(Y_1^{\mathrm{M}\prime} vv' Y_1^{\mathrm{M}}\right) = E(\{p_{ij}\})$ $i, j = 1, \ldots, k$ is an $S \times S$ matrix with the following elements:

$$\begin{aligned} E(\{p_{ij}\}) = {} & T_{\mathrm{M}} \gamma_{i,i} \left(\sum_{l=0}^{h-1} \kappa_l^2 \right) \\ & + 2 \sum_{r=1}^{h-1} (T_{\mathrm{M}} - r) \gamma_{i,j+r} \left(\sum_{l=0}^{h-r-1} \chi\, \kappa_{l,l+r} \right) \quad i, j = 1, \ldots, S, \end{aligned}$$

where $\gamma_{i,i}$ is the element of covariance matrix Γ_y occupying the position (i, j). Then, after some manipulation, it can be concluded that $E\left(Y_1^{\mathrm{M}\prime} vv' Y_1^{\mathrm{M}}\right) = \Gamma_y \mathcal{B}_h^{(T_{\mathrm{M}})}$, where $\mathcal{B}_h^{(T_{\mathrm{M}})} = [\{b_{ij}\}]$ is an $S \times S$ Toeplitz symmetric matrix with

$$\begin{aligned} [\{b_{ii}\}] &= T_{\mathrm{M}} \left(\sum_{l=0}^{h-1} \kappa_l^2 \right), \quad i = 1, \ldots, S \\ [\{b_{ij}\}] &= 2(T_{\mathrm{M}} - 1) \left(\sum_{l=0}^{h-j} \kappa_{l,l+j} \right), \quad i = 1, \ldots, S; \;\; j = i+1, \ldots, i+h-1 \\ [\{b_{ij}\}] &= 0, \;\; i = 1, \ldots, S; \;\; j = i+h, \ldots, S. \end{aligned}$$

Therefore,

$$\begin{aligned} \operatorname{trace}\left\{ T_{\mathrm{M}}^{-1} \Gamma_y^{-1} E\left(Y_1^{\mathrm{M}\prime} vv' Y_1^{\mathrm{M}} \right) \right\} &= \operatorname{trace}\left[T_{\mathrm{M}}^{-1} \Gamma_y^{-1} \Gamma_y \mathcal{B}_h^{(T_{\mathrm{M}})} \right] \\ &= S \sigma_{\mathrm{M}}^2(h), \end{aligned}$$

and hence

$$\operatorname{trace}\left\{ T_{\mathrm{M}}^{-1} \Gamma_y^{-1} E\left[Y_1^{\mathrm{M}\prime} v^{(L)} v^{(L)\prime} Y_1^{\mathrm{M}} \right] \right\} = S \sigma_{\mathrm{M}}^2(h) + o\left(T_{\mathrm{M}}^{-1/3} \right). \tag{48}$$

We should now solve the second term at the right-hand side of eqn (43). Let us denote as ξ_{ij} to an element of the $S \times S$ matrix Γ_y^{-1} and φ_{rs} to an element of the $S \times S$ matrix. Then,

$$E\left[\left(\hat{\Gamma}_y^{-1} - \Gamma_y^{-1} \right) Y_1^{\mathrm{M}\prime} v^{(L)} v^{(L)\prime} Y_1^{\mathrm{M}} \right] = O\left\{ S \sup_{i,j,r,s} E\left[\left(\hat{\xi}_{ij} - \xi_{ij} \right) \varphi_{rs} \right] \right\}.$$

By Holders' inequality,

$$E\left[\left\|\left(\hat{\xi}_{ij}-\xi_{ij}\right)\varphi_{rs}\right\|\right] \le E\left(\left\|\hat{\xi}_{ij}-\xi_{ij}\right\|^2\right)^{1/2} E\left(\|\varphi_{rs}\|^2\right)^{1/2}.$$

Since

$$\left\|\hat{\Gamma}_y^{-1}-\Gamma_y^{-1}\right\|^2 = \left\|\hat{\Gamma}_y^{-1}(\hat{\Gamma}_y-\Gamma_y)\Gamma_y^{-1}\right\|^2 \le \left\|\hat{\Gamma}_y^{-1}\right\|^2 \|(\hat{\Gamma}_y-\Gamma_y)\|^2 \left\|\Gamma_y^{-1}\right\|^2,$$

we have

$$E\left(\left\|\hat{\Gamma}_y^{-1}-\Gamma_y^{-1}\right\|^2\right) \le \left\|\Gamma_y^{-1}\right\|^2 \left[E\left(\left\|\hat{\Gamma}_y^{-1}\right\|^4\right)\right]^{1/2} \left[E\left(\|(\hat{\Gamma}_y-\Gamma_y)\|^4\right)\right]^{1/2}. \tag{49}$$

Since $\left\|\Gamma_y^{-1}\right\|$ is uniformly bounded above by a positive constant, and from Lewis and Reinsel (1985, p. 397), we have $\|\Gamma_y^{-1}\|^2 = O(S)$ and then $\|\xi_{ij}\|^2 = O(1)$, $E\left(\left\|\hat{\xi}_{ij}\right\|^4\right) = O(1)$. Moreover, by the asymptotic properties of OLS we have

$$E\left(\|(\hat{\gamma}_{i,j}-\gamma_{i,j})\|^4\right) = O(T_{\mathrm{M}}^{-2})$$

(Lewis and Reinsel, 1985a, 1988; Bhansali, 1981). Then, applying these results to eqn (49), we obtain

$$E\left(\left\|\hat{\xi}_{ij}-\xi_{ij}\right\|^2\right) = O(T_{\mathrm{M}}^{-1}).$$

Moreover, from eqn (48), we have $E(\|\varphi_{rs}\|^2) = O(1)$. Then,

$$E\left[\left\|\left(\hat{\xi}_{ij}-\xi_{ij}\right)\varphi_{rs}\right\|\right] = O(T_{\mathrm{M}}^{-1/2}).$$

Therefore, also applying Assumption 2,

$$\mathrm{trace}\left\{T_{\mathrm{M}}^{-1}E\left[\left(\hat{\Gamma}_y^{-1}-\Gamma_y^{-1}\right)Y_1^{\mathrm{M}\prime}v^{(L)}v^{(L)\prime}Y_1^{\mathrm{M}}\right]\right\} = O\left(ST_{\mathrm{M}}^{-3/2}\right) = o\left(T_{\mathrm{M}}^{-1}\right). \tag{50}$$

As a result, from eqns (43), (48) and (50), we have

$$E\left[v^{(L)\prime}Y_1^{\mathrm{M}}(Y_1^{\mathrm{M}\prime}Y_1^{\mathrm{M}})^{-1}Y_1^{\mathrm{M}\prime}v^{(L)}\right] = S\sigma_{\mathrm{M}}^2(h) + o\left(T_{\mathrm{M}}^{-1/3}\right) + o\left(T_{\mathrm{M}}^{-1}\right) \tag{51}$$

Finally, from eqns (40), (42) and (51)

$$E\left[\hat{\sigma}_{\mathrm{M}}^2(h)\right] = T_{\mathrm{M}}^{-1}(T_{\mathrm{M}}+S)\sigma_{\mathrm{M}}^2(h) + T_{\mathrm{M}}^{-1}S\sigma_{\mathrm{M}}^2(h) + o(T_{\mathrm{M}}^{-1}) = \sigma_{\mathrm{M}}^2(h) + o(T_{\mathrm{M}}^{-1}),$$

and the theorem is proved. □

ACKNOWLEDGEMENTS

The authors are thankful to a referee for valuable comments. This research was supported by MCYT grant no. SEJ2004-03303, Spain.

NOTE

Corresponding author: Ismael Sánchez, Avd de la Universidad 30, 28911, Leganés, Madrid, Spain. E-mail: ismael.sanchez@uc3m.es

REFERENCES

Akaike, H. (1970) Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* 21, 243–7.

Baillie, R. T. (1980) Predictions from ARMAX models. *Journal of Econometrics* 12, 365–74.

Baxter, G. (1963) A norm inequality for a "Finite Section" Wiener-Hopf equation. *Illinois Journal of Mathematics* 22, 203–17.

Bhansali, R. J. (1981) Effects of not knowing the order of an autoregressive process on the mean squared error of prediction-I. *Journal of the American Statistical Association* 76, 588–97.

Bhansali, R. J. (1993) Estimation of the prediction error variance and R^2 measure by autoregressive model fitting. *Journal of Time Series Analysis* 14, 125–46.

Bhansali, R. J. (1997) Direct autoregressive predictors for multistep prediction: order selection and performance relative to the plug in predictor. *Statistica Sinica* 7, 425–49.

Bhansali, R. J. (1999) Autoregressive model selection for multistep prediction. *Journal of Statistical Planning and Inference* 78, 295–305.

Box, G. and Jenkins, G. (1976) *Time Series Analysis*: *Forecasting and Control*. San Francisco: Holden Burg.

Box, G. and Tiao, G. (1977) A canonical analysis of multiple time series. *Biometrika* 64, 355–65.

Cheng, R. and Pourahmadi, M. (1993) Baxter's inequality and convergence of finite predictors of multivariate stochastic processess. *Probability Theory and Related Fields* 95, 115–24.

Fuller, W. A. and Hasza, D. P. (1981) Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association* 76, 155–61.

Granger, C. W. J. and Newbold, P. (1986) *Forecasting Economic Time Series*. Orlando, FL: Academic Press.

Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Lewis, R. and Reinsel, G. C. (1985a) Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16, 393–411.

Lewis, R. and Reinsel, G. C. (1985b) Prediction error of multivariate time series with mis-specified models. *Journal of Time Series Analysis* 9, 43–57.

Peña, D. and Box, G. (1987) Identifying a simplifying structure in time series. *Journal of the American Statistical Association* 82, 836–43.

Peña, D. Tiao, G. C. and Tsay, R. S. (2001) *A course in Time Series Analysis*. New York: John Wiley.

Pierce, D. A. (1979) R^2 measures for time series. *Journal of the American Statistical Association* 74, 901–10.

Reinsel, G. C. (1993) *Elements of Multivariate Time Series Analysis*. New York: Springer Verlag.

Reinsel, G. C. and Velu, R. P. (1998) *Multivariate Reduced-Rank Regression*. New York: Springer Verlag.

Tiao, G. C. and Tsay, R. S. (1989) Model specification in multivariate time series. *Journal of the Royal Statistical Society, Series* B 51, 157–213.

Yaglom, A. M. (1963) Stationary Gaussian processes satisfying the strong mixing condition and best predictable functions. Proc. Int. Res. Sem. of the Statistical Laboratory, University of California, Berkeley, CA, 241–52.

Yamamoto, T. (1976) Asymptotic mean squared prediction error for an autoregressive model with estimated coefficients. *Applied Statistics* 2, 123–7.

Yamamoto, T. (1981) Predictions of multivariate autoregressive-moving average models. *Biometrika* 68, 485–92.

Zellner, A. and Palm, F. (1974) Time series analysis and simultaneous equation econometric Models. *Journal of Econometrics* 2, 17–54.