Working Papers

07/2010

Determining the MSE-optimal cross section to forecast

Ignacio Arbués

First draft: December 2010

This draft: December 2010

Determining the MSE-optimal cross section to forecast

Abstract

We address the problem of which subset of time series to select among a given set in order to forecast another series. The forecasts are evaluated in terms of Mean Squared Error. We propose a family of criteria for which weak and strong consistency results are proved. The criteria are compared to some well-known hypothesis tests by means of Monte Carlo experimentation and a real-data example.

Keywords

forecasting, model selection, VARMA models

Authors and Affiliations

Ignacio Arbués

Dirección General de Metodología, Calidad y Tecnologías de la Información y las Comunicaciones, Instituto Nacional de Estadística

Determining the MSE-optimal cross section to forecast

Ignacio Arbués*

December 17, 2010

Abstract

In this paper, we address the question of which subset of time series should be selected among a given set in order to forecast another series. We evaluate the quality of the forecasts in terms of Mean Squared Error. We propose a family of criteria to estimate the optimal subset. Consistency results are proved, both in the weak (in probability) and strong (almost sure) sense. The results are generalized for the case in which there are more than one series to forecast. We present the results of a Monte Carlo experiment and a real data example in which the criteria are compared to some hypothesis tests such as the ones by Diebold and Mariano (1995), Clark and McCracken (2001 and 2007) and Giacomini and White (2006).

Key words: Forecasting; Model Selection; VARMA models.

JEL classification: C32; C52; C53.

1 Introduction

If we want to forecast a time series x_t^0 using time series models, we have to decide whether to use a univariate or a multivariate model, and in this latter case, which variables to include in the model. Once the composition of the vector time series has been decided, there are many tools to identify and estimate the model. Consequently, in this paper, we focus on the first two decisions.

*Ignacio Arbués, D. G. de Metodología, Calidad y Tecnologías de la Información y las Comunicaciones, Instituto Nacional de Estadística, Castellana, 183. 28071, Madrid, Spain, Tel: 34 915834641, E-mail address: iarbues@ine.es

More precisely, if we have a certain set of time series, which subset (hereinafter, 'subset' will mean a certain subset of the whole set of time series available) is the most convenient to forecast x_t^0? An easy answer to this question is to use all the time series, but if the number of series is large, the number of parameters of the models is also large (usually growing faster than linearly). In that case, the parameters are computationally difficult to estimate and even if the computational difficulties are overcome, the estimates may have large variances that render the models useless. Methods based in factor models try to limit the proliferation of parameters while using all the information of the panel. See, for example, Stock and Watson (2002) or Forni, Hallin, Lippi and Reichlin (2005). On the other hand, many classes of models like VARMA do not allow to include many series, since the number of parameter grows too fast.

If we want to use a model that does not allow large cross sections, a usual approach to the problem is to use hypothesis tests. For example, a two-sided test is described in Diebold and Mariano (1995), to compare the predictive efficiency of two models. In Clark and McCracken (2001, 2007), one-side tests are presented, that allow us to decide between nested models, rejecting the null when the most parsimonious one does not encompass the other. Granger-Causality tests (see Granger, 1969) are designed to determine if some series included in a certain subset are indeed useful to produce forecasts. Giacomini and White (2006) proposed a test of conditional predictive ability. In Peña and Sánchez (2007), a method to compare univariate and multivariate forecasts was presented.

In this paper, we present a different approach. We select the subset, or cross section, using a selection criterion rather than a hypothesis test. A great variety of model selection criteria have been proposed, for example, the AIC by Akaike (1973 and 1974), Schwarz's (1978) SBC or the HQ criterion by Hannan and Quinn (1979). The case of misspecification has been analyzed, among others, by Nishii (1988) and Sin and White (1996).

Instead of the penalized log-likelihood, our criteria consist of the logarithm of the mean squared $h-$step prediction error of x_t^0 plus penalty terms that take into account, not the number of parameters, but the size of the cross section. Thus, it is not covered by most of the references cited above. The exception is section 6 of Sin and White (1996), but we impose less stringent conditions on the penalty terms.

In section 2, we describe in detail the proble of the optimal cross section selection. We present in section 3 the class of criteria. Strong and weak consistency results are proved for a relatively general class of models in section 4 and in section 5 we show that the assumptions are satisfied in the case of VARMA models.

In section 6, some further generalizations are discussed; we consider in subsection 6.1 the case that the subset is selected among a random class and then, in subsection 6.2 the case that there are more than one series to forecast.

In order to assess the performance of the method, we have used Monte Carlo simulations to compare the criteria to some hypothesis tests. Specifically, we have considered the test by Diebold and Mariano and the ENC-T and ENC-NEW test of Clark and McCracken and the conditional predictive ability test by Giacomini and White. The results of this experiment are discussed in section 7. Finally, section 8 reports the results of an empirical application.

2 The optimal cross section

Suppose we want to forecast a certain time series x_t^0 at horizon h. For this purpose, besides x_t^0 itself, we have at our disposal a set of time series $\{x_t^i\}$ with $i = 1, \ldots, N$.

We also assume that for any subset $I \subset \mathcal{S} = \{0, \ldots, N\}$, such that $0 \in I$, there is a forecast $x_{t+h|t}^{0,I}$ of x_{t+h}^0 computed with the information contained in the series of I. In other words, $x_{t+h|t}^{0,I}$ is $\mathcal{F}_t(I)-$measurable, where $\mathcal{F}_t(I)$ is the $\sigma-$field generated by $\{x_s^I : s \leq t\}$, $x_s^I = (x_s^{i_1}, \ldots, x_s^{i_n})'$ and $I = \{i_1, \ldots, i_n\}$. In particular, $x_{t+h|t}^{0,I}$ is chosen as the optimal predictor in the sense that it minimizes

$$\sigma_h^2(I) = \mathbf{E}[(x_{t+h}^0 - x_{t+h|t}^{0,I})^2] \tag{1}$$

among a certain class of predictors (later, it will be the class of the linear predictors). The generalization to other loss functions remains for future investigation. The expression (1) is finite if x_{t+h}^0 and $x_{t+h|t}^{0,I}$ have bounded second-order moments and it is independent from t if x_t^I is strictly stationary. In the case of linear predictors, the condition can be relaxed to weak stationarity.

It is possible that some choices of I are ruled out in advance. Consequently, the selection is restricted to a certain class of subsets $\mathcal{I} \subset P(\{0, \ldots, N\})$. Now, we can state our problem, that is, to minimize σ_h^2 in $\mathcal{I}$. Let us denote by $\mathcal{I}_0$

the class of minimizers. In general, for any $I \in \mathcal{I}_0$, $I \subset J$ implies $J \in \mathcal{I}_0$, so the solution will not be unique. Therefore, it is natural to choose, among the multiple solutions, those most parsimonious in some sense. Let $\delta(I)$ be an integer function of I, such that if $I \subset J$, then $\delta(I) \leq \delta(J)$ and if $I \subsetneq J$, then $\delta(I) < \delta(J)$. For the sake of generality, we allow different possibilities for δ, but in our experiments we use the cardinality of I.

Consequently, our aim is to consistently estimate a subset I_0 that minimizes δ in $\mathcal{I}_0$. We call $\mathcal{I}_{00}$ the set of such minimizers. In general, even $\mathcal{I}_{00}$ may have more that one element, but in some cases uniqueness can be proved.

3 Criteria

In real life, rather than the optimal predictor of x^0_{t+h}, we will have an approximation, $\hat{x}^{0,I}_{t+h|t}$, typically computed with an estimated model, say for $t = 1, \ldots, T$.

We can define,

$$\hat{\varepsilon}^{0,I}_{t,h} = x^0_{t+h} - \hat{x}^{0,I}_{t+h|t}, \tag{2}$$

$$\hat{\sigma}^2_h(I) = \frac{1}{T}\sum_{t=1}^{T-h}(\hat{\varepsilon}^{0,I}_{t,h})^2 \tag{3}$$

and the family of criteria

$$\mathrm{FC}(I) = \log \hat{\sigma}^2_h(I) + \delta(I)\frac{S_T}{T}, \tag{4}$$

where S_T is a nondecreasing function of T whose properties will be prescribed in the following sections.

With this criteria, we choose the set $\hat{I}_T$ as

$$\hat{I}_T = \arg\min_{I \in \mathcal{I}} \mathrm{FC}(I). \tag{5}$$

The necessity of restraining the choice of $\hat{I}_T$ in (5) to a certain class $\mathcal{I}$ is due to the fact that the growth of $\#P(\{1, \ldots, N\}) = 2^N$ makes, even for moderate values of N, unfeasible to try all subsets. On the other hand, the assumption that $\mathcal{I}$ is always fixed in advance is not realistic. In some cases, $\mathcal{I}$ will be determined using the data of the series and thus it will be random. Nevertheless, in order to introduce the main ideas of the consistency results, we will present in section 4 the case of deterministic $\mathcal{I}$ and in 6.1 we describe the changes necessary to deal with the random case.

We have excluded the possibility of using more than one model for each subset I. In that case, a natural extension would be to replace $\hat{\sigma}_h^2(I)$ by the minimum MSE across models. This variation remains for future research.

4 Consistency

In this section we will establish some conditions under which the estimate $\hat{I}_T$ described in the previous section is consistent. Given that the set of optimal values, $\mathcal{I}_{00}$, may contain more than one element, we say that $\hat{I}_T$ is almost sure or strongly consistent if there exists with probability 1 some T_0 such that for any $T > T_0$, $\hat{I}_T \in \mathcal{I}_{00}$. Then, we write $\hat{I}_T \xrightarrow{a.s.} \mathcal{I}_{00}$. We say that $\hat{I}_T$ is consistent in probability if $P[\hat{I}_T \in \mathcal{I}_{00}] \to 1$ and we write $\hat{I}_T \xrightarrow{p} \mathcal{I}_{00}$.

Assumption 1. *The class $\mathcal{I}$ is closed with respect to union.*

This assumption is not unreasonable. If two sets, I and J contain relevant information to predict x_{t+h}^0, it is natural to try $I \cup J$, so that the predictions use both the information from I and J.

Assumption 2. *All x_t^I are weakly stationary and linearly regular*[1].

If assumption 2 holds, the Wold decomposition of x_t^I can be written as

$$x_t^I = \Psi(L)\varepsilon_{t-k}^I = \sum_{k=0}^{\infty} \Psi_k^I \varepsilon_{t-k}^I, \tag{6}$$

where ε_t^I are the linear innovations of x_t^I and L is the lag operator.

Assumption 3. *The following holds,*

(a) ε_t^I is ergodic, with bounded fourth-order moments, $\mathbf{E}[\varepsilon_t^I | \mathcal{F}_{t-1}(I)] = 0$, $\mathbf{E}[\varepsilon_t^I \varepsilon_t^{I\prime} | \mathcal{F}_{t-1}(I)] = \Sigma^I$, with $\Sigma^I > 0$.

(b) The continuation of $\Psi(z)$ to the unit circle has no unit modulus roots.

Under these assumptions, the best linear predictor of x_{t+h}^0 using $\{x_s^I : 1 \leq s \leq t\}$ is the best predictor, in the sense of mean squared error. We may, at

[1] As defined in Hannan and Deistler (1988). This property is also known as "linearly, purely nondeterministic".

least theoretically, also consider the best predictor using x_s^I with s from $-\infty$ to t. Let us write both predictors as

$$\mathbb{P}_{t,h}^{0,I} = \sum_{k=0}^{t-1} P_{t,h,k}^{0,I} L^k, \qquad \mathbb{P}_{h}^{0,I} = \sum_{k=0}^{\infty} P_{h,k}^{0,I} L^k, \tag{7}$$

where the coefficients in the second expression do not depend on t due to the stationarity of x_t^I. We can now write $x_{t+h|t}^{0,I}$ as $\mathbb{P}_{t,h}^{0,I}(L)x_t^I$.

On the other hand, if we do not know the true model, but rather an estimated one, we can compute the predictor coefficients as functions of the estimated model coefficients and then, the estimated predictors can be written as

$$\hat{x}_{t+h|t}^{0,I} = \hat{\mathbb{P}}_{t,h}^{0,I}(L)x_t^I = \sum_{k=0}^{t-1} \hat{P}_{t,h,k}^{0,I} x_{t-k}^I, \qquad \tilde{x}_{t+h|t}^{0,I} = \hat{\mathbb{P}}_{h}^{0,I}(L)x_t^I = \sum_{k=0}^{\infty} \hat{P}_{h,k}^{0,I} x_{t-k}^I, \tag{8}$$

where, of course, $\tilde{x}_{t,h}^{0,I}$ is a theoretical construct and we only introduce it for the proof of the consistency results.

Some conditions on the coefficients above are required.

Assumption 4. *The following holds,*

(a) $\hat{P}_{t,h,k}^{0,I} = P_{h,k}^{0,I} + v_{T,k} w_T$, *where* $|v_{T,k}| \leq r_k$ *uniformly in* T, *with* $\sum_k r_k < +\infty$ *and* $\sum_{k>s} r_k = O(s^{-\alpha})$, $\alpha > 0$ *and* $w_T = O(Q_T)$, *for* $Q_T = [\log\log T/T]^{1/2}$.

(b) With probability 1, uniformly for large T, $|\hat{P}_{t,h,k}^{0,I} - \hat{P}_{h,k}^{0,I}| \leq u_{t,k}, \forall t \leq T$, *where* $\sum_t^{\infty} \sum_k^{t-1} u_{t,k} < +\infty$.

Here we use $O(\cdot)$ and $o(\cdot)$ for almost sure order (later, we write $O_p(\cdot)$ for order in probability). This assumption is related to the consistency of the model estimates and to the decay of the predictor terms. The condition $\sum_k r_k < +\infty$ is restrictive, but allows for some forms of long memory, since the decay of the tail of the sum is allowed to be hyperbolic.

If almost sure convergence is not guaranteed, assumption 4 can be replaced by a weaker version, in probability (of course, with weak consistency), with $w_T = O_p(T^{-1/2})$. We denote the weak version as 4bis.

NOTE: we do not impose any relationship between the data used to compute the estimates and the data used to forecast, besides that the convergence of the estimated predictors depends on T. Thus, if we have a time series at our

disposal, we can use the whole series to estimate the model and to obtain the forecasting residuals of (3) or we can split the series into a length-T_e part to estimate the model and another one of length T to obtain the forecasting residuals. The assumptions hold as long as T_e and T are in an adequate relationship, e.g. $\lim_T T_e/T \in (0, +\infty)$. On the other hand, the models can be nested or nonnested and they can be identified by whatever method is preferred, as long as consistency is ensured.

We establish first the following rate of convergence.

Lemma 1. *If assumptions 1, 2, 3 and 4 hold, then for any* $I, J \in \mathcal{I}_0$, $\hat{\sigma}_h^2(I) - \hat{\sigma}_h^2(J) = O(Q_T^2)$. *If assumption 4bis holds instead of 4, then* $\hat{\sigma}_h^2(I) - \hat{\sigma}_h^2(J) = O_p(T^{-1})$.

We can state now the main result in this section,

Proposition 1. *The following holds.*

(i) If the assumptions of lemma 1 hold and $S_T/T \to 0$, $S_T/\log\log T \to +\infty$, *then* $\hat{I} \xrightarrow{a.s} \mathcal{I}_{00}$.

(ii) If assumption 4bis holds instead of 4 and $S_T/T \to 0$, $S_T \to +\infty$, *then* $\hat{I} \xrightarrow{p} \mathcal{I}_{00}$.

NOTE: if we extended our framework to allow for an infinite time series class $\mathcal{I}$, we could analyze the case that there is not a finite optimal cross section. In that case, the asymptotic optimality of the predictors could have more practical relevance than consistency. Something similar happens in the context of order determination of autoregressive models when the process is an AR(∞) (see Shibata, 1980).

With an additional assumption we can also prove uniqueness.

Proposition 2. *If assumptions 1–3 hold and in addition,* $\mathcal{I}$ *is closed with respect to intersection, then* $\mathcal{I}_{00}$ *has only one element.*

5 The VARMA case

In this section we show that in the framework of VARMA models, with quite usual methods of identification and estimation, we can use the criteria to consistently estimate the optimal cross section. We will consider models of the

form

$$\Phi(L)x_t^I = \Theta(L)\varepsilon_t^I, \tag{9}$$

where $\Phi(L)$ and $\Theta(L)$ are matrix polynomials and ε_t^I is a vector of innovations. For identification purposes, the class of the ARMA models is usually partitioned in subclasses indexed by a multi-index α in some different ways, for example,

(a) $\alpha = (p, q)$, where p and q are the degrees Φ and Θ.

(b) $\alpha = (\alpha_0, \ldots, \alpha_s)$ the Kronecker indices.

We denote the subclasses of the partition as $M(I, \alpha)$. Then, the identification of the ARMA model for x_t^I consists of estimating the multi-index $\hat{\alpha}(I)$ among a certain set $A(I)$. For example, we can use a consistent information criterion such as BIC. Then, $\hat{I}$ is estimated as described in section 3, $\hat{x}_{t+h|t}^{0,I}$ being the best linear predictor corresponding to the VARMA model estimated by maximum likelihood in $M(I, \hat{\alpha}(I))$.

Let us now express the assumption of good specification as,

Assumption 5. *For any $I \in \mathcal{I}$, there exist a multi-index $\alpha_0 \in A(I)$ and matrix polynomials $\Phi_0^I(z), \Theta_0^I(z)$ in $M(I, \alpha)$ such that,*

(a) $\Phi_0^I(L)x_t^I = \Theta_0^I(L)\varepsilon_t^I$.

(b) For any z such that $|z| \leq 1$, $|\Phi_0(z)|$, $|\Theta_0(z)| \neq 0$.

We need a technical lemma, before establishing the consistency of $\hat{I}$. In this lemma we omit for simplicity the superscript I.

Lemma 2. *Let us assume that Φ, Θ are such that $|\Phi(z)|, |\Theta(z)| \neq 0$ for any z, $|z| \leq 1$ and, $\hat{\Phi} = \Phi + O(Q_T)$, $\hat{\Theta} = \Theta + O(Q_T)$. Then,*

$$\|\hat{P}_{h,k}^0 - P_{h,k}^0\| \leq u_T v_k, \tag{10}$$

where $u_T = O(Q_T)$ and $v_k = O(\rho^k)$ and $0 < \rho < 1$.

With this lemma, we can prove the following proposition.

Proposition 3. *If assumptions 1, 2, 3 and 5 hold, then assumption 4 also holds and $\hat{I}_T \xrightarrow{a.s.} \mathcal{I}_{00}$.*

Lemma 2 and proposition 3 can be easily adapted to the case of convergence in probability.

6 Generalizations

In this section, we consider some variations of the framework described in the previous sections.

In 6.1, we consider the case that instead of choosing the subset among the elements a fixed $\mathcal{I}$, we have a random $\hat{\mathcal{I}}$. This generalization is necessary if there are so many series that a preliminary work is done in order to discard some subsets before using FC. If the pre-selection is done using the data of the series, then the subset is in fact selected among a random class. We provide some natural assumptions under which the FC-selected $\hat{I}_T$ is still consistent.

In 6.2, we analyze the case that we are interested in forecasting several of the time series available with the same multivariate model.

6.1 Random $\mathcal{I}$

We will now choose $\hat{I}$ as $\arg\min_{I \in \mathcal{I}_T} \mathrm{FC}(I)$, where $\mathcal{I}_T$ is random and possibly depending on the time series, whence the subscript. In order to achieve consistency, it is necessary to impose some constraints in the behavior of $\mathcal{I}_T$. In particular, we have to avoid the case that there are optimal subsets, say I, such that $I \in \mathcal{I}_T$ and $I \notin \mathcal{I}_T$ infinitely many times. For the sake of brevity, we restrict the analysis to the strong convergence results, but it can be easily adapted to the weak case.

Let us define,

$$\overline{\mathcal{I}}^p_\infty = \{I \in P(\{0,\dots,N\}) : P[I \in \limsup_T \mathcal{I}_T] = p\}, \tag{11}$$

$$\underline{\mathcal{I}}^p_\infty = \{I \in P(\{0,\dots,N\}) : P[I \in \liminf_T \mathcal{I}_T] = p\}, \tag{12}$$

where

$$\limsup_T A_T = \bigcap_{T=1}^{\infty} \bigcup_{S=T}^{\infty} A_S, \tag{13}$$

$$\liminf_T A_T = \bigcup_{T=1}^{\infty} \bigcap_{S=T}^{\infty} A_S. \tag{14}$$

We can now express more precisely the condition on $\mathcal{I}_T$.

Assumption 6. *$\underline{\mathcal{I}}^1_\infty \neq \emptyset$ and for any $J \in \bigcup_{p>0} \overline{\mathcal{I}}^p_\infty \setminus \underline{\mathcal{I}}^1_\infty$, $\sigma^2_h(I) \geq \sigma^2_{h,*}$, where*

$$\sigma^2_{h,*} = \min_{I \in \underline{\mathcal{I}}^1_\infty} \sigma^2_h(I). \tag{15}$$

If the inequality holds as equality, then $\delta(J) > \min\{\delta(I) : I \in \underline{\mathcal{I}}^1_\infty\}$.

With this assumption, we can focus on the set $\underline{\mathcal{I}}^1_\infty$. Let us denote by $\mathcal{I}_{\infty,0}$ the minimizers of σ^2_h in $\underline{\mathcal{I}}^1_\infty$ and by $\mathcal{I}_{\infty,00}$ the minimizers of δ in $\mathcal{I}_{\infty,0}$. Then, we can state the following proposition,

Proposition 4. *If 1, 2, 3, 4 and 6 hold, then* $\hat{I}_T \xrightarrow{a.s.} \mathcal{I}_{\infty,00}$.

With the generalization to the random $\mathcal{I}_T$, we can analyze some examples. We present one in which a scheme to build $\mathcal{I}_T$ using the data is combined with FC to provide a consistent estimate of the —in this case unique— element of $\mathcal{I}_{\infty,00}$. Then, we analyze other method that produces inconsistent estimates.

Example 1. *A consistent method to select* $\hat{I}_T$.

Let us consider the following scheme to build $\mathcal{I}_T$.

(i) $\tilde{I}^0 = \{0\}$.

(ii) For $k > 0$, $\tilde{I}^k = \tilde{I}^{k-1} \cup \{j\}$, where $j = \arg\min_{j \notin \tilde{I}^{k-1}} \hat{\sigma}^2_h(\tilde{I}^{k-1} \cup \{j\})$.

The process ends with $\tilde{I}^N = \{0, \ldots, N\}$ and then, $\mathcal{I}_T = \{\tilde{I}^0, \ldots, \tilde{I}^N\}$. In order to study the asymptotic behavior of $\mathcal{I}_T$, we define $I^0, \ldots, I^N$ according to (i) and (ii) but with σ^2_h instead of $\hat{\sigma}^2_h$. Let us write $\sigma^2_h(I^0) \geq \sigma^2_h(I^1) \geq \ldots \geq \sigma^2_h(I^{k_0}) = \ldots = \sigma^2_h(I^N)$, where for the sake of simplicity we also assume that the inequalities are strict.

It is easy to see that under the assumptions of section 4, w.p. 1, for any $j \leq k_0$, $\tilde{I}^j \to I^j$. Thus, $\{I^0, \ldots, I^{k_0}\} \subset \underline{\mathcal{I}}^1_\infty$ and for $\delta(I) = \#I$, assumption 6 holds. Then, $\hat{I}_T$ chosen among the elements of $\mathcal{I}_T$ according to FC is a consistent estimate of I^{k_0}.

Example 2. *An almost surely inconsistent method to select* $\hat{I}_T$.

We want to forecast x^0_{t+1} using information up to t with VAR models. We can use the scheme (i)-(ii) from the previous example, but now, we make a hypothesis test on $H_0 : \Phi_{0,k,l} = 0$, for $l = 1, \ldots, p$. That is, we test that the coefficients of x^l_t in the equation of x^0_t are null. Then if we reject H_0, we go on to the next iteration of (ii), but if we accept, then the process terminates and $\hat{I}_T = \tilde{I}^k$.

We can see that with probability 1, $\hat{I}_T \nrightarrow I^{k_0}$, where k_0 is as in example 1. Under the assumptions of proposition 1, the estimates $\hat{\phi}_{0,k,l}$ satisfy a Law of the Iterated Logarithm and then, w.p. 1,

$$\limsup_T \frac{\hat{\Phi}_{0,k,l} - \Phi_{0,k,l}}{Q_T} = a^{1/2}, \tag{16}$$

where a is the variance of the asymptotic (gaussian) distribution of $T^{1/2}(\hat{\Phi}_{0,k,l} - \Phi_{0,k,l})$. On the other hand, we reject at $100 \times (1-\alpha)\%$, when

$$\left| \frac{\hat{\Phi}_{0,k,l}}{T^{-1/2} a^{1/2}} \right| > \xi_\alpha, \tag{17}$$

where ξ_α is the value such that $P[Z > \xi_\alpha] = \alpha/2$ when Z is a a zero-mean, unit-variance Gaussian. From (16), we know that with probability 1, even if the null hypothesis holds, there exists a subsequence such that the left side of (17) diverges to infinity as $\log\log T$. Thus, with probability 1, $\hat{I}_T \neq I^{k_0}$ infinitely many times as $T \to \infty$.

6.2 Forecasting Multiple Series

Let us now introduce the case that we want to forecast several time series. In order to maintain as much as possible the notation of the previous sections, in this section the symbol x_t^0 mean $(z_t^1, \ldots, z_t^m)'$ and thus, $x_t^I = (z_t^1, \ldots, z_t^m, x_t^{i_1}, \ldots, x_t^{i_n})'$.

If we intend to apply the techniques of the previous sections to this case, it is necessary to adapt the criterion of optimality $\sigma_h^2(I) = \min_{J \in \mathcal{I}} \sigma_h^2(J)$. We can use the order relationship defined in the set S^+ of the symmetric positive semidefinite matrices by

$$A \prec B \Longleftrightarrow \exists C \in S^+, B = A + C. \tag{18}$$

Now, we can adapt the results of the $m = 1$ case. If $\Sigma_h(I)$ is now the matrix

$$\Sigma_h(I) = \mathbf{E}[\varepsilon_{t+h|t}^0 {\varepsilon_{t+h|t}^0}'], \tag{19}$$

where $\varepsilon_{t+h|t}^0$ is as in section 2, but now a vector, the symbol $\mathcal{I}_0$ denotes the set all $I \in \mathcal{I}$ such that $\Sigma_h(I) \prec \Sigma_h(J)$ for any J. If $\{0, \ldots, N\} \in \mathcal{I}$, then $\mathcal{I}_0$ is nonempty. Again, $\mathcal{I}_{00}$ is defined as in section 2.

The relationship $\prec$ in S^+ does not directly provide a criteria to select $\hat{I}$ because it is not a total order relationship, so we have to summarize the information of $\hat{\Sigma}_h(I)$ into some scalar. We can do it with a scalar function $\nu : S^+ \mapsto \mathbb{R}$.

We will study the asymptotic behavior of $\hat{I}$ with a quite general family of such functions. We only restrict ν in the following way,

Assumption 7. *ν satisfies the following properties,*

(a) It is strictly increasing, that is, for any $A, B \in S^+$ such that $A \prec B$, $\nu(A) \leq \nu(B)$ and if $A \prec B$ and $\nu(A) = \nu(B)$, then $A = B$.

(b) In any region $\mathcal{A}$ such that $\forall A \in \mathcal{A}, \det A \geq \mu > 0$, ν is Lipschitz with respect to some matrix norm $\|\dots\|$, that is, there exists $L(\mathcal{A}) > 0$ such that $|\nu(A) - \nu(B)| \leq L(\mathcal{A})\|A - B\|$.

It is obvious that the elements of $\mathcal{I}_0$ are minimizers of ν, but also if $\mathcal{I}_0 \neq \emptyset$, then every minimizer of ν, belongs to $\mathcal{I}_0$. Hence, if we define the criteria

$$\mathrm{FC}(I) = \nu(\hat{\Sigma}_h(I)) + \delta(I)\frac{S_T}{T}, \tag{20}$$

then we can consistently estimate $\mathcal{I}_{00}$ with $\hat{I}$ as in section 3.

Proposition 5. *The following holds,*

(i) If assumptions 1, 2, 3, 4 and 7 hold and $S_T/T \to 0$, $S_T/\log\log T \to +\infty$, then $\hat{I} \xrightarrow{a.s.} \mathcal{I}_{00}$.

(ii) If assumption 4bis holds instead of 4 and $S_T/T \to 0$, $S_T \to +\infty$, then $\hat{I} \xrightarrow{p} \mathcal{I}_{00}$.

Some examples of ν are,

$\nu_1(\Sigma) = \log\det\Sigma$.

$\nu_2(\Sigma) = \log\mathrm{tr}\Sigma$.

$\nu_3(\Sigma) = \log\Pi_{j=1}^m \Sigma_{jj}$.

Proposition 6. *Functions ν_1, ν_2 and ν_3 satisfy assumption 7.*

7 Monte Carlo experimentation

We have made a simulation experiment to assess the performance of the criteria in selecting the optimal cross section. Let us consider the following bivariate Data Generating Process,

$$\mathrm{DGP}_1 \left\{ \begin{array}{rcccc} x_t^0 & = & .5x_{t-1}^0 + & bx_{t-1}^1 & +\varepsilon_t^0 \\ x_t^1 & = & & .5x_{t-1}^1 & +\varepsilon_t^1 \end{array} \right. , \tag{21}$$

where $(\varepsilon_t^0, \varepsilon_t^1)'$ is a zero-mean Gaussian white noise with unit covariance matrix. When $b > 0$, the optimal cross section to forecast x_t^0 at horizons $h = 1, 2, 3$ is $I = \{0, 1\}$ and when $b = 0$, it is $I = \{0\}$. We want to assess the performance of the criteria for the selection of the optimal cross section by comparing them to several tests, namely the S_1 test of Diebold and Mariano (1995); the ENC-T and ENC-NEW tests described by Clark and McCracken (2001); the conditional predictive ability test by Giacomini and White (2006, GW) and a Granger-Causality test (GC) by Granger (1969). Note that the ENC-T and ENC-NEW cannot be applied to forecasting horizons greater than 1, whereas the Granger-causality test does not make an explicit distinction between forecasting horizons.

The performance of the tests is usually measured in terms of power and empirical size, but we are interested in measuring the frequency of correct selection of the optimal cross section. The acceptance/rejection is related to the choice of the cross section in different ways depending on the specific test. In the case of the encompassing tests and the Granger-causality test, we select $\hat{I} = \{0, 1\}$ if the test rejects the null and $\hat{I} = \{0\}$ if it does not. In the case of the equal forecast accuracy test S_1, we select $\hat{I} = \{0, 1\}$ if the test rejects and the error with $I = \{0\}$ is larger and we select $\hat{I} = \{0\}$ otherwise. For the GW tests of conditional predictive ability, we use the decision rule proposed in Giacomini and White (2006) with $c = 0$. We have to specify in advance the significance levels of the tests. For each of them, we set significance levels at 90% and at 95%.

Regarding the criteria, we have chosen $\delta(I)$ as the cardinality of I. Two penalty functions are considered, the BIC-like, $S_T = \log T$ and the HQ-like, $S_T = 2 \log \log T$. We denote the first as FC_1 and the second as FC_2. For the latter, we have not established strong consistency but only consistency in probability, but this is not relevant to the experiment.

As we explained in the note before lemma 1, the criteria can be computed either using the whole series to estimate and forecast or splitting the series into in-sample and out-of-sample parts. In this section and in the following one, we denote by FC_1^* and FC_2^* the criteria with out-of-sample forecasts. The parameter estimates for FC_1^* and FC_2^* and for tests S_1, ENC-T, ENC-NEW are obtained with the first 5/7ths of the observations and the out-of-sample forecasts with the last 2/7ths. The GW test is designed for a fixed or at least bounded estimation window. In our simulations, the window comprises the first 40 observations for

$T \leq 100$ and the first 100 for $T > 100$.

In order to obtain the forecasts, we have to determine models for $I = \{0\}$ and for $I = \{0, 1\}$. We have run the simulations in two different ways, (a) fitting AR(1) and VAR(1) models for $I = \{0\}$ and $I = \{0, 1\}$ respectively and (b) fitting AR(p) and VAR(p), with p selected by the BIC. When the order is selected by BIC, the order of the univariate model may be greater than the order of the multivariate one. Then, the models are nonnested and some of the tests cannot be applied. Furthermore, the results do not show a significantly different behavior with respect to the ones with fixed p. Thus, we have not included their results but they can be obtained from the author.

We generated $M = 5,000$ realizations of DGP_1 for $b = 0, 0.05, 0.1, 0.2$ and for different series lengths (50, 100, 200, 400, 800). In table 2, we represent the frequencies of selecting $\{0, 1\}$ for each of the combinations of b and length and for each of the tests or criteria.

We have designed additional scenarios to check the performance of the criteria under different conditions. In order to check the effect of heavy-tailed noise, we have run simulations of a process with the same autoregressive structure as DGP_1, but ε_t^0 and ε_t^1 are $t-$distributed with 4 degrees of freedom. We call this, DGP_2.

Part (a) of assumption 4 involves consistency of the estimated predictors to the optimal ones. As we saw in section 5, this holds for well-specified ARMA models. However, we want to assess the performance of the criteria when the models are misspecified and thus, consistency is not guaranteed by our theoretical results.

In the first misspecifiacion scenario, the DGP is a VAR(2). In this case, it only makes sense the fixed order ($p = 1$) estimation in order to preserve the misspecification condition.

$$\text{DGP}_3 \begin{cases} x_t^0 = 1/3x_{t-1}^0 + bx_{t-1}^1 + 2/9x_{t-2}^0 + \varepsilon_t^0 \\ x_t^1 = 1/3x_{t-1}^1 + 2/9x_{t-2}^1 + \varepsilon_t^1. \end{cases} \tag{22}$$

The next case, DGP_4 is as DGP_1, but ε_t^0 and ε_t^1 are GARCH,

$$\varepsilon_t^j = \sqrt{h_t}\xi_t \qquad h_t = \mu_h + \varphi h_{t-1} + \alpha(\varepsilon_{t-1}^j)^2, \tag{23}$$

with $\mu_h = .05$, $\varphi = .8$, $\alpha = .15$ and $\xi_t \sim WN(0, 1)$.

Finally, we generate a MA(1),

$$\text{DGP}_5 \begin{cases} x_t^0 &= \varepsilon_t^0 &+.5\varepsilon_{t-1}^0 &+b\varepsilon_{t-1}^1 \\ x_t^1 &= \varepsilon_t^1 & &+.5\varepsilon_{t-1}^1. \end{cases} \tag{24}$$

In tables 2-6 we present the frequency of selecting $\hat{I} = \{0, 1\}$.

The comparison between different selection methods, either tests or criteria involves both the probability of correctly selecting $\{0, 1\}$ when $b > 0$ and the probability of wrongly selecting $\{0, 1\}$ when $b = 0$. If both probabilities are greater for method A than for method B, we say that it is less conservative. If the probabilities of correct selection are greater for method A in both cases, we can say that A outperforms B.

By looking at the tables 2 and 3 we see that FC_1 is more conservative than FC_2 (but for $b = 0.2$), ENC-T at 90%, ENC-NEW at 90% (but for b=0.2), and GC and outperforms ENC-NEW and ENC-T at 95%, GW and DM except for $h = 3$ and b, T small.

On the other hand, FC_2 is less conservative than DM at 90% for $h = 1, 2$, DM at 95% and ENC-NEW, more conservative than GC and outperforms ENC-T and GW. The criteria FC_1^* and FC_2^*, not surprisingly, are more conservative versions of FC_1 and FC_2.

When the model is not correctly specified (tables 4-6), we see that, the relations are generally not very different, but if we are interested in forecasting at horizon $h > 0$, the GC can produce extremely bad results when the model is misspecified as in DGP_5 (in fact, this example was intentionally included to illustrate the risks of using the GC-test for this purpose).

8 Empirical example

In this section, following Clark and McCracken (2001), we will try to determine whether the unemployment rate is useful to improve the 1-step forecast of inflation. More specifically, the variable to forecast x_t^0 will now be the second difference of the logarithm of the USA CPI index without food and energy and x_t^1 will be the first difference of the unemployment rate among males between 25 and 54. The quarterly series run from 1958:Q3 to 1998:Q1.

Using the notation of the previous sections, we want to choose among $I = \{0\}$ and $I = \{0, 1\}$ as the optimal cross section. We will make our choice using the same criteria and tests as in section 7.

As in the previous section, we have obtained the criteria both using the whole series to estimate and forecast and splitting it into in-sample and out-of-sample parts.

All the tests but the Granger-causality require out-of-sample forecasts. We divide the series into periods 1958:Q3 to 1987:Q1 and 1987:Q2 to 1998:Q1, so that their lengths are in relationship of 1 to 0.4. We use a fixed scheme, that is, the models are identified and the parameters estimated with the data of the in-sample period and they remain fixed for all the out-of-sample period.

Instead of using only autoregressive models, we use ARMA and VARMA with autoregressive and moving average orders up to 3. According to the BIC criterion, with the data of the in-sample period we choose a MA(1) model for $I = \{0\}$ and a VMA(1) for $I = \{0, 1\}$. This is very convenient because then, the models are nested and we can use the encompassing tests ENC-T and ENC-NEW. The models identified and estimated with the in-sample period data are,

$$x_t^0 = \varepsilon_t^0 - 0.3298(0.0959)\varepsilon_{t-1}^0,$$

where the variance of ε_t^0 is 2.7882 and

$$\begin{cases} x_t^0 = \varepsilon_t^0 - 0.4415(0.0943)\varepsilon_{t-1}^0 - 1.3051(0.3798)\varepsilon_{t-1}^1 \\ x_t^1 = \varepsilon_t^1 + 0.0446(0.0187)\varepsilon_{t-1}^0 + 0.6864(0.0710)\varepsilon_{t-1}^1 \end{cases}$$

where the covariance matrix of $(\varepsilon_t^0, \varepsilon_t^1)'$ is,

$$\Sigma_\varepsilon = \begin{pmatrix} 2.5170 & -0.0844 \\ -0.0844 & 0.1178 \end{pmatrix}.$$

The numbers in parentheses are the standard errors of the estimates. The results of the tests and criteria that are computed with out-of-sample forecasts are in the upper part of table 1.

On the other hand, for the GC test and FC_1 and FC_2 criteria, we have identified and estimated the models with the whole series from 1958:Q3 to 1998:Q1. According to the BIC criterion, we choose now a MA(1) model for the univariate and a VAR(1) for the multivariate. The estimated univariate model is

$$x_t^0 = \varepsilon_t^0 - 0.3470(0.0546)\varepsilon_{t-1}^0,$$

where the variance of ε_t^0 is 1.4281 and the multivariate one is

$$\begin{cases} x_t^0 &= -0.3450(0.0749)x_{t-1}^0 &-1.1229(0.2888)x_{t-1}^1 + \varepsilon_t^0 \\ x_t^1 &= \quad 0.0550(0.0164)x_{t-1}^0 &+0.6548(0.0629)x_{t-1}^1 + \varepsilon_t^1, \end{cases}$$

with covariance matrix

$$\Sigma_\varepsilon = \begin{pmatrix} 1.9083 & -0.0629 \\ -0.0629 & 0.0911 \end{pmatrix}.$$

In the lower part of table 1 we include the results of the CG test and the criteria FC_1 and FC_2 with parameters and forecasts computed using the full sample.

The results agree with Clark and McCracken (2001) even if the conditions are slightly different. The encompassing tests and our criteria indicate that the unemployment is relevant to forecast inflation based on the out-of-sample forecasts (upper part of table 1). On the other hand, the GW and S_1 tests do not reject their respective null hypotheses of conditional and unconditional equal predictive ability.

The analysis without splitting of the series, which is summarized in the lower part of table 1, points in the same direction as the encompassing tests. Both criteria FC_1 and FC_2 yield lower values for $I = \{0, 1\}$ and the Granger-Causality test clearly rejects the null.

	test/crit.	statistic	10%-CV	crit. $I = \{0\}$	crit. $I = \{0, 1\}$
	FC_1^*			-1.018	-1.045
	FC_2^*			-1.042	-1.092
in-sample estimates	ENC-T	1.656	1.645		
out-of-sample forecasts	ENC-NEW	9.697	1.003		
	S_1	0.5481	1.645		
	GW	3.7862	4.605		
full sample	FC_1			0.7376	0.7083
for estimation	FC_2			0.7163	0.6859
and forecasting	GC	15.115	2.705		

Table 1: The first column indicates whether the series is split into in-sample (to estimate) and out-of-sample (to forecast) periods or we use the full-length series to estimate and forecast; the second column is the test or criteria; the third is the value of the test statistic; the fourth is the 10% critical value and the last two are the values of the criteria for the two possible cross sections.

A Proofs

Proof of Lemma 1. If we prove the lemma for $I \subset J$, then it is easy to see that it holds for any $I, J \in \mathcal{I}_0$. We just have to apply it in turn to $I \subset I \cup J$ and

$J \subset I \cup J$. Thus, with no loss of generality, we assume that $I \subset J = I \cup K$, where $I \cap K = \emptyset$.

We can decompose the predictor as $\mathbb{P}_h^{0,J}$ as $[\mathbb{P}_h^{0,J,1} : \mathbb{P}_h^{0,J,2}]$ in such way that $\mathbb{P}_h^{0,J}(L)x_t^J = \mathbb{P}_h^{0,J,1}(L)x_t^I + \mathbb{P}_h^{0,J,2}(L)x_t^K$. Since $\mathbb{P}_h^{0,J}$ is the least squares predictor, then the minimum of the quadratic functional,

$$(P, Q) \mapsto q(P, Q) = \mathbf{E}[x_{t+h}^0 - P(L)x_t^I - Q(L)x_t^K]^2, \tag{25}$$

is attained at $(\mathbb{P}_h^{0,J,1}, \mathbb{P}_h^{0,J,2})$ and the minimal value is $\sigma_h^2(J)$, but this value is also attained at $(\mathbb{P}_{t,h}^{0,I}, 0)$, because I is also in $\mathcal{I}_0$. If the functional q is strictly convex, then the minimum is unique and,

$$\mathbb{P}_h^{0,J,1} = \mathbb{P}_h^{0,I}, \quad \mathbb{P}_h^{0,J,2} = 0.$$

We can see that the strict convexity of q is equivalent to the condition that $P(L)x_t^I + Q(L)x_t^K = 0$ implies $P, Q = 0$, but this property holds when x_t^J is linearly regular (see Hannan and Poskitt, 1987) and Ψ does not have unit modulus roots.

Let us turn now to $\hat{\sigma}_h^2(J) - \hat{\sigma}_h^2(I)$. First, we can see that only $O(T^{-1})$ terms are neglected if we replace $\hat{\sigma}_h^2(I)$ by $\dot{\sigma}_h^2(I)$, where

$$\dot{\sigma}_h^2(I) = \frac{1}{T}\sum_{t=1}^{T-h}(\dot{\varepsilon}_{t,h}^{0,I})^2, \tag{26}$$

and

$$\dot{\varepsilon}_{t,h}^{0,I} = x_{t+h}^0 - \hat{\mathbb{Q}}_h^{0,I}(L)x_t^I = x_{t+h}^0 - \sum_{k=0}^{t-1}\hat{P}_{h,k}^{0,I}x_{t-k}^I. \tag{27}$$

Let us consider the difference

$$\dot{\sigma}_h^2(I) - \hat{\sigma}_h^2(I) = \frac{1}{T}\sum_{t=1}^{T-h}\left[(\dot{\varepsilon}_{t,h}^{0,I})^2 - (\hat{\varepsilon}_{t,h}^{0,I})^2\right]. \tag{28}$$

We can write

$$|(\dot{\varepsilon}_{t,h}^{0,I})^2 - (\hat{\varepsilon}_{t,h}^{0,I})^2| = |\dot{\varepsilon}_{t,h}^{0,I} - \hat{\varepsilon}_{t,h}^{0,I}| \cdot |\dot{\varepsilon}_{t,h}^{0,I} + \hat{\varepsilon}_{t,h}^{0,I}|, \tag{29}$$

and consequently,

$$|(\dot{\varepsilon}_{t,h}^{0,I})^2 - (\hat{\varepsilon}_{t,h}^{0,I})^2| \leq \sum_{k=0}^{t-1}|\hat{P}_{h,k}^{0,I} - \hat{P}_{t,h,k}^{0,I}| \cdot |x_t^I(\dot{\varepsilon}_{t,h}^{0,I} + \hat{\varepsilon}_{t,h}^{0,I})| \leq \tag{30}$$

$$\leq \sum_{k=0}^{t-1}u_{t,k}|x_t^I(\dot{\varepsilon}_{t,h}^{0,I} + \hat{\varepsilon}_{t,h}^{0,I})|. \tag{31}$$

Using assumption 4(a), we can bound $\dot{\varepsilon}^{0,I}_{t,h}$ and $\hat{\varepsilon}^{0,I}_{t,h}$ by random variables with uniformly bounded second-order moments. This implies that

$$|(\dot{\varepsilon}^{0,I}_{t,h})^2 - (\hat{\varepsilon}^{0,I}_{t,h})^2| \leq \zeta_t \tag{32}$$

$$\sum_{t}^{\infty} \mathbf{E}\zeta_t < +\infty. \tag{33}$$

Then, by theorem 2, page 66, in Gihman and Skorohod (1974), we get that $\sum_t[\ldots]$ in (28) is bounded with probability 1 and then $\dot{\sigma}^2_h(I) - \hat{\sigma}^2_h(I) = O(T^{-1})$. We can now use the $\dot{\sigma}^2$ terms instead of the $\hat{\sigma}^2$ ones, in the difference $\hat{\sigma}^2_h(J) - \hat{\sigma}^2_h(I)$. Then, we proceed as

$$\frac{1}{T}\sum_{t=1}^{T-h}\left[(\dot{\varepsilon}^{0,J}_{t,h})^2 - (\dot{\varepsilon}^{0,I}_{t,h})^2\right] = \frac{1}{T}\sum_{t=1}^{T-h}\left[\dot{\varepsilon}^{0,J}_{t,h} - \dot{\varepsilon}^{0,I}_{t,h}\right]\left[\dot{\varepsilon}^{0,J}_{t,h} + \dot{\varepsilon}^{0,I}_{t,h}\right]. \tag{34}$$

On the other hand, we can write

$$\dot{\varepsilon}^{0,I}_{t,h} = x^0_{t+h} - \hat{\mathbb{Q}}^{0,I,*}_h(L)x^J_t, \tag{35}$$

with $\hat{\mathbb{Q}}^{0,I,*}_h = [\hat{\mathbb{Q}}^{0,I}_h : 0]$ and $\hat{\mathbb{Q}}^{0,I}_h(L) = \sum_{k=1}^{t-1} P^{0,I}_{h,k}L^k$. Then, $\dot{\sigma}^2_h(J) - \dot{\sigma}^2_h(I)$ equals

$$\frac{1}{T}\sum_{t=1}^{T-h}\left[\sum_{k=1}^{t-1}\left(\hat{P}^{0,J}_{h,k} - \hat{P}^{0,I,*}_{h,k}\right)x^J_{t-k}\right]\left[2x^0_{t+h} - \sum_{l=1}^{t-1}\left(\hat{P}^{0,J}_{h,l} + \hat{P}^{0,I,*}_{h,l}\right)x^J_{t-l}\right]. \tag{36}$$

We will denote the first $[\ldots]$ factor as a_t whereas the second is decomposed as $b_t + c_t$, with

$$\begin{aligned} b_t &= 2\varepsilon^{0,J}_{t,h}, \\ c_t &= \sum_{l=0}^{t-1}\left(2P^{0,J}_{h,l} - \hat{P}^{0,J}_{h,l} - \hat{P}^{0,I,*}_{h,l}\right)x^J_{t-l}, \end{aligned}$$

where $\varepsilon^0_{t,h} = x^0_{t+h} - x^{0,J}_{t+h|t}$. Let us deal first with the product $(1/T)\sum_t a_t b_t$,

$$\frac{1}{T}\sum_{t=1}^{T-h} a_t b_t = 2\frac{1}{T}\sum_{t=1}^{T-h}\sum_{k=1}^{t-1}\left(\hat{P}^{0,J}_{h,k} - \hat{P}^{0,I,*}_{h,k}\right)x^J_{t-k}\varepsilon^{0,J}_{t,h}. \tag{37}$$

We can swap the order of summation and use that the difference in parentheses does not depend on t. Thus, it becomes

$$2\sum_{k=1}^{T-h-1}\left\{\left(\hat{P}^{0,J}_{h,k} - \hat{P}^{0,I,*}_{h,k}\right)\frac{1}{T}\sum_{t=k+1}^{T-h} x^J_{t-k}\varepsilon^{0,J}_{t,h}\right\} =: 2\sum_{k=1}^{T-h-1}\Delta_k s_{k,T}. \tag{38}$$

Let us write now

$$\left|\sum_k \frac{\Delta_k}{Q_T}\frac{s_{k,T}}{Q_T}\right| \leq M\sum_k r_k \frac{|s_{k,T}|}{Q_T} \leq \tag{39}$$

$$\leq M\left[\sum_{k\leq g(T)} r_k \frac{|s_{k,T}|}{Q_T} + \sum_{k>g(T)} r_k \frac{|s_{k,T}|}{Q_T}\right], \tag{40}$$

where the first inequality is due to assumption 4. If $g(T) = (\log T)^a$, then by lemma 5.3.5 in Hannan and Deistler (1988, hereafter, HD), we have that $\sup_{k\leq g(T)} |s_{k,T}| = O(Q_T)$. Thus, the first term inside the brackets in (40) is $O(1)$. On the other hand, $\sup_{0\leq k<\infty} |s_{k,T}| = O([\log T/T]^{1/2})$ by theorem 7.4.3, again in HD,

$$\sum_{k>(\log T)^a} r_k \frac{|s_{k,T}|}{Q_T} \leq M(\log T)^{-\alpha a}\left(\frac{\log T}{\log\log T}\right)^{1/2}. \tag{41}$$

Thus, if $a > 1/(2\alpha)$ then (39) is bounded with probability 1.

We put now $(1/T)\sum_t a_t c_t = \sum_{k,l}\Delta_k G_{k,l,T}\tilde{\Delta}'_l$, where $G_{k,l,T}$ is defined as $(1/T)\sum_t x^J_{t-k}{x^J_{t-l}}'$ and $\tilde{\Delta}_l := 2P^{0,J}_{h,l} - \hat{P}^{0,J}_{h,l} - \hat{P}^{0,I,*}_{h,l}$. Using that $G_{k,l,T}$ is almost surely uniformly bounded (this is implied by Theorem 5.3.2 in HD), then

$$\left|\sum_{k,l}\frac{\Delta_k}{Q_T}G_{k,l,T}\frac{\tilde{\Delta}'_l}{Q_T}\right| \leq M\left(\sum_k r_k\right)^2. \tag{42}$$

With this, the first part of the lemma is proved. For the order in probability, it is only necessary to replace Q_T by $T^{-1/2}$ and use that $T^{-1/2}\mathbf{E}|s_{k,T}|$ is uniformly bounded. Let us see this.

$$\mathbf{E}|s_{k,T}| = \mathbf{E}\left|\sum_{j=0}^{\infty}\Psi^J_j\frac{1}{T}\sum_{t=k+1}^{T-h}\varepsilon^J_{t-k-j}\varepsilon^{0,J}_{t,h}\right| \leq \tag{43}$$

$$\leq \left(\mathbf{E}\left[\sum_{j=0}^{\infty}\Psi^J_j\frac{1}{T}\sum_{t=k+1}^{T-h}\varepsilon^J_{t-k-j}\varepsilon^{0,J}_{t,h}\right]^2\right)^{1/2} \tag{44}$$

The term inside $(\cdot)^{1/2}$ can be written as

$$\frac{1}{T^2}\sum_{j,l}\sum_{t,s}\mathbf{E}\left[{\varepsilon^J_{t-k-j}}'\Psi^{J\prime}_j\Psi^J_l\varepsilon^J_{s-k-l}\varepsilon^{0,J}_{t,h}\varepsilon^{0,J}_{s,h}\right] \lessapprox \frac{\|\Sigma\|^2}{T}\left(\sum_{\nu=1}^{h}\|\Psi^J_\nu\|\right)\sum_j^{\infty}\|\Psi^J_j\|^2, \tag{45}$$

because of assumption 3 and the fact that $\varepsilon^{0,J}_{t,h}$ is the first component of vector $\sum_{\nu=0}^{h-1}\Psi^J_\nu\varepsilon^J_{t+h-\nu}$. □

Proof of Prop. 1. In order to prove strong consistency of $\hat{I}_T$ it suffices to prove that w. p. 1, every convergent subsequence converges to an element of $\mathcal{I}_{00}$. We avoid cumbersome notation by using $\hat{I}_T$ for a convergent subsequence. If $\hat{I}_T \to J$, we will show that necessarily $J \in \mathcal{I}_{00}$. Let us consider first the case $J \notin \mathcal{I}_0$ and then, $J \in \mathcal{I}_0 \setminus \mathcal{I}_{00}$.

If $J \notin \mathcal{I}_0$, then for any $I \in \mathcal{I}_0$, $\sigma_h^2(J) > \sigma_h^2(I)$. For large T, $\hat{I}_T = J$, so

$$\mathrm{FC}(\hat{I}_T) - \mathrm{FC}(I) = \log \hat{\sigma}_h^2(J) - \log \hat{\sigma}_h^2(I) + [\delta(J) - \delta(I)]\frac{S_T}{T}, \tag{46}$$

and the first difference in the right hand side converges to a strictly positive value, whereas the last term converges to zero. Thus, w.p. 1, for large T , $\mathrm{FC}(\hat{I}_T) - \mathrm{FC}(I) > 0$.

For the case $J \in \mathcal{I}_0 \setminus \mathcal{I}_{00}$ we need the order of convergence of $\log \hat{\sigma}_h^2(J) - \log \hat{\sigma}_h^2(I)$ established in lemma 1 for $I \in \mathcal{I}_0$, $I \subset J$. We can write

$$\log \hat{\sigma}_h^2(J) - \log \hat{\sigma}_h^2(I) = \log \left\{ 1 + \frac{\hat{\sigma}_h^2(J) - \hat{\sigma}_h^2(I)}{\sigma_h^2(I)} \right\}, \tag{47}$$

and by a first-order Taylor expansion, we obtain

$$\log \hat{\sigma}_h^2(J) - \log \hat{\sigma}_h^2(I) = [1 + o(1)] \left\{ \frac{\hat{\sigma}_h^2(J) - \hat{\sigma}_h^2(I)}{\sigma_h^2(I)} \right\}. \tag{48}$$

Since $\hat{\sigma}_h^2(J) - \hat{\sigma}_h^2(I) = O(Q_T^2)$,

$$\frac{\mathrm{FC}(J) - \mathrm{FC}(I)}{Q_T^2} = O(1) + \frac{S_T}{\log \log T}, \tag{49}$$

that diverges to $+\infty$ and then, for large T, $\hat{I}_T \neq J$.

The same arguments can be easily adapted to prove consistency in probability. □

Proof of Prop. 2. Let us assume that there exist I and J in $\mathcal{I}_{00}$ such that $I \neq J$. Let us consider a partition of the predictor of $K = I \cup J$ as $\mathbb{P}_h^{0,K} = (\mathbb{P}_h^{0,I \cap J}, \mathbb{P}_h^{0,I \setminus J}, \mathbb{P}_h^{0,J \setminus I})$. Now, proceeding as in the proof of lemma 1 we get that $\mathbb{P}_h^{0,J \setminus I} = 0$ and $\mathbb{P}_h^{0,I \setminus J} = 0$. Consequently $I \cap J$ has the same predictor as I, so $\sigma_h(I \cap J) = \sigma_h(I)$ and $I \cap J \in \mathcal{I}_0$, but $I \cap J \subsetneq I$. This implies that $\delta(I \cap J) < \delta(I)$, which contradicts the assumption that $I \in \mathcal{I}_{00}$. □

Proof of Lemma 2. If β is a vector containing the free coefficients of Φ and Θ, and $p_{h,k}^0 = \mathrm{vec}(P_{h,k}^0)$, $\hat{p}_{h,k}^0 = \mathrm{vec}(\hat{P}_{h,k}^0)$ we can write

$$\hat{p}_{h,k}^0 = p_{h,k}^0 + \frac{\partial p_{h,k}^0}{\partial \beta'}(\xi_{\hat{\beta}})(\hat{\beta} - \beta), \tag{50}$$

with $\xi_{\hat{\beta}}$ between β and $\hat{\beta}$. We will see now that $\partial p_{h,k}/\partial\beta'(\xi_{\hat{\beta}}) = O(\rho^k)$ uniformly in T. By introducing the rational function $\Psi_{[h]}(L) = \sum_{j=0}^{\infty}\Psi_{h+j}L^j$, we can express the $h-$step predictor of the vector satisfying the ARMA model as $\mathbb{P}_h = \Psi_{[h]}\Pi$, with $\Pi(L) = \Theta^{-1}(L)\Phi(L)$. We can differentiate $\mathbb{P}_h$ with respect to β (using rule 7 from the Appendix A13 in Lütkepohl, 1991) as

$$\frac{\partial \mathrm{vec}(\mathbb{P}_h)}{\partial\beta'} = (\mathbb{I}\otimes\Psi_{[h]})\frac{\partial\mathrm{vec}(\Pi)}{\partial\beta'} + (\Pi'\otimes I)\frac{\partial\mathrm{vec}(\Psi_{[h]})}{\partial\beta'}, \tag{51}$$

where $\mathbb{I}$ is the identity matrix. Applying in turn the same property to $\Psi_{[h]} = L^{-h}[\Phi(L)^{-1}\Theta(L) - \sum_{\nu=0}^{h}\Psi_\nu L^\nu]$ we get that

$$\begin{aligned} L^h\frac{\partial\mathrm{vec}(\Psi_{[h]})}{\partial\beta'} &= (\mathbb{I}\otimes\Phi^{-1})\frac{\partial\mathrm{vec}(\Theta)}{\partial\beta'} + \\ &+ (\Theta'\otimes\mathbb{I})\frac{\partial\mathrm{vec}(\Phi^{-1})}{\partial\beta'} - \sum_{\nu=0}^{h}\frac{\partial\mathrm{vec}(\Psi_\nu)}{\partial\beta'}L^\nu. \end{aligned} \tag{52}$$

Let us now consider the expressions above with Φ, Θ, etc. as functions of $\xi_{\hat{\beta}}$ and let $1 < r < \min\{r_\Phi, r_\Theta\}$ where r_Φ and r_Θ are the radii of convergence of $|\Phi(z)^{-1}|$ and $|\Theta(z)^{-1}|$ evaluated at $\xi_{\hat{\beta}} = \beta$. With probability 1, there exists a certain T_1 such that for $T > T_1$, $\xi_{\hat{\beta}}$ is so near to β that the radii of $|\Phi(z)^{-1}|$ and $|\Theta(z)^{-1}|$ are greater that r. Since the decay of the power series appearing in (51) and (52) depends only on Φ and Θ, and given that $\mathbb{P}_h^0$ as a vector is the first row of $\mathbb{P}_h$ we conclude that $\partial p_{h,k}^0/\partial\beta'(\xi_{\hat{\beta}}) = O((1/r')^k)$ for any $r' < r$.

□

Proof of Prop. 3. For any I, by theorem 5.5.1, page 205, in Hannan and Deistler (1988), $\hat{\alpha}(I) \to \alpha_0(I)$. Therefore, with probability 1, there exists some T_1 such that for any $T > T_1$, $\hat{\alpha}(I) = \alpha_0(I)$. Then, the estimates $\hat{\Phi}(z)$ and $\hat{\Theta}(z)$ satisfy a Law of Iterated Logarithm and we can apply lemma 2 and then, assumption 4(a) is granted.

We can use the state-space representation of the VARMA model to prove that assumption 4(b) holds,

$$\begin{aligned} x_t^I &= HZ_t \\ Z_t &= FZ_{t-1} + \xi_t. \end{aligned} \tag{53}$$

With this representation, we obtain that the predictions of x_{t+h}^I using either $\{x_s^I : -\infty < s \le t\}$ or $\{x_s^I : 1 \le s \le t\}$ satisfy

$$\begin{aligned} x_{t+h\,|t} &= HF^hZ_{t|t} \\ Z_{t|t} &= (F - K_tH)Z_{t-1|t-1} + K_tx_t^I. \end{aligned} \tag{54}$$

For the case of the estimates using all the past from $-\infty$ the relationships hold with the limit Kalman gain $K_t = K$. Then, it can be proved that if $\tilde{Z}_{t|t} = \mathbb{P}(L)x_t^I$ is the prediction of Z_t with $\{x_s^I : -\infty < s \leq t\}$ and $Z_{t|t} = \mathbb{P}_t(L)x_t^I$ is the prediction with $\{x_s^I : 1 \leq s \leq t\}$, then

$$\begin{aligned} \mathbb{P}_t(L) &= K_t + (F - K_t H)\mathbb{P}_{t-1}(L)L, \quad (55) \\ \mathbb{P}(L) &= K + (F - KH)\mathbb{P}(L)L. \quad (56) \end{aligned}$$

Under the assumptions on $\Phi(L)$ and $\Theta(L)$, the eigenvalues of $F - KH$ are inside the unit circle and $K_t \to K$ exponentially. We put $\|K_t - K\| = O(\rho^t)$, with $\rho \in (0,1)$. Then, if we denote by $\|\cdot\|_1$ the norm $\|\sum_k A_k L^k\| = \sum_k \|A_k\|$, then,

$$\begin{aligned} \|\mathbb{P}_t - \mathbb{P}\|_1 &\leq \|K_t - K\| + \|F - K_t H\| \cdot \|\mathbb{P}_{t-1} - \mathbb{P}\|_1 + \|K - K_t\| \cdot \|H\mathbb{P}\|_1 \leq \\ &\leq O(\rho^t) + \rho\|\mathbb{P}_{t-1} - \mathbb{P}\|_1 \leq \ldots \leq O(t\rho^t). \quad (57) \end{aligned}$$

Thus, the series $\sum_t^\infty \|F^h(\mathbb{P}_t - \mathbb{P})\|_1$ is convergent. If we consider the predictors with the estimated model, $\hat{\mathbb{P}}$ and $\hat{\mathbb{P}}_t$, then by continuity of the state-space representation, it can be proved that the bound is uniform for sufficiently large T. Consequently, assumption 4(b) yields.

□

Proof of Prop. 4. With probability 1, there exists T_1 such that for all $T > T_1$,

$$\underline{\mathcal{I}}_\infty^1 \subset \mathcal{I}_T \subset \bigcup_{p>0} \overline{\mathcal{I}}_\infty^p. \quad (58)$$

Let us assume with no loss of generality that $\hat{I} \to I_0$. By assumption 6, we can discard with probability 1 all elements in $\mathcal{P}(\{0,\ldots,N\}) \setminus \underline{\mathcal{I}}_\infty^1$ as possible limits. On the other hand, any I such that $\sigma_h^2(I) > \sigma_{h,*}^2$ can be ruled out. We conclude by applying lemma 1 with $\mathcal{I}_0 = \mathcal{I}_{\infty,0}$.

□

Proof of Prop. 5. First of all, lemma 1 applies to the $m > 1$ case without any changes. Then, the proof of proposition 1 can be easily adapted. It is only necessary to see that

$$\mathrm{FC}(J) - \mathrm{FC}(I) \geq -L\|\hat{\Sigma}_h^2(J) - \hat{\Sigma}_h^2(I)\| + [\delta(J) - \delta(I)]\frac{S_T}{T}, \quad (59)$$

where $\|\hat{\Sigma}_h^2(J) - \hat{\Sigma}_h^2(I)\| = O(Q_T^2)$. Then, for $I \in \mathcal{I}_{00}, J \in \mathcal{I}_0 \setminus \mathcal{I}_{00}$, $\mathrm{FC}(J) - \mathrm{FC}(I) \xrightarrow{a.s} +\infty$. □

Proof of Prop. 6. Let us see that ν_1, ν_2 and ν_3 satisfy assumption 7. The fact that ν_2 and ν_3 are nondecreasing is obvious, since the diagonal elements of any matrix in S^+ cannot be negative. On the other hand, for any $A \in S^+$, $\mathrm{tr}A = 0$ implies $A = 0$ and thus, they are strictly increasing. The property (a) for ν_1 is a consequence of theorem 18.1.1 in Harville (1997).

The Lipschitz condition is again obvious for ν_2, ν_3, whereas for ν_1 is due to the fact that the derivative of $\log\det\Sigma$ is Σ^{-1} and the norm of Σ^{-1} is bounded when Σ is bounded away from zero.

□

References

[1] Akaike H., 1973, Information Theory and an extension of the Maximum Likelihood Principle, in: B. N. Petrov and F. Csaki, (eds.), Second international symposium on information theory. Academiai Kiado: Budapest, pp. 267-281.

[2] Akaike, H., 1974, A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

[3] Clark T. E. and McCracken M. W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85-110.

[4] Clark T. E. and McCracken M. W., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291-311.

[5] Diebold F. and Mariano R., 1995. Comparing Predictive Accuracy. Journal of Business and Economics Statistics 13, 252-263.

[6] Forni M., Hallin M., Lippi F. and Reichlin L., 2005. The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. Journal of the American Statistical Association 100, 830-840.

[7] Giacomini R. and White H., 2006. Tests of conditional predictive ability. Econometrica 74(6), 1545-1578.

[8] Gihman I. I. and Skorohod A. V., 1974. The theory of Stochastic Processes. Springer, New York.

[9] Granger C. W. J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424-438.

[10] Hannan E. J. and Deistler M., 1988. The Statistical Theory of Linear Systems. John Wiley and Sons, New York.

[11] Hannan E. J. and Poskitt D. S., 1987. Unit canonical correlation between future and past. The Annals of Statistics 16, 784-79.

[12] Hannan E. J. and Quinn B. G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society B 41, 190-195.

[13] Harville, D. A., 1997. Matrix Algebra From a Statistician's Perspective. Springer, New York.

[14] Lütkepohl H., 1991. Introduction to Multiple Time Series. Springer, Berlin.

[15] Nishii, R., 1988. Maximum Likelihood Principle and Model Selection when the True Model Is Unspecified. Journal of Multivariate Analysis 27, 392-403.

[16] Peña D. and Sánchez I., 2007. Measuring the Advantages of Multivariate versus Univariate Forecasts. Journal of Time Series Analysis 28, 886-909.

[17] Schwarz G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461-464.

[18] Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. The Annals of Statistics 8, 147-164.

[19] Sin Ch.-Y. and White H., 1996. Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics 71, 207-225.

[20] Stock J. H. and Watson, M. W., 2002. Forecasting Using Principal Components From a Large Number of Predictors. Journal of the American Statistical Association 97, 1167-1179.

B Tables

	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
$FC_1(h=1)$	0.054	0.035	0.021	0.014	0.010	0.076	0.062	0.071	0.099	0.165	0.131	0.156	0.252	0.434	0.748	0.358	0.550	0.816	0.982	1.000
$FC_1(h=2)$	0.061	0.043	0.029	0.020	0.016	0.075	0.064	0.071	0.092	0.148	0.123	0.140	0.206	0.344	0.589	0.277	0.415	0.643	0.887	0.991
$FC_1(h=3)$	0.033	0.022	0.015	0.009	0.006	0.043	0.030	0.033	0.041	0.067	0.067	0.074	0.100	0.172	0.331	0.155	0.232	0.382	0.618	0.856
$FC_2(h=1)$	0.112	0.082	0.068	0.061	0.051	0.138	0.131	0.163	0.230	0.372	0.215	0.268	0.426	0.659	0.905	0.485	0.697	0.915	0.996	1.000
$FC_2(h=2)$	0.113	0.092	0.077	0.073	0.064	0.130	0.123	0.150	0.197	0.309	0.192	0.234	0.347	0.521	0.764	0.377	0.548	0.771	0.944	0.996
$FC_2(h=3)$	0.069	0.056	0.047	0.039	0.036	0.081	0.074	0.083	0.120	0.188	0.115	0.143	0.210	0.331	0.533	0.241	0.355	0.540	0.757	0.919
$FC_1^*(h=1)$	0.013	0.005	0.002	0.001	0.001	0.018	0.011	0.009	0.010	0.017	0.033	0.034	0.045	0.082	0.205	0.112	0.164	0.310	0.557	0.815
$FC_1^*(h=2)$	0.016	0.006	0.003	0.001	0.001	0.021	0.012	0.010	0.014	0.017	0.030	0.033	0.045	0.076	0.172	0.088	0.131	0.250	0.441	0.692
$FC_1^*(h=3)$	0.007	0.004	0.002	0.001	0.000	0.008	0.004	0.005	0.004	0.005	0.015	0.015	0.018	0.030	0.067	0.041	0.059	0.115	0.232	0.428
$FC_2^*(h=1)$	0.027	0.014	0.007	0.005	0.005	0.037	0.028	0.026	0.042	0.071	0.060	0.068	0.108	0.195	0.397	0.177	0.257	0.447	0.694	0.883
$FC_2^*(h=2)$	0.030	0.017	0.013	0.009	0.007	0.040	0.028	0.033	0.044	0.075	0.053	0.064	0.101	0.175	0.344	0.147	0.215	0.381	0.583	0.792
$FC_2^*(h=3)$	0.017	0.009	0.007	0.004	0.004	0.022	0.013	0.017	0.023	0.034	0.034	0.034	0.053	0.088	0.193	0.077	0.114	0.223	0.381	0.588
$\mathrm{DM}_{90}(h=1)$	0.043	0.034	0.029	0.034	0.032	0.047	0.052	0.053	0.067	0.105	0.061	0.078	0.098	0.149	0.235	0.110	0.151	0.225	0.340	0.528
$\mathrm{DM}_{90}(h=2)$	0.074	0.054	0.041	0.040	0.036	0.076	0.067	0.059	0.065	0.089	0.086	0.080	0.097	0.120	0.170	0.126	0.131	0.174	0.244	0.374
$\mathrm{DM}_{90}(h=3)$	0.097	0.073	0.055	0.049	0.041	0.099	0.081	0.066	0.060	0.075	0.106	0.091	0.090	0.091	0.124	0.134	0.124	0.138	0.163	0.229
$\mathrm{DM}_{95}(h=1)$	0.022	0.017	0.014	0.017	0.015	0.024	0.028	0.029	0.036	0.058	0.034	0.045	0.056	0.091	0.149	0.060	0.086	0.140	0.234	0.410
$\mathrm{DM}_{95}(h=2)$	0.048	0.029	0.020	0.020	0.017	0.049	0.038	0.033	0.034	0.051	0.057	0.045	0.055	0.067	0.104	0.083	0.077	0.103	0.152	0.262
$\mathrm{DM}_{95}(h=3)$	0.072	0.047	0.032	0.025	0.023	0.069	0.053	0.035	0.032	0.041	0.076	0.057	0.052	0.052	0.071	0.097	0.079	0.083	0.091	0.138
$\mathrm{GW}_{90}(h=1)$	0.052	0.045	0.041	0.040	0.040	0.056	0.047	0.052	0.066	0.083	0.064	0.062	0.085	0.117	0.176	0.098	0.125	0.171	0.241	0.418
$\mathrm{GW}_{90}(h=2)$	0.105	0.086	0.092	0.151	0.289	0.105	0.084	0.092	0.172	0.352	0.112	0.102	0.115	0.214	0.403	0.129	0.153	0.172	0.286	0.527
$\mathrm{GW}_{90}(h=3)$	0.113	0.129	0.143	0.190	0.344	0.108	0.139	0.149	0.218	0.374	0.115	0.148	0.154	0.248	0.419	0.128	0.169	0.185	0.265	0.464
$\mathrm{GW}_{95}(h=1)$	0.025	0.021	0.017	0.018	0.018	0.028	0.021	0.026	0.035	0.045	0.034	0.031	0.041	0.068	0.107	0.053	0.068	0.098	0.153	0.307
$\mathrm{GW}_{95}(h=2)$	0.077	0.054	0.049	0.082	0.201	0.077	0.055	0.050	0.097	0.249	0.080	0.069	0.065	0.128	0.289	0.098	0.103	0.102	0.167	0.386
$\mathrm{GW}_{95}(h=3)$	0.091	0.091	0.089	0.120	0.258	0.089	0.100	0.093	0.133	0.283	0.091	0.109	0.099	0.158	0.320	0.102	0.125	0.121	0.160	0.350
$\mathrm{ENC\text{-}T}_{90}$	0.120	0.111	0.098	0.104	0.102	0.134	0.142	0.163	0.214	0.320	0.175	0.223	0.319	0.466	0.675	0.318	0.455	0.663	0.877	0.982
$\mathrm{ENC\text{-}T}_{95}$	0.064	0.053	0.050	0.052	0.050	0.074	0.080	0.091	0.128	0.203	0.099	0.134	0.200	0.326	0.534	0.195	0.309	0.518	0.781	0.961
$\mathrm{ENC\text{-}NEW}_{90}$	0.050	0.039	0.031	0.033	0.031	0.064	0.063	0.091	0.145	0.260	0.101	0.147	0.268	0.485	0.786	0.278	0.473	0.774	0.965	0.999
$\mathrm{ENC\text{-}NEW}_{95}$	0.024	0.017	0.012	0.011	0.011	0.034	0.032	0.044	0.077	0.155	0.057	0.088	0.165	0.339	0.676	0.190	0.343	0.662	0.935	0.998
GC_{90}	0.130	0.115	0.109	0.107	0.102	0.161	0.162	0.218	0.314	0.495	0.239	0.312	0.502	0.745	0.945	0.506	0.742	0.948	0.998	1.000
GC_{95}	0.079	0.066	0.057	0.054	0.050	0.102	0.105	0.137	0.214	0.371	0.166	0.222	0.381	0.636	0.904	0.403	0.645	0.907	0.996	1.000

Table 2: Results of the simulations of DGP_1. The value of the parameter b is indicated in the first row and the length of the series in the second. The leftmost column indicates the test or criteria (with forecasting horizon h between parentheses; if h is not specified, $h = 1$ or it is not applicable). The figures in the remaining places are the frequencies of of selecting $\{0, 1\}$ for each combination of b, length and test/criteria.

	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
$FC_1(h=1)$	0.060	0.037	0.021	0.016	0.011	0.080	0.066	0.072	0.100	0.172	0.137	0.175	0.262	0.444	0.744	0.376	0.551	0.800	0.969	0.999
$FC_1(h=2)$	0.060	0.045	0.028	0.021	0.017	0.075	0.068	0.070	0.095	0.147	0.119	0.147	0.217	0.349	0.583	0.291	0.425	0.642	0.871	0.987
$FC_1(h=3)$	0.030	0.021	0.013	0.009	0.006	0.040	0.035	0.035	0.046	0.067	0.066	0.078	0.108	0.177	0.329	0.164	0.238	0.394	0.606	0.854
$FC_2(h=1)$	0.112	0.085	0.069	0.063	0.048	0.145	0.135	0.164	0.230	0.376	0.221	0.293	0.433	0.657	0.895	0.491	0.685	0.899	0.992	1.000
$FC_2(h=2)$	0.109	0.092	0.079	0.075	0.062	0.134	0.137	0.151	0.207	0.311	0.189	0.249	0.351	0.527	0.755	0.388	0.553	0.765	0.936	0.995
$FC_2(h=3)$	0.064	0.057	0.044	0.039	0.032	0.079	0.080	0.088	0.122	0.183	0.116	0.149	0.213	0.334	0.528	0.247	0.358	0.541	0.742	0.921
$FC_1^*(h=1)$	0.015	0.005	0.003	0.001	0.000	0.020	0.013	0.011	0.011	0.020	0.036	0.038	0.052	0.094	0.214	0.116	0.182	0.314	0.551	0.801
$FC_1^*(h=2)$	0.016	0.009	0.004	0.002	0.001	0.018	0.015	0.012	0.013	0.020	0.034	0.038	0.051	0.086	0.173	0.094	0.147	0.248	0.444	0.688
$FC_1^*(h=3)$	0.009	0.004	0.002	0.001	0.001	0.012	0.009	0.006	0.006	0.008	0.016	0.017	0.023	0.036	0.068	0.049	0.065	0.118	0.231	0.436
$FC_2^*(h=1)$	0.027	0.014	0.009	0.006	0.005	0.039	0.028	0.026	0.041	0.075	0.064	0.075	0.116	0.209	0.399	0.177	0.273	0.448	0.684	0.875
$FC_2^*(h=2)$	0.030	0.020	0.012	0.010	0.007	0.038	0.032	0.033	0.046	0.073	0.060	0.073	0.109	0.190	0.337	0.152	0.230	0.374	0.584	0.785
$FC_2^*(h=3)$	0.021	0.013	0.008	0.005	0.004	0.023	0.020	0.020	0.023	0.035	0.035	0.038	0.058	0.100	0.190	0.086	0.122	0.216	0.381	0.591
$\mathrm{DM}_{90}(h=1)$	0.037	0.036	0.029	0.034	0.032	0.044	0.050	0.054	0.071	0.108	0.059	0.077	0.108	0.163	0.235	0.101	0.159	0.232	0.346	0.534
$\mathrm{DM}_{90}(h=2)$	0.071	0.050	0.040	0.036	0.037	0.070	0.060	0.057	0.068	0.090	0.085	0.082	0.095	0.125	0.169	0.117	0.135	0.173	0.243	0.375
$\mathrm{DM}_{90}(h=3)$	0.095	0.072	0.054	0.044	0.040	0.097	0.076	0.071	0.065	0.074	0.106	0.095	0.091	0.098	0.120	0.125	0.122	0.129	0.154	0.233
$\mathrm{DM}_{95}(h=1)$	0.019	0.015	0.015	0.014	0.015	0.021	0.024	0.025	0.037	0.060	0.031	0.040	0.056	0.094	0.152	0.053	0.086	0.139	0.238	0.409
$\mathrm{DM}_{95}(h=2)$	0.043	0.027	0.018	0.016	0.017	0.044	0.032	0.029	0.035	0.047	0.051	0.047	0.053	0.071	0.101	0.072	0.075	0.099	0.143	0.260
$\mathrm{DM}_{95}(h=3)$	0.065	0.043	0.031	0.020	0.020	0.066	0.047	0.041	0.033	0.039	0.076	0.056	0.052	0.053	0.068	0.087	0.075	0.073	0.086	0.146
$\mathrm{GW}_{90}(h=1)$	0.043	0.038	0.040	0.038	0.035	0.050	0.043	0.047	0.063	0.086	0.057	0.062	0.085	0.122	0.178	0.092	0.118	0.170	0.264	0.426
$\mathrm{GW}_{90}(h=2)$	0.090	0.079	0.077	0.116	0.217	0.089	0.088	0.091	0.149	0.279	0.108	0.102	0.120	0.195	0.332	0.131	0.149	0.165	0.266	0.439
$\mathrm{GW}_{90}(h=3)$	0.101	0.115	0.125	0.169	0.278	0.099	0.133	0.135	0.189	0.317	0.108	0.143	0.148	0.206	0.351	0.121	0.159	0.183	0.252	0.403
$\mathrm{GW}_{95}(h=1)$	0.022	0.018	0.017	0.016	0.016	0.024	0.019	0.021	0.030	0.042	0.027	0.031	0.042	0.064	0.104	0.050	0.069	0.094	0.167	0.308
$\mathrm{GW}_{95}(h=2)$	0.065	0.048	0.041	0.061	0.145	0.068	0.057	0.044	0.083	0.192	0.080	0.067	0.069	0.101	0.221	0.095	0.098	0.095	0.158	0.304
$\mathrm{GW}_{95}(h=3)$	0.083	0.083	0.078	0.099	0.189	0.079	0.095	0.078	0.113	0.221	0.087	0.101	0.088	0.121	0.247	0.093	0.116	0.112	0.154	0.280
$\mathrm{ENC\text{-}T}_{90}$	0.118	0.113	0.103	0.103	0.105	0.135	0.148	0.175	0.231	0.332	0.180	0.242	0.332	0.484	0.679	0.324	0.478	0.669	0.861	0.974
$\mathrm{ENC\text{-}T}_{95}$	0.059	0.056	0.050	0.051	0.050	0.071	0.082	0.095	0.132	0.215	0.100	0.140	0.211	0.349	0.546	0.192	0.321	0.517	0.760	0.945
$\mathrm{ENC\text{-}NEW}_{90}$	0.055	0.043	0.036	0.033	0.032	0.071	0.073	0.094	0.147	0.268	0.114	0.169	0.282	0.491	0.779	0.287	0.482	0.750	0.946	0.997
$\mathrm{ENC\text{-}NEW}_{95}$	0.029	0.020	0.014	0.014	0.013	0.039	0.038	0.048	0.083	0.167	0.070	0.100	0.182	0.360	0.673	0.202	0.366	0.640	0.908	0.994
GC_{90}	0.172	0.143	0.126	0.117	0.103	0.192	0.180	0.223	0.306	0.489	0.250	0.331	0.491	0.742	0.946	0.482	0.719	0.940	0.998	1.000
GC_{95}	0.126	0.091	0.078	0.068	0.059	0.138	0.122	0.152	0.213	0.369	0.184	0.233	0.376	0.628	0.900	0.384	0.611	0.892	0.994	1.000

Table 3: Results of the simulations of DGP_2. The value of the parameter b is indicated in the first row and the length of the series in the second. The leftmost column indicates the test or criteria (with forecasting horizon h between parentheses; if h is not specified, $h = 1$ or it is not applicable). The figures in the remaining places are the frequencies of of selecting $\{0, 1\}$ for each combination of b, length and test/criteria.

	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
$FC_1(h=1)$	0.068	0.045	0.034	0.021	0.015	0.100	0.106	0.138	0.214	0.400	0.210	0.290	0.486	0.774	0.972	0.515	0.774	0.965	0.999	1.000
$FC_1(h=2)$	0.045	0.033	0.026	0.017	0.012	0.089	0.106	0.165	0.281	0.532	0.222	0.349	0.594	0.881	0.994	0.591	0.870	0.991	1.000	1.000
$FC_1(h=3)$	0.031	0.021	0.016	0.009	0.007	0.054	0.065	0.092	0.158	0.325	0.139	0.216	0.388	0.695	0.948	0.384	0.656	0.924	0.998	1.000
$FC_2(h=1)$	0.128	0.103	0.088	0.077	0.069	0.173	0.194	0.261	0.401	0.639	0.314	0.434	0.658	0.898	0.994	0.647	0.872	0.988	1.000	1.000
$FC_2(h=2)$	0.091	0.077	0.066	0.061	0.057	0.150	0.192	0.294	0.479	0.738	0.316	0.491	0.744	0.954	0.999	0.703	0.929	0.998	1.000	1.000
$FC_2(h=3)$	0.060	0.053	0.051	0.043	0.040	0.100	0.129	0.195	0.328	0.565	0.215	0.338	0.569	0.850	0.986	0.501	0.781	0.970	1.000	1.000
$FC_1^*(h=1)$	0.017	0.007	0.004	0.001	0.001	0.026	0.017	0.020	0.031	0.067	0.057	0.071	0.126	0.248	0.490	0.176	0.308	0.533	0.788	0.952
$FC_1^*(h=2)$	0.011	0.006	0.004	0.001	0.001	0.020	0.018	0.027	0.045	0.101	0.052	0.079	0.162	0.349	0.680	0.194	0.383	0.686	0.928	0.994
$FC_1^*(h=3)$	0.005	0.002	0.001	0.000	0.000	0.009	0.009	0.006	0.016	0.029	0.022	0.033	0.064	0.158	0.418	0.088	0.186	0.424	0.779	0.967
$FC_2^*(h=1)$	0.032	0.018	0.012	0.007	0.006	0.049	0.041	0.054	0.089	0.184	0.099	0.127	0.226	0.420	0.665	0.262	0.421	0.659	0.857	0.970
$FC_2^*(h=2)$	0.026	0.015	0.015	0.007	0.007	0.042	0.043	0.068	0.120	0.250	0.093	0.143	0.289	0.533	0.824	0.289	0.509	0.796	0.960	0.997
$FC_2^*(h=3)$	0.011	0.007	0.005	0.003	0.002	0.021	0.021	0.028	0.055	0.121	0.046	0.069	0.147	0.333	0.667	0.151	0.295	0.589	0.880	0.985
$DM_{90}(h=1)$	0.048	0.043	0.043	0.041	0.037	0.056	0.061	0.080	0.106	0.163	0.091	0.109	0.161	0.231	0.340	0.139	0.204	0.303	0.481	0.728
$DM_{90}(h=2)$	0.090	0.064	0.059	0.051	0.047	0.107	0.097	0.121	0.168	0.250	0.147	0.175	0.249	0.366	0.565	0.244	0.350	0.514	0.748	0.945
$DM_{90}(h=3)$	0.107	0.080	0.065	0.057	0.052	0.113	0.110	0.111	0.149	0.217	0.152	0.162	0.222	0.318	0.502	0.212	0.283	0.405	0.640	0.878
$DM_{95}(h=1)$	0.025	0.023	0.023	0.020	0.020	0.031	0.034	0.047	0.062	0.099	0.052	0.063	0.097	0.150	0.243	0.082	0.129	0.206	0.360	0.612
$DM_{95}(h=2)$	0.061	0.039	0.035	0.027	0.023	0.071	0.060	0.075	0.107	0.167	0.102	0.113	0.160	0.258	0.449	0.167	0.241	0.380	0.638	0.901
$DM_{95}(h=3)$	0.079	0.054	0.037	0.031	0.028	0.081	0.073	0.065	0.090	0.139	0.110	0.107	0.137	0.211	0.375	0.148	0.184	0.277	0.496	0.796
$GW_{90}(h=1)$	0.065	0.052	0.060	0.069	0.071	0.066	0.068	0.075	0.114	0.171	0.083	0.102	0.135	0.206	0.303	0.129	0.156	0.246	0.382	0.648
$GW_{90}(h=2)$	0.118	0.089	0.080	0.095	0.176	0.122	0.119	0.102	0.184	0.343	0.140	0.175	0.204	0.319	0.546	0.213	0.292	0.415	0.629	0.876
$GW_{90}(h=3)$	0.121	0.124	0.094	0.115	0.189	0.118	0.135	0.126	0.169	0.325	0.136	0.175	0.196	0.285	0.478	0.176	0.254	0.331	0.493	0.770
$GW_{95}(h=1)$	0.033	0.025	0.032	0.035	0.040	0.036	0.033	0.037	0.064	0.106	0.045	0.052	0.077	0.120	0.205	0.077	0.089	0.153	0.262	0.521
$GW_{95}(h=2)$	0.089	0.058	0.044	0.049	0.111	0.095	0.083	0.061	0.105	0.224	0.109	0.122	0.127	0.211	0.410	0.172	0.220	0.303	0.496	0.810
$GW_{95}(h=3)$	0.099	0.089	0.058	0.062	0.116	0.091	0.098	0.079	0.098	0.203	0.107	0.127	0.127	0.182	0.332	0.139	0.190	0.235	0.361	0.667
ENC-T_{90}	0.130	0.119	0.116	0.119	0.113	0.155	0.177	0.224	0.318	0.468	0.236	0.305	0.461	0.654	0.865	0.407	0.604	0.828	0.970	0.999
ENC-T_{95}	0.070	0.066	0.066	0.064	0.058	0.088	0.100	0.137	0.210	0.339	0.144	0.197	0.322	0.519	0.772	0.264	0.444	0.710	0.932	0.998
ENC-NEW_{90}	0.054	0.047	0.045	0.042	0.039	0.081	0.099	0.153	0.269	0.487	0.156	0.255	0.472	0.764	0.961	0.399	0.679	0.930	0.996	1.000
ENC-NEW_{95}	0.028	0.020	0.021	0.016	0.015	0.045	0.049	0.084	0.167	0.352	0.097	0.164	0.345	0.652	0.927	0.287	0.558	0.881	0.993	1.000
GC_{90}	0.149	0.139	0.132	0.126	0.124	0.200	0.235	0.335	0.503	0.742	0.343	0.499	0.734	0.943	0.998	0.714	0.923	0.996	1.000	1.000
GC_{95}	0.096	0.082	0.079	0.070	0.069	0.133	0.157	0.234	0.385	0.644	0.257	0.389	0.635	0.900	0.995	0.620	0.880	0.992	1.000	1.000

Table 4: Results of the simulations of DGP_3. The value of the parameter b is indicated in the first row and the length of the series in the second. The leftmost column indicates the test or criteria (with forecasting horizon h between parentheses; if h is not specified, $h = 1$ or it is not applicable). The figures in the remaining places are the frequencies of of selecting $\{0, 1\}$ for each combination of b, length and test/criteria.

	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
$FC_1(h=1)$	0.066	0.041	0.024	0.016	0.010	0.088	0.076	0.084	0.103	0.191	0.161	0.197	0.285	0.460	0.740	0.404	0.557	0.775	0.947	0.998
$FC_1(h=2)$	0.066	0.046	0.033	0.022	0.016	0.081	0.073	0.081	0.094	0.152	0.135	0.161	0.224	0.352	0.588	0.300	0.429	0.634	0.845	0.980
$FC_1(h=3)$	0.038	0.023	0.016	0.010	0.008	0.048	0.038	0.040	0.045	0.071	0.078	0.086	0.119	0.187	0.336	0.170	0.246	0.392	0.609	0.834
$FC_2(h=1)$	0.123	0.099	0.074	0.062	0.048	0.156	0.153	0.179	0.231	0.385	0.251	0.309	0.440	0.656	0.890	0.514	0.683	0.874	0.980	1.000
$FC_2(h=2)$	0.117	0.099	0.085	0.065	0.063	0.139	0.141	0.164	0.203	0.307	0.210	0.256	0.356	0.529	0.754	0.397	0.544	0.751	0.914	0.991
$FC_2(h=3)$	0.070	0.056	0.048	0.040	0.035	0.086	0.083	0.097	0.119	0.188	0.127	0.156	0.226	0.335	0.527	0.243	0.361	0.533	0.737	0.907
$FC_1^*(h=1)$	0.013	0.008	0.003	0.002	0.001	0.020	0.014	0.013	0.014	0.024	0.040	0.044	0.067	0.114	0.227	0.139	0.199	0.334	0.538	0.788
$FC_1^*(h=2)$	0.016	0.011	0.005	0.004	0.001	0.021	0.017	0.014	0.016	0.025	0.037	0.042	0.062	0.100	0.192	0.107	0.154	0.272	0.442	0.668
$FC_1^*(h=3)$	0.010	0.007	0.003	0.002	0.000	0.012	0.009	0.009	0.008	0.011	0.022	0.021	0.029	0.044	0.083	0.049	0.079	0.133	0.246	0.427
$FC_2^*(h=1)$	0.029	0.017	0.009	0.006	0.004	0.039	0.031	0.034	0.046	0.082	0.071	0.083	0.129	0.222	0.407	0.206	0.280	0.453	0.661	0.863
$FC_2^*(h=2)$	0.030	0.022	0.014	0.010	0.006	0.038	0.033	0.036	0.046	0.075	0.067	0.077	0.118	0.197	0.349	0.158	0.229	0.380	0.572	0.769
$FC_2^*(h=3)$	0.023	0.016	0.007	0.007	0.004	0.026	0.023	0.023	0.025	0.043	0.042	0.046	0.065	0.109	0.207	0.087	0.132	0.230	0.384	0.575
$\mathrm{DM}_{90}(h=1)$	0.040	0.038	0.034	0.033	0.028	0.047	0.046	0.056	0.070	0.106	0.064	0.082	0.113	0.167	0.238	0.115	0.158	0.240	0.358	0.543
$\mathrm{DM}_{90}(h=2)$	0.072	0.054	0.043	0.038	0.037	0.079	0.059	0.057	0.069	0.086	0.090	0.087	0.098	0.132	0.175	0.119	0.137	0.177	0.255	0.377
$\mathrm{DM}_{90}(h=3)$	0.094	0.075	0.051	0.048	0.041	0.102	0.080	0.066	0.063	0.075	0.110	0.093	0.089	0.101	0.121	0.126	0.124	0.132	0.165	0.227
$\mathrm{DM}_{95}(h=1)$	0.021	0.020	0.017	0.017	0.014	0.026	0.025	0.031	0.037	0.063	0.035	0.047	0.067	0.106	0.157	0.067	0.096	0.157	0.259	0.427
$\mathrm{DM}_{95}(h=2)$	0.045	0.031	0.021	0.018	0.018	0.050	0.034	0.028	0.036	0.048	0.059	0.052	0.056	0.074	0.108	0.078	0.082	0.107	0.163	0.261
$\mathrm{DM}_{95}(h=3)$	0.066	0.048	0.026	0.025	0.019	0.074	0.051	0.039	0.033	0.038	0.079	0.059	0.050	0.054	0.065	0.088	0.079	0.079	0.096	0.141
$\mathrm{GW}_{90}(h=1)$	0.050	0.038	0.045	0.038	0.037	0.060	0.049	0.050	0.071	0.088	0.067	0.075	0.088	0.119	0.182	0.099	0.118	0.179	0.269	0.435
$\mathrm{GW}_{90}(h=2)$	0.102	0.078	0.077	0.108	0.191	0.102	0.089	0.082	0.130	0.251	0.111	0.104	0.117	0.171	0.307	0.141	0.149	0.161	0.243	0.423
$\mathrm{GW}_{90}(h=3)$	0.116	0.117	0.123	0.150	0.241	0.110	0.122	0.125	0.168	0.268	0.110	0.132	0.139	0.178	0.314	0.130	0.151	0.162	0.228	0.364
$\mathrm{GW}_{95}(h=1)$	0.022	0.017	0.019	0.016	0.018	0.032	0.023	0.022	0.029	0.044	0.037	0.038	0.041	0.061	0.108	0.055	0.066	0.110	0.177	0.320
$\mathrm{GW}_{95}(h=2)$	0.074	0.051	0.044	0.057	0.114	0.074	0.057	0.046	0.070	0.156	0.080	0.067	0.069	0.092	0.195	0.105	0.100	0.101	0.150	0.288
$\mathrm{GW}_{95}(h=3)$	0.093	0.079	0.076	0.086	0.158	0.084	0.092	0.084	0.103	0.174	0.088	0.099	0.090	0.104	0.206	0.103	0.111	0.103	0.133	0.241
$\mathrm{ENC\text{-}T}_{90}$	0.118	0.116	0.102	0.105	0.095	0.134	0.142	0.176	0.222	0.328	0.189	0.246	0.335	0.486	0.680	0.341	0.473	0.658	0.834	0.963
$\mathrm{ENC\text{-}T}_{95}$	0.063	0.061	0.051	0.053	0.047	0.073	0.079	0.101	0.135	0.215	0.111	0.150	0.222	0.360	0.547	0.221	0.345	0.527	0.747	0.929
$\mathrm{ENC\text{-}NEW}_{90}$	0.058	0.051	0.043	0.039	0.034	0.075	0.079	0.112	0.155	0.276	0.125	0.189	0.292	0.492	0.763	0.317	0.482	0.722	0.913	0.990
$\mathrm{ENC\text{-}NEW}_{95}$	0.031	0.027	0.019	0.018	0.013	0.041	0.046	0.062	0.092	0.177	0.080	0.119	0.203	0.373	0.658	0.231	0.374	0.624	0.869	0.985
GC_{90}	0.201	0.184	0.163	0.135	0.124	0.214	0.211	0.240	0.300	0.476	0.266	0.331	0.469	0.717	0.946	0.474	0.695	0.929	0.996	1.000
GC_{95}	0.150	0.138	0.114	0.090	0.074	0.163	0.158	0.176	0.218	0.361	0.210	0.250	0.369	0.606	0.896	0.386	0.591	0.876	0.993	1.000

Table 5: Results of the simulations of DGP_4. The value of the parameter b is indicated in the first row and the length of the series in the second. The leftmost column indicates the test or criteria (with forecasting horizon h between parentheses; if h is not specified, $h = 1$ or it is not applicable). The figures in the remaining places are the frequencies of of selecting $\{0, 1\}$ for each combination of b, length and test/criteria.

	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
$FC_1(h=1)$	0.065	0.042	0.030	0.018	0.012	0.064	0.051	0.042	0.048	0.062	0.090	0.087	0.108	0.154	0.270	0.176	0.231	0.373	0.625	0.906
$FC_1(h=2)$	0.046	0.032	0.021	0.014	0.010	0.042	0.031	0.022	0.016	0.014	0.043	0.031	0.021	0.019	0.013	0.047	0.035	0.026	0.018	0.013
$FC_1(h=3)$	0.006	0.004	0.001	0.001	0.000	0.007	0.004	0.002	0.001	0.001	0.008	0.005	0.003	0.002	0.003	0.013	0.009	0.007	0.007	0.010
$FC_2(h=1)$	0.121	0.100	0.083	0.065	0.060	0.121	0.113	0.111	0.134	0.176	0.160	0.162	0.213	0.318	0.513	0.267	0.367	0.556	0.803	0.973
$FC_2(h=2)$	0.090	0.077	0.064	0.053	0.045	0.078	0.075	0.062	0.054	0.048	0.087	0.069	0.061	0.055	0.046	0.086	0.072	0.063	0.045	0.029
$FC_2(h=3)$	0.020	0.013	0.010	0.007	0.004	0.020	0.015	0.012	0.008	0.010	0.022	0.016	0.017	0.017	0.022	0.030	0.026	0.034	0.040	0.051
$FC_1^*(h=1)$	0.014	0.006	0.003	0.001	0.000	0.014	0.008	0.005	0.006	0.006	0.020	0.013	0.013	0.021	0.038	0.049	0.052	0.080	0.162	0.348
$FC_1^*(h=2)$	0.012	0.004	0.002	0.001	0.001	0.009	0.006	0.003	0.002	0.001	0.010	0.005	0.004	0.003	0.003	0.016	0.011	0.010	0.010	0.014
$FC_1^*(h=3)$	0.003	0.001	0.000	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.004	0.001	0.001	0.000	0.000	0.003	0.002	0.002	0.001	0.003
$FC_2^*(h=1)$	0.030	0.016	0.010	0.007	0.004	0.030	0.020	0.016	0.021	0.029	0.038	0.032	0.042	0.063	0.121	0.082	0.097	0.162	0.309	0.547
$FC_2^*(h=2)$	0.026	0.013	0.008	0.006	0.004	0.021	0.012	0.010	0.009	0.010	0.025	0.016	0.013	0.014	0.018	0.033	0.028	0.028	0.036	0.048
$FC_2^*(h=3)$	0.009	0.003	0.002	0.001	0.001	0.007	0.003	0.002	0.001	0.002	0.008	0.005	0.003	0.004	0.005	0.011	0.009	0.009	0.014	0.025
$\mathrm{DM}_{90}(h=1)$	0.045	0.040	0.039	0.033	0.035	0.043	0.043	0.048	0.051	0.068	0.052	0.050	0.068	0.093	0.129	0.068	0.096	0.132	0.195	0.285
$\mathrm{DM}_{90}(h=2)$	0.087	0.063	0.047	0.043	0.039	0.086	0.067	0.049	0.042	0.041	0.078	0.058	0.044	0.036	0.029	0.069	0.055	0.033	0.024	0.016
$\mathrm{DM}_{90}(h=3)$	0.102	0.080	0.062	0.054	0.046	0.099	0.084	0.061	0.051	0.048	0.106	0.082	0.063	0.048	0.040	0.097	0.081	0.049	0.039	0.031
$\mathrm{DM}_{95}(h=1)$	0.023	0.020	0.019	0.015	0.017	0.023	0.021	0.025	0.025	0.039	0.028	0.028	0.037	0.053	0.079	0.039	0.053	0.079	0.125	0.194
$\mathrm{DM}_{95}(h=2)$	0.057	0.036	0.025	0.024	0.021	0.055	0.040	0.026	0.022	0.021	0.053	0.036	0.023	0.018	0.014	0.045	0.029	0.016	0.011	0.007
$\mathrm{DM}_{95}(h=3)$	0.073	0.052	0.037	0.028	0.022	0.071	0.054	0.034	0.026	0.026	0.077	0.054	0.037	0.026	0.022	0.073	0.053	0.029	0.020	0.015
$\mathrm{GW}_{90}(h=1)$	0.045	0.044	0.038	0.047	0.038	0.047	0.045	0.040	0.048	0.064	0.058	0.048	0.059	0.075	0.114	0.067	0.079	0.112	0.162	0.234
$\mathrm{GW}_{90}(h=2)$	0.103	0.091	0.084	0.114	0.192	0.099	0.094	0.080	0.110	0.187	0.103	0.089	0.080	0.095	0.171	0.104	0.081	0.063	0.073	0.128
$\mathrm{GW}_{90}(h=3)$	0.115	0.124	0.123	0.148	0.259	0.102	0.119	0.115	0.134	0.236	0.112	0.120	0.124	0.142	0.228	0.105	0.125	0.106	0.120	0.195
$\mathrm{GW}_{95}(h=1)$	0.021	0.020	0.018	0.020	0.018	0.022	0.021	0.015	0.022	0.027	0.033	0.025	0.028	0.038	0.062	0.035	0.041	0.061	0.096	0.152
$\mathrm{GW}_{95}(h=2)$	0.075	0.057	0.045	0.062	0.118	0.075	0.063	0.044	0.060	0.118	0.082	0.062	0.045	0.051	0.109	0.085	0.058	0.032	0.035	0.078
$\mathrm{GW}_{95}(h=3)$	0.093	0.089	0.082	0.089	0.178	0.079	0.090	0.073	0.082	0.162	0.088	0.085	0.080	0.081	0.152	0.085	0.095	0.069	0.069	0.128
$\mathrm{ENC\text{-}T}_{90}$	0.126	0.116	0.118	0.109	0.105	0.130	0.121	0.137	0.149	0.193	0.144	0.158	0.202	0.268	0.392	0.200	0.266	0.388	0.570	0.784
$\mathrm{ENC\text{-}T}_{95}$	0.065	0.064	0.062	0.053	0.056	0.068	0.067	0.075	0.087	0.115	0.081	0.084	0.119	0.170	0.267	0.117	0.167	0.263	0.427	0.663
$\mathrm{ENC\text{-}NEW}_{90}$	0.055	0.044	0.040	0.036	0.036	0.053	0.049	0.056	0.078	0.117	0.068	0.082	0.124	0.204	0.377	0.135	0.207	0.369	0.643	0.900
$\mathrm{ENC\text{-}NEW}_{95}$	0.027	0.019	0.017	0.014	0.013	0.026	0.024	0.024	0.038	0.062	0.035	0.040	0.067	0.117	0.254	0.081	0.124	0.246	0.512	0.834
GC_{90}	0.143	0.130	0.125	0.112	0.113	0.141	0.146	0.158	0.198	0.262	0.180	0.197	0.275	0.411	0.627	0.293	0.421	0.634	0.871	0.988
GC_{95}	0.090	0.074	0.070	0.061	0.061	0.087	0.085	0.092	0.120	0.173	0.119	0.128	0.188	0.296	0.513	0.209	0.313	0.523	0.801	0.975

Table 6: Results of the simulations of DGP_5. The value of the parameter b is indicated in the first row and the length of the series in the second. The leftmost column indicates the test or criteria (with forecasting horizon h between parentheses; if h is not specified, $h = 1$ or it is not applicable). The figures in the remaining places are the frequencies of of selecting $\{0, 1\}$ for each combination of b, length and test/criteria.