Statistica Sinica **15**(2005), 447-476

ASYMPTOTIC SECOND MOMENT PROPERTIES OF OUT-OF-SAMPLE FORECAST ERRORS OF MISSPECIFIED REGARIMA MODELS AND THE OPTIMALITY OF GLS

David F. Findley

U.S. Bureau of the Census

Abstract: Under minimal assumptions, it is established that the sample second moments of the errors of out-of-sample (real time) forecasts of possibly incorrect regARIMA models have asymptotic limits with useful frequency domain formulas. Both OLS and GLS estimates of the mean function are considered. With misspecified regressors, under additional assumptions that do not appear to exclude any regressors of interest, the asymptotic formulas are used to show that GLS has minimal asymptotic mean square error for one-step-ahead forecasting relative to OLS and other alternatives.

Key words and phrases: ARIMA, model selection, real-time forecasting.

1. Introduction

For modeling and forecasting most economic indicators and many other time series, it is appropriate to model a time-varying mean function as well as the autocovariance structure. Suppose that, after any needed data transformations, one has observations $Y_t, 1 \leq t \leq T$ of a time series of the form

$$Y_t = AX_t + y_t, \tag{1}$$

where X_t is a sequence of column vectors and y_t is a process with convergent sample second moments that is asymptotically orthogonal to the sequence X_t in the double sense of (31) and (32) below. With monthly or quarterly economic data for example, components of the regressor sequence X_t might describe holiday effects (Bell and Hillmer (1983)) and trading day effects (Findley, Monsell, Bell, Otto and Chen (1998)). Or X_t could include economic variables helpful for forecasting Y_t. Such series Y_t are candidates for regARMA modeling: the modeler considers a regressor X_t^M that is a subvector of X_t which need not coincide with X_t, and proceeds as though, for some A^M to be estimated, the residual process $y_t^M = Y_t - A^M X_t^M$ has the autocovariance structure of an ARMA (r, s) model, which need not be the case.

For model selection in this situation, Findley (1990, 1991) suggests graphical and other diagnostics that can show whether one of the model choices provides persistently better h-step-ahead forecasts $Y^M_{t+h|t}$ of data Y_{t+h}, $t_0 \leq t \leq T-h$, for some $h \geq 1$ and some $t_0 < T-h$. Dawid (1984), Rissanen (1986), Hjorth (1994) and Section 4.3 of Findley Monsell, Bell, Otto and Chen (1998) emphasize comparisons of "out-of-sample" (real time) forecasts, meaning those obtained when the model coefficients used to calculate $Y^M_{t+h|t}$ are estimated from $Y_s, 1 \leq s \leq t$ for $t_0 \leq t \leq T-h$, because the object of interest for forecasting is generally the future course of the observed series, not the future course of some statistical replicate. The diagnostics for such comparisons that are implemented in the X-12-ARIMA time series modeling and seasonal adjustment program discussed in the last reference often suggest that the accumulating squared errors $\sum_{t=t_0}^{\tau}(Y_{t+h}-Y^M_{t+h|t})^2$ increase roughly linearly in τ, or, more concretely, that the averages of the squared out-of-sample forecast errors converge as τ increases, to a finite, positive limit, even for models that are far from correct.

In this article, under minimal assumptions on y_t, and under practically general assumptions on X_t given in Section 3, we show that, for a realization on which appropriately weighted sample second moment sequences of the series X_t and y_t converge, if X^M_t is a subvector of X_t that includes all unbounded regressors of X_t (e.g., polynomials), then

$$\lim_{T\to\infty} \frac{1}{T-h-t_0+1} \sum_{t=t_0}^{T-h} \left(Y_{t+h} - Y^M_{t+h|t}\right)^2 \tag{2}$$

exists and, moreover, has a frequency domain formula that describes large-sample effects of any misspecification, either of the regressor or of the model for the asymptotic second moments of y_t; see (58) below. In these and in all other limiting formulas of the paper, only a single realization is considered, so the limits are ordinary limits, not probabilistic limits. For simplicity, we use $t_0 = 1$ in sums and T in place of $T-h$ in denominators hereafter.

The analysis required to obtain convergence in (2) is made somewhat delicate by the fact that each $Y^M_{t+h|t}$, and therefore each term of the sum (2), is determined by a different estimate of the model parameters: the estimate that determines $Y^M_{t+h|t}$ is a function of the information set $Y_1, \ldots Y_t$, which increases with increasing t. (When the coefficients of X^M_t are estimated with Generalized Least Squares (GLS), there can be situations in which each term in (2) depends on two model parameter estimates, as we shall explain.) The earliest convergence result for an out-of-sample forecast error second moment like (2) is that of Rissanen (1986) for one-step-ahead (i.e., $h = 1$) forecasts of a zero mean autoregressive process Y_t for whose order a finite upper bound is known. Rissanen uses the convergence with probability one in (2) to justify the choice of the autoregressive model order for which the average on the left in (2) with $t_0 = 1$ is minimized. Our focus on out-of-sample forecast errors was partly stimulated by Rissanen's article. We later became aware of the article Dawid (1984) which promotes out-of-sample forecast error criteria for model selection in a context without model correctness assumptions.

A basic theme of our paper is the inheritance of the asymptotic stationarity property assumed for y_t by the out-of-sample forecast error sequences that arise from regARMA and regARIMA modeling with regressors that have more general asymptotic second moment properties than y_t, as described in Section 3. Inheritance is first obtained for ARMA models in Theorem 4 through an adaptation of a recursive estimation result of Lai and Ying (1991), presented as Proposition 3, and by using some general inheritance results that are, in essence, special cases of results of Findley, Pötscher and Wei (2001, 2004) (hereafter FPW 2001, 2004) for within-sample forecast errors. The extensions to regARMA models and regARIMA models, in Sections 5 and 7 respectively, cover GLS estimation procedures of X-12-ARIMA and other contemporary regARIMA modeling software. Theorem 5 describes the asymptotic bias of these GLS estimates, generalizing the simple Ordinary Least Squares (OLS) result (42) of Subsection 4.1.

The formula for the value of (2) is provided by Theorem 6 using the general formula of Theorem 4. In conjunction with a GLS optimality result for within-sample forecasting taken from Findley (2003), this leads to Theorem 8 which describes an asymptotic optimality property of GLS estimates relative to OLS estimates for out-of-sample one-step-ahead forecasting when X_t^M omits regressors that are asymptotically correlated with it. Results for nonstationary regARIMA models are given in Section 7. Proofs that are not immediate consequences of the discussion preceding the asserted result are given in the Appendix. Some extensions of the results of this article are described in Section 8.

2. ARMA Modeling and Forecasting

2.1. Mode of convergence

As indicated above, our results are presented for a single realization of the time series mainly because forecasts of an observed realization are the focus. Restricting the discussion to a single realization makes it possible to ensure more often that a sequence of ARMA model parameter estimates has a unique asymptotic limit, a requirement of our foundational result Proposition 3. In incorrect model situations, it can happen that several parameter values minimize asymptotic average squared forecast error. Then the parameter estimates converge to the set of such values, not necessarily to a single value. However, for some estimation methods it can be shown that, although on different realizations the parameter estimates can have different limit values, on a given realization the sequence of estimates cannot have multiple cluster points and therefore has a unique limit value, in which case our results apply, see Theorem 2.2.2. and Corollary 2.2.1 of Chen (2002) and Findley (2002). The latter reference and Pötscher (1991) cite the main references on non-unique minimizers.

Because only one realization is considered, our convergence assumptions and results are stated in terms of simple, non-stochastic limits, as in the functional approach to time series analysis of Wiener described in Brillinger (1975, pp.41-43), except that our limits refer only to the infinitely distant future, not to both the infinitely distant past and future. A conventional statistical interpretation of our results can be obtained by assuming that all limits assumed to exist do so with probability one over the probability space of all realizations of the time series, a property known to hold broadly for sample second moments of stochastic data, see Theorem IV.3.6 of Hannan (1970) and Subsection 3.1.1 of FPW (2001).

As in Pötscher (1987), a vector sequence $V_t, t \geq 1$, is said to be *asymptotically stationary* (A.S.) if the limiting lagged second moments $\Gamma_k^V = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T-k} V_{t+k}V_t'$ exist (finitely) for all $k \geq 0$. In this case, negatively lagged scaled sample second moments also converge: for $k > 0$, $\Gamma_{-k}^V = \lim_{T\to\infty} T^{-1} \sum_{t=k+1}^{T} V_{t-k}V_t' = (\Gamma_k^V)'$. The matrix sequence Γ_k^V, $k = 0, \pm 1, \ldots$, called the asymptotic second moment sequence of $V_t, t \geq 1$, is positive semidefinite. Due to this property, there is a nondecreasing, positive semidefinite matrix valued function $G_V(\lambda)$ such that $\Gamma_k^V = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_V(\lambda)$, holds for all k, see Grenander (1954) or Chapter II of Hannan (1970). $G_V(\lambda)$ is called the asymptotic spectral distribution matrix of $V_t, t \geq 1$.

2.2. Autoregressive parameterization of ARMA models

We start by describing the model parameterization that will be used to obtain forecasts of an A.S. sequence, beginning with univariate sequences. After presenting two types of ARMA forecast functions, we obtain formulas for the asymptotic second moments of the out-of-sample forecast errors.

Let v_t denote an A.S. scalar time series with asymptotic second moments γ_k^v and asymptotic spectral distribution $G_v(\lambda)$. Suppose v_t is modeled as an invertible ARMA(r, s) model for some $r, s \geq 0$, with autoregressive polynomial $a(z) = 1 + a_1 z + \cdots + a_r z^r$ and moving average polynomial $c(z) = 1 + c_1 z + \cdots + c_s z^s$ satisfying

$$a(z) \neq 0 \neq c(z),\ |z| \leq 1. \tag{3}$$

We do not require $a_r \neq 0$ or $c_s \neq 0$: r and s are upper bounds on the AR and MA orders. Define $\theta(z) = a(z)/c(z)$ and $\tilde{\theta}(z) = c(z)/a(z) = \theta(z)^{-1}$. The coefficients of the power series expansions $\theta(z) = \sum_{j=0}^{\infty} \theta_j z^j$ and $\tilde{\theta}(z) =$

$\sum_{j=0}^{\infty}\tilde{\theta}_j z^j$ in $\{|z|<1\}$ are the model's autoregressive representation coefficients and moving average representation coefficients, respectively. These coefficients are uniquely defined, even when $a(z)$ and $c(z)$ have common zeros. They are related to one another by a simple recursion formula,

$$\tilde{\theta}_0 = \theta_0 = 1,\ \tilde{\theta}_j = -\sum_{i=0}^{j-1}\tilde{\theta}_i\theta_{j-i},\ (j=1,2,\dots), \tag{4}$$

and its inverse. Because of (3), θ_j and $\tilde{\theta}_j$ decay exponentially to zero and so are absolutely summable. In fact, $\sum_{j=0}^{\infty}(1+\varepsilon)^j|\theta_j|<\infty$ and $\sum_{j=0}^{\infty}(1+\varepsilon)^j|\tilde{\theta}_j|<\infty$ hold for some $\varepsilon>0$.

Let $\Theta^{r,s}$ denote the set of all such autoregressive representation coefficient sequences $\theta=(1,\theta_1,\dots)$. These will be shown to determine the models' forecast functions. For $\theta\in\Theta^{r,s}$, in order that the pair of polynomials $a(z)=1+a_1z+\cdots+a_rz^r$ and $c(z)=1+c_1z+\cdots+c_sz^s$ with the properties (3) and $a(z)/c(z)=\theta(z)$ be unique, it is necessary and sufficient that $\theta\notin\Theta^{r-1,s-1}$, in which case $a(z)$ and $c(z)$ have no zeros in common and $a_r\neq 0$ or $c_s\neq 0$. Thus, the difference set $\Theta^{r,s}_{\max}=\Theta^{r,s}\setminus\Theta^{r-1,s-1}$ describes the set of uniquely identified models in $\Theta^{r,s}$. Note that $\Theta^{r,s}_{\max}=\Theta^{r,s}$ if $r=0$ or $s=0$.

Appendix A of Pötscher (1991) summarizes why $\tilde{\theta}=(1,\tilde{\theta}_1,\dots)$ is a convenient model parameter for ARMA model estimation theory to circumvent problems from pairs $a(z)$, $c(z)$ with a common zero. Convergence of AR and MA coefficient sequences leads to coordinatewise convergence of the associated $\tilde{\theta}$ (and θ). The converse holds when $\theta\in\Theta^{r,s}_{\max}$; see Lemma 2, For invertible models, the θ can play the same role as the $\tilde{\theta}$ because of the coordinatewise continuity of the transformations $\theta\mapsto\tilde{\theta}$ and $\tilde{\theta}\mapsto\theta$, properties that follows from (4) and its inverse; see also Section 3 of Findley, Pötscher and Wei (2004). We say that a θ-model (an ARMA model with $\theta(z)=a(z)/c(z)$) is the *correct model* for the limiting second moment ratios γ_k^v/γ_0^v of the A.S. time series v_t if $G_v(\lambda)$ is differentiable with derivative proportional to $|\theta(e^{i\lambda})|^{-2}$. We *never* need to assume that any θ-model is correct.

2.3. Basic forecast formulas

For any $h\geq 1$ and $\theta\in\Theta^{r,s}$, define $\tilde{\theta}_{h-1}(z)=\sum_{j=0}^{h-1}\tilde{\theta}_jz^j$. Let B denote the backshift operator, $Bv_t=v_{t-1}$. If v_t is a stationary Gaussian ARMA(r,s) process with AR and MA polynomials $a(z)$ and $c(z)$, respectively, such that $\theta(z)=a(z)/c(z)$, then $v_{t+h|t}^{opt}=E\{v_{t+h}|v_u,-\infty<u\leq t\}$, the mean square optimal forecast of v_{t+h} from $v_u,-\infty<u\leq t$, is produced by the filter

$$\pi(h,\theta)(B)=B^{-h}\left(\tilde{\theta}(B)-\tilde{\theta}_{h-1}(B)\right)\theta(B), \tag{5}$$

(see Hannan (1970, p.147)). That is, $v^{opt}_{t+h|t}$ coincides with the θ-model's forecast

$$v^{\infty}_{t+h|t}(\theta) = \pi(h,\theta)(B)\, v_t = \sum_{j=0}^{\infty} \pi_j(h,\theta)\, v_{t-j}. \tag{6}$$

The associated forecast error filter is

$$\eta(h,\theta)(B) = \tilde{\theta}_{h-1}(B)\,\theta(B), \tag{7}$$

i.e., $v_{t+h} - v^{\infty}_{t+h|t}(\theta) = \eta(h,\theta)(B)\, v_{t+h}$. Of course, $\eta(1,\theta)(B) = \theta(B)$.

More generally, for any series v_t, ARMA or not, that has stationary second moments $Ev_{t+k}v_t = \gamma^v_k$, and for every $\theta \in \Theta^{r,s}$, it is easy to check that the infinite sum in (6) converges in mean square and so defines an *infinite-past* forecast, with (generally nonminimal) mean square error given by

$$E\left(v_{t+h} - v^{\infty}_{t+h|t}(\theta)\right)^2 = \int_{-\pi}^{\pi} \left|\eta(h,\theta)\left(e^{i\lambda}\right)\right|^2 dG_v(\lambda), \tag{8}$$

where $G_v(\lambda)$ denotes the spectral distribution function of v_t.

In the situation with which this paper is concerned, instead of assuming that v_t has second moments, it is assumed that v_t is A.S. with asymptotic spectral distribution $G_v(\lambda)$. Then Theorem 4 below shows that the r.h.s. of (8) arises as the *limit* of the sample mean squared errors of forecasts based on finitely many observations $v_1, \ldots, v_t$ as $t \to \infty$. Thus it is the asymptotic average squared forecast error. Because we do not assume that any θ-model is correct, the models most of interest for forecasting are those that minimize this asymptotic quantity for some $h \geq 1$ over the set Θ of models being considered.

2.4. Truncated and exact finite-past forecast functions

Our interest is in forecasts of univariate data of the form (1) using a θ-model for y_t. Usually the estimates of the coefficients of the chosen regression vector (generally a subvector of X_t) are the θ-model's GLS estimates. This will be seen to give rise to forecasts of A.S. vector series $V_t, t \geq 1$, whose coordinates are y_t and rescaled entries of X_t, in which the same θ-model is used to forecast every coordinate series $v_t, t \geq 1$. Two varieties of forecast functions are in common use. We start with the simpler one, which uses zero for all unavailable values in the infinite-past forecast function. For any $\theta \in \Theta^{r,s}$, this truncation of the vector analogue of (6) produces the *truncated* forecast functions,

$$V_{t+h|t}(\theta) = \sum_{j=0}^{t-1} \pi_j(h,\theta)\, V_{t-j}. \tag{9}$$

Results for these predictors will serve as stepping stones to results for the more commonly used (exact) *finite-past* forecast functions defined for any $h \geq 1$ by

$$\hat{V}_{t+h|t}(\theta) = \sum_{j=0}^{t-1} \pi_{t,j}(h,\theta) V_{t-j}, \tag{10}$$

where the coefficient vector $[\pi_{t,j}(h,\theta)]_{0\leq j\leq t-1}$ is the solution of the linear system

$$[\pi_{t,j}(h,\theta)]_{0\leq j\leq t-1} [\rho_{j-k}(\theta)]_{0\leq j,k\leq t-1} = [\rho_{k+h}(\theta)]_{1\leq k\leq t}, \tag{11}$$

with $\rho_k(\theta) = \sum_{j=0}^{\infty} \tilde{\theta}_{j+k}\tilde{\theta}_j / \sum_{j=0}^{\infty} \tilde{\theta}_j^2$, $k \geq 0$; see, for example, Newton and Pagano (1983). We set $V_{t+h|t}(\theta) = \hat{V}_{t+h|t}(\theta) = 0$ for $1-h \leq t \leq 0$. The finite-past forecast error coefficients $\eta_{t,j}(h,\theta), 0 \leq j \leq t-1$, are defined so that $V_t - \hat{V}_{t|t-h}(\theta) = \sum_{j=0}^{t-1} \eta_{t,j}(h,\theta) V_{t-j}$ for $1 \leq t \leq T$.

If a coordinate series v_t of V_t is stationary and Gaussian, and if its first $t+h$ autocorrelations are $\rho_k(\theta), 1 \leq j \leq t+h$, then $E\{v_{t+h}|v_u, 1 \leq u \leq t\} = \sum_{j=0}^{t-1} \pi_{t,j}(h,\theta) v_{t-j}$. However, we do not require the autocorrelations $\rho_k(\theta)$ specified by a θ in $\Theta^{r,s}$ to have any relation to the ratios γ_k^v/γ_0^v of a coordinate series v_t being forecasted. No optimality properties are assumed of the predictors (9) and (10), or of analogously defined predictors, as in (43) and (44) below.

Remark 1. For GLS estimation defined as in Amemiya (1973), and for maximum Gaussian likelihood estimation, the finite-past one-step-ahead forecast errors, or equivalently their coefficients, are usually normalized by dividing the coefficients by the square root of the mean square forecast error quantity $w_{t|t-1}(\theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} |\sum_{j=0}^{t-1} \eta_{t,j}(1,\theta) e^{-ij\lambda}|^2 |\theta(e^{i\lambda})|^{-2} d\lambda$. For the kinds of convergent sequences $\theta^t \to \theta$ we consider below, it follows from the proof of (5.17) of FPW (2004), that $w_{t|t-1}(\theta^t) \to 1$. As a consequence, this normalization has no effect on the asymptotic formulas obtained below, see (b) of Lemma 11. Therefore, for simplicity, we omit the normalization.

2.5. Åström's Recursion Formula for $V_{t+h|t}(\theta)$

For ARMA(r,s) processes, Åström (1970) established a useful recursion for the h-step-ahead infinite-past forecasts (6) that follows from the polynomial division algorithm applied to the ratio $\tilde{\theta}(z) = c(z)/a(z) = \sum_{j=0}^{\infty} \tilde{\theta}_j z^j$ of the s degree MA polynomial $c(z)$ and the r degree AR polynomial $a(z)$. For any $h \geq 1$, this algorithm yields

$$c(z) = \tilde{\theta}_{h-1}(z) a(z) + z^h g_h(z), \tag{12}$$

where

$$g_h(z) = z^{-h} \left\{ \tilde{\theta}(z) - \tilde{\theta}_{h-1}(z) \right\} a(z) \tag{13}$$

is a polynomial of degree at most $q = \max\{r-1, s-h\}$. Because the forecast filter (5) can be expressed as $\pi(h,\theta)(B) = g_h(B)\, c(B)^{-1}$, the series of predictors $v^{\infty}_{t+h|t}(\theta)$ defined by (6) satisfies the difference equation $c(B)\, v^{\infty}_{t+h|t}(\theta) = g_h(B)\, v_t$. This is *Åström's Recursion Formula*. The next proposition, whose proof is in the Appendix, establishes an analogue for the truncated predictors defined by (9),

$$\begin{aligned} &V_{t+h|t}(\theta) + c_1 V_{t+h-1|t-1}(\theta) + \cdots + c_s V_{t+h-s|t-s}(\theta) \\ &= g_{h,0} V_t + g_{h,1} V_{t-1} + \cdots + g_{h,q} V_{t-q}. \end{aligned} \tag{14}$$

For this result, we allow $\theta(z)$ to have zeros on $|z| = 1$. This is convenient because estimates of invertible models need not be invertible; see Anderson and Takemura (1986). Further, the asymptotic limits of AR polynomials of incorrect ARMA models can have zeros of magnitude one (in order to achieve optimal one-step-ahead forecasts within the incorrect model class) according to Pötscher (1987, 1991). We use $\bar{\Theta}^{r,s}$ to denote the superset of $\Theta^{r,s}$ obtained by weakening (3) to

$$a(z) \neq 0 \neq c(z),\ |z| < 1. \tag{15}$$

Proposition 1. *Let $\theta \in \bar{\Theta}^{r,s}$ be given, as well as a vector sequence $V_t, t \geq 1$. For $-q+1 \leq t \leq 0$, define $V_t = 0$. Then, for any $h \geq 1$, the sequence of truncated predictors $V_{t+h|t}(\theta)$, $t \geq 1$ of* (9) *is the solution of* (14) *determined by the initial conditions $V_{t+h|t}(\theta) = 0, -s+1 \leq t \leq 0$.*

The recursion (14) is our key to obtaining results for out-of-sample forecast errors. Before presenting them, we need some basic facts about the convergence of sequences of parameters θ. Given $\theta^t = (1, \theta_1^t, \theta_2^t, \ldots)$, $t \geq 1$, and $\theta = (1, \theta_1, \theta_2, \ldots)$, we write $\theta^t \to \theta$ to indicate coordinatewise convergence: $\lim_{t\to\infty} \theta_j^t = \theta_j$ for all $j \geq 0$. Although we allow some θ^t's to belong to $\bar{\Theta}^{r,s} \setminus \Theta^{r,s}$, meaning $\theta^t(z)$ has a unit magnitude zero or pole, for the reasons indicated above, in order to ensure that AR and MA coefficients also have limits we usually require the limit θ to be in $\bar{\Theta}^{r,s}_{\max} = \bar{\Theta}^{r,s} \setminus \bar{\Theta}^{r-1,s-1}$ or in $\Theta^{r,s}_{\max}$.

For every $\xi \geq 0, \varepsilon \geq 0$, let $\Theta_{\xi,\varepsilon}$ denote the subset of $\bar{\Theta}^{r,s}$ consisting of θ such that $\theta(z) = a(z)/c(z)$, where $a(z) = 0$ only for $|z| \geq 1+\xi$, and $c(z) = 0$ only for $|z| \geq 1+\varepsilon$. Note that $\Theta_{0,0} = \bar{\Theta}^{r,s}$ and, if $\xi > 0$ and $\varepsilon > 0$, that $\Theta_{\xi,\varepsilon} \subseteq \Theta^{r,s}$. These sets provide the properties we require of the convergent parameter sequences θ^t that define the out-of-sample forecasts $V_{t+h|t}(\theta^t)$ and $\hat{V}_{t+h|t}(\theta^t)$. The basic facts are summarized in the next lemma, an elementary result whose proof can be obtained from the arguments of Pötscher (1991, p.447) and the continuity of the transformation $\theta \mapsto \tilde{\theta}$.

Lemma 2. (a) *The sets $\Theta_{\xi,\varepsilon}$ are compact in the sense that every sequence $\theta^t \geq 1$ in $\Theta_{\xi,\varepsilon}$ has a coordinatewise convergent subsequence whose limit θ belongs to $\Theta_{\xi,\varepsilon}$.*

(b) *Suppose the sequence θ^t, $t \geq 1$ in $\bar{\Theta}^{r,s}$ is such that $\theta^t \to \theta \in \bar{\Theta}^{r,s}_{\max}$. Then for polynomials $a_t(z) = 1 + \sum_{j=1}^{r} a_{t,j} z^j$, $a(z) = 1 + \sum_{j=1}^{r} a_j z^j$, $c_t(z) = 1 + \sum_{j=1}^{s} c_{t,j} z^j$, and $c(z) = 1 + \sum_{j=1}^{s} c_j z^j$ such that $\theta^t(z) = a_t(z)/c_t(z)$ and $\theta(z) = a(z)/c(z)$, and for the corresponding polynomials $g_{h,t}(z)$ and $g_h(z)$ defined in accord with* (12)−(13), *we have*

$$\lim_{t\to\infty} a_{t,j} = a_j,\ \lim_{t\to\infty} c_{t,,j} = c_j,\ \lim_{t\to\infty} g_{h,t,j} = g_{h,j}, \tag{16}$$

for $0 \leq j \leq r, s$ and q, respectively. Further, if the zeros of $a(z)$, resp. $c(z)$ belong to $\{|z| > 1\}$, then there is a $\Theta_{\xi,\varepsilon}$ with $\xi > 0$, resp. $\varepsilon > 0$, such that for some $m > 0$, the set $\Theta = \{\theta, \theta^t,\ t \geq m\}$ satisfies

$$\Theta \subseteq \Theta_{\xi,\varepsilon}. \tag{17}$$

If neither $a(z)$ nor $c(z)$ has a unit magnitude zero, then (17) *holds with $\xi > 0$ and $\varepsilon > 0$ for some m.*

The next theorem is a key result that makes it possible to derive properties of out-of-sample forecast errors $V_{t+h} - \hat{V}_{t+h|t}(\theta^t)$ from those of the limit model, $V_{t+h} - \hat{V}_{t+h|t}(\theta)$. Its proof, given in the Appendix, uses (16) and other properties of $\Theta_{\xi,\varepsilon}$ presented in Lemma 10 of the Appendix, together with an adaptation of the proof of Lemma 5 of Lai and Ying (1991), to establish the assertion (19). Then (20) is obtained via a uniform Baxter inequality from FPW (2004).

Proposition 3. *Let $V_t, t \geq 1$ be an A.S. sequence. Suppose a sequence θ^t, $t \geq 1$ in $\bar{\Theta}^{r,s}$ is given such that*

$$\theta^t \to \theta \in \bar{\Theta}^{r,s}_{\max} \cap \Theta_{\xi,\varepsilon} \tag{18}$$

for some $\xi \geq 0$ and $\varepsilon > 0$. Then, for each $h \geq 1$, the truncated predictors $V_{t+h|t}(\theta^t)$ have the sample mean square convergence property

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-h} \left(V_{t+h|t}\left(\theta^t\right) - V_{t+h|t}(\theta)\right)' \left(V_{t+h|t}\left(\theta^t\right) - V_{t+h|t}(\theta)\right) = 0. \tag{19}$$

If ξ is also positive, then the finite-past predictors $\hat{V}_{t+h|t}(\theta^t)$ have the analogous property,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-h} \left(\hat{V}_{t+h|t}\left(\theta^t\right) - \hat{V}_{t+h|t}(\theta)\right)' \left(\hat{V}_{t+h|t}\left(\theta^t\right) - \hat{V}_{t+h|t}(\theta)\right) = 0. \tag{20}$$

Because the differences appearing in (19) and (20) are differences of forecast errors, e.g.,

$$\hat{V}_{t+h|t}\left(\theta^t\right) - \hat{V}_{t+h|t}(\theta) = \left(V_{t+h} - \hat{V}_{t+h|t}(\theta)\right) - \left(V_{t+h} - \hat{V}_{t+h|t}\left(\theta^t\right)\right), \tag{21}$$

it follows from Proposition 2.1 of FPW (2001) (see also Section 5.2.(iii) of this reference) that the asymptotic properties of the sample second moments of the forecast errors $V_{t+h} - \hat{V}_{t+h|t}(\theta^t)$ and $V_{t+h} - V_{t+h|t}(\theta^t)$ coincide with those of $V_{t+h} - V_{t+h|t}(\theta)$, which were established in Theorem 5.2 of FPW (2004). The next theorem is an immediate consequence.

Although (19) and the related results below do not require θ^t to be a function of $V_1, \ldots, V_t$, this is the most natural type of θ^t to occur with forecast errors of the form $V_{t+h} - V_{t+h|t}(\theta^t)$ and their finite-past analogues. Hence, we call all forecast errors of this form out-of-sample forecast errors.

Theorem 4. *Let V_t, $t \geq 1$ be an A.S. sequence with asymptotic spectral distribution function $G_V(\lambda)$, and let θ^t, $t \geq 1$ in $\Theta_{\xi,\varepsilon}$ with $\varepsilon > 0$ converge to a limit $\theta \in \bar{\Theta}^{r,s}_{\max}$. Then given $h, l \geq 1$, for all $k \geq 0$, we have*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} \left(V_{t+k} - V_{t+k|t+k-h}\left(\theta^{t+k}\right)\right)\left(V_t - V_{t|t-l}\left(\theta^t\right)\right)' = \Gamma^V_k(h, l, \theta), \quad (22)$$

where, for any $\theta \in \Theta_{\xi,\varepsilon}$,

$$\Gamma^V_k(h, l, \theta) = \int_{-\pi}^{\pi} e^{-ik\lambda} \eta(h, \theta)\left(e^{i\lambda}\right) \eta(l, \theta)\left(e^{-i\lambda}\right) dG_V(\lambda). \quad (23)$$

If also $\xi > 0$, then the same conclusions hold for the series $V_t - \hat{V}_{t|t-h}(\theta^t)$, $t \geq 1, h \geq 1$.

When the θ^t are maximum likelihood estimates, Theorem 2.1 of Pötscher (1991) shows that their limit θ belongs to $\Theta^{r,s}_{\max}$, both in the correct (not over-parameterized) model situation and in quite general situations in which ARMA (r, s) spectral densities imperfectly approximate a true spectral density that is continuous and positive.

For scalar autoregressive models estimated by least squares, precise rates of convergence are available for the finite-past version of (22) for weakly stationary linear processes satisfying higher moment conditions and other restrictions: see Wei (1992) for the case $h = 1$, where earlier work on rates of convergence going back to Rissanen (1986) is generalized and, in some cases, corrected; see Ing (2004) for the case $h > 1$.

3. Basic Data and Regressor Assumptions and Some Consequences

3.1. Extended asymptotic stationarity

To formulate the data assumptions for X_t and y_t in (1), it is convenient to introduce two generalizations of the concept of asymptotic stationarity. Let U_t, $t \geq 1$ be a real-valued column vector sequence, some of whose entries might be realizations of stochastic variates while others are deterministic, e.g., polynomials, sinusoids, or trading day or holiday effect regressors. Let I_U denote the identity matrix whose order is the dimension $\dim U$ of U_t. The sequence U_t is said to be *scalably asymptotically stationary* (S.A.S.) if there exists a decreasing *scaling* sequence $D_{U,1} \geq D_{U,2} \geq \ldots$ of positive definite diagonal matrices $D_{U,T} = \mathrm{diag}(d_{1,T}^{-1}, \ldots, d_{\dim U,T}^{-1})$ satisfying

$$\lim_{T\to\infty} D_{U,T+k}^{-1} D_{U,T} = I_U \ \ (k = 0, 1, \ldots)\,, \tag{24}$$

and having the property that for all $k \geq 0$,

$$\Gamma_k^U = \lim_{T\to\infty} D_{U,T} \sum_{t=1}^{T-k} U_{t+k} U_t' D_{U,T} \tag{25}$$

exists (finitely). Under (24)–(25), negatively lagged scaled sample second moments also converge: for $k > 0$, $\Gamma_{-k}^U = \lim_{T\to\infty} D_{U,T} \sum_{t=k+1}^{T} U_{t-k} U_t' D_{U,T} = (\Gamma_k^U)'$, and the matrix sequence Γ_k^U, $k = 0, \pm 1, \ldots$ is positive semidefinite, so there is a nondecreasing, positive semidefinite matrix valued function $G_U(\lambda)$ such that $\Gamma_k^U = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_U(\lambda)$. The Γ_k^U are the asymptotic second moment matrices of the sequence U_t and $G_U(\lambda)$ is its asymptotic spectral distribution matrix. As a synonym for the S.A.S. property, we say the entries of U_t are *jointly* S.A.S. This mode of stationarity was introduced for regressors without a formal name in Grenander (1954) to encompass polynomials: if $U_t = t^p$, $p \geq 0$, one can define $D_{U,T} = T^{-(p+1/2)}$ and obtain $\Gamma_k^U = (2p+1)^{-1}$ for all k. Then $G_U(\lambda)$ can be defined to be 0 for $\lambda < 0$ and $(2p+1)^{-1}$ for $\lambda \geq 0$. U_t is A.S. when $D_{U,T} = T^{-1/2} I_U$.

If $U_t, t \geq 1$ is S.A.S. and has the further property that the sequence $\mathring{U}_t = t^{1/2} D_{U,t} U_t$, $t \geq 1$ is A.S., then U_t will be said to be *extendedly asymptotically stationary* (E.A.S.). The asymptotic lag k second moment of $\mathring{U}_t$ will be denoted by $\mathring{\Gamma}_k^U$, and its asymptotic spectral distribution function will be denoted by $\mathring{G}_U(\lambda)$. Of course, if U_t is A.S., it is E.A.S. with $\mathring{\Gamma}_k^U = \Gamma_k^U$ and $\mathring{G}_U(\lambda) = G_U(\lambda)$. If $U_t = t^p$, $p \geq 0$, then $\mathring{U}_t = 1$, $\mathring{\Gamma}_k^U = 1$ and $\mathring{G}_U(\lambda) = (2p+1)\, G_U(\lambda)$.

3.2. The basic assumptions

We assume that the observed series Y_t in (1) results from realizations of y_t and X_t with asymptotic stationarity properties of the sort just defined. Specifically, we require y_t to be A.S., i.e., the limits

$$\gamma_k^y = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} y_{t+k} y_t \tag{26}$$

exist for all $k \geq 0$. Their asymptotic spectral distribution is denoted by $G_y(\lambda)$.

The regressor sequence X_t, $t \geq 1$, in (1) is required to be E.A.S. with

$$D_{X,T} \searrow 0. \tag{27}$$

Partition X_t as

$$X_t = \begin{bmatrix} X_t^M \\ X_t^N \end{bmatrix}, \tag{28}$$

where the superscript N designates the regressors *not* in the model. Let the corresponding partition of A in (1) be $A = \begin{bmatrix} A^M & A^N \end{bmatrix}$, and those of $D_{X,T}$, Γ_k^X and $G_X(\lambda)$ be

$$D_{X,T} = \begin{bmatrix} D_{M,T} & 0 \\ 0 & D_{N,T} \end{bmatrix},$$

$$\Gamma_k^X = \begin{bmatrix} \Gamma_k^{MM} & \Gamma_k^{MN} \\ \Gamma_k^{NM} & \Gamma_k^{NN} \end{bmatrix}, G_X(\lambda) = \begin{bmatrix} G^{MM}(\lambda) & G^{MN}(\lambda) \\ G^{NM}(\lambda) & G^{NN}(\lambda) \end{bmatrix}, \tag{29}$$

respectively. We need Γ_0^{MM} to be positive definite,

$$\Gamma_0^{MM} > 0. \tag{30}$$

Further, we require both X_t and its associated A.S. sequence $\mathring{X}_t = t^{1/2} D_{X,t} X_t$, to be *asymptotically orthogonal* to the series y_t, meaning

$$\lim_{T\to\infty} T^{-\frac{1}{2}} \sum_{t=k+1}^{T-k} y_t X'_{t\pm k} D_{X,T} = 0\ ,\ (k = 0, 1, \ldots), \tag{31}$$

for X_t, and

$$\lim_{T\to\infty} T^{-1} \sum_{t=k+1}^{T-k} y_t \mathring{X}_{t\pm k} = 0\ ,\ (k = 0, 1, \ldots), \tag{32}$$

for $\mathring{X}_t$.

For example, if $y_t = \sum b_j \varepsilon_{t-j}$, where ε_t is an independent white noise process with $\sup_t E\left|\varepsilon_t\right|^r < \infty$, then, for a set of realizations having probability one, (31) holds when $\Gamma_0^X > 0$, and (32) holds when

$$\mathring{\Gamma}_{0,ii}^X > 0, 1 \leq i \leq \dim X_t. \tag{33}$$

It suffices that $r > 2$ if the spectral density of y_t is bounded, or that $r > 4$ if the spectral density is unbounded but square integrable, see Appendix B of FPW (2001). Hereafter, we refer to (26)−(27) and (30)−(32) as the assumptions of Section 3.

Note that (31) yields $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} y_t = 0$ when X_t contains a coordinate that is one for all t because this coordinate's scaling factor in $D_{X,T}$ can be taken to be $T^{-1/2}$. In this sense, y_t in (1) can be thought of as an asymptotically mean zero process.

The data decomposition being modeled is

$$Y_t = A^M X_t^M + y_t^M, \tag{34}$$

where, from (1),

$$y_t^M = A^N X_t^N + y_t. \tag{35}$$

We require X_t^N to be asymptotically stationary, i.e.,

$$D_{N,T} = T^{-1/2} I_N, \tag{36}$$

with I_N being the identity matrix of order $\dim X_t^N$. (Omitted regressor variables of larger order would give rise to model residuals that would become infinite in magnitude as T increases and so would be seen to lack the A.S. property for large enough T.) Hence, the analogue of (29) for $\mathring{X}_t$ is

$$\mathring{\Gamma}_k^X = \begin{bmatrix} \mathring{\Gamma}_k^{MM} & \mathring{\Gamma}_k^{MN} \\ \mathring{\Gamma}_k^{NM} & \Gamma_k^{NN} \end{bmatrix}, \quad \mathring{G}_X(\lambda) = \begin{bmatrix} \mathring{G}_{MM}(\lambda) & \mathring{G}_{MN}(\lambda) \\ \mathring{G}_{NM}(\lambda) & G_{NN}(\lambda) \end{bmatrix}. \tag{37}$$

Under (36), it follows from (31) that y_t^M is A.S.: for each $k \geq 0$,

$$\gamma_k^M = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} y_{t+k}^M y_t^M = A^N \Gamma_k^{NN} A^{N\prime} + \gamma_k^y.$$

In general, y_t^M and X_t^M will not be asymptotically orthogonal, e.g.,

$$\lim_{T\to\infty} T^{-1/2} \sum_{t=1}^{T-k} y_{t+k}^M X_t^{M\prime} D_{M,T} = \lim_{T\to\infty} T^{-1/2} \sum_{t=1}^{T-k} A^N X_{t+k}^N X_t^{M\prime} D_{M,T} = A^N \Gamma_k^{NM} \tag{38}$$

will generally be non-zero for some k and some $A^N \neq 0$ unless the sequences X_t^M and X_t^N are asymptotically orthogonal in the sense that $\Gamma_k^{NM} = 0$ for $k = 0, \pm 1, \ldots$.

Here and below, asymptotic formulas for the correct regressor case, $X_t^M = X_t$, can be obtained by setting $A^N = 0$.

3.3. Examples of E.A.S. regressors

Grenander (1954) and Grenander and Rosenblatt (1984, Chap.7) verify the joint S.A.S. property for regressors whose components X_{it} are polynomials t^p with $p \geq 0$ (scaling sequence $T^{-p-1/2}$), linear combinations of $\cos \lambda t$ and $\sin \lambda t$ with $0 < \lambda \leq \pi$, e.g., periodic functions (scaling sequence $T^{-1/2}$), and products of polynomials t^p and sinusoids (scaling sequence $T^{-p-1/2}$). Their formulas show that (30) holds for vectors X_t whose coordinates are regressors of this type. Because $(t^{1/2}t^{-p-1/2})t^p = 1$, the E.A.S. property for polynomial regressors and their products with sinusoids follows from the A.S. property of the constant regressor and of the sinusoids. Note that $\mathring{\Gamma}_0^X$ will be singular (but (33) will hold) if two coordinates of X_t involving a power of t differ only in the power p, because the corresponding coordinates of $\mathring{X}_t$ will be identical and therefore linearly dependent.

3.4. Vector reformulation of the basic assumptions

Our basic assumptions beyond (30) can be usefully reformulated in vector form as follows. The vector sequence $U_t = \left[y_t\, X_t^{M\prime}\, X_t^{N\prime} \right]'$, $t \geq 1$ is S.A.S. with $D_{U,T} = \mathrm{diag}\left(T^{-1/2}, D_{M,T}, T^{-1/2} I_N \right)$, wherein $D_{M,T} \searrow 0$. Further

$$G_U(\lambda) = \begin{bmatrix} G_y(\lambda) & 0 & 0 \\ 0 & G_{MM}(\lambda) & G_{MN}(\lambda) \\ 0 & G_{NM}(\lambda) & G_{NN}(\lambda) \end{bmatrix}.$$

The sequence

$$\mathring{U}_t = \begin{bmatrix} y_t \\ t^{1/2} D_{M,t} X_t^M \\ X_t^N \end{bmatrix},\ t \geq 1 \tag{39}$$

is A.S. with asymptotic spectral distribution

$$\mathring{G}_U(\lambda) = \begin{bmatrix} G_y(\lambda) & 0 & 0 \\ 0 & \mathring{G}_{MM}(\lambda) & \mathring{G}_{MN}(\lambda) \\ 0 & \mathring{G}_{NM}(\lambda) & G_{NN}(\lambda) \end{bmatrix}.$$

4. Estimation of A^M

4.1. OLS estimation of A^M

For each $t \geq 1$, given $Y_s, 1 \leq s \leq t$, the associated OLS estimator of A^M in (34) is

$$A_t^M = \sum_{s=1}^{t} Y_s X_s^{M\prime} \Big[\sum_{s=1}^{t} X_s^M X_s^{M\prime} \Big]^{-1}, 1 \leq t \leq T. \tag{40}$$

Thus

$$A_t^M - A^M = \sum_{s=1}^{t} y_s X_s^{M\prime} \Big[\sum_{s=1}^{t} X_s^M X_s^{M\prime} \Big]^{-1} + A^N \sum_{s=1}^{t} X_s^N X_s^{M\prime} \Big[\sum_{s=1}^{t} X_s^M X_s^{M\prime} \Big]^{-1}.$$

Hence, setting

$$C^{NM} = \Gamma_0^{NM} \left(\Gamma_0^{MM} \right)^{-1}, \tag{41}$$

(31), (36) and (38) yield the asymptotic bias formula

$$\lim_{t\to\infty}\left(A_t^M - A^M\right)t^{-1/2}D_{M,t}^{-1} = A^N C^{NM}. \tag{42}$$

The r.h.s. of (42) is zero when $\Gamma_0^{NM} = 0$, e.g., when X_t^M and X_t^N are asymptotically orthogonal.

4.2. GLS estimation of A^M

GLS estimates of regression coefficients are usually obtained by applying a transformation to the data that yields uncorrelated data with constant variance, followed by application of the same transformation to the regressors, and then by calculation of the OLS estimate of A^M from these transformed quantities. For data conforming to a given ARMA model, the finite-past one-step-ahead forecast errors are uncorrelated and, with the rescaling described in Remark 1, have constant variance. Therefore, given a candidate ARMA model, its GLS estimates are obtained from its one-step-ahead forecast error formulas. For each $t \geq 1$, given Y_s, $1 \leq s \leq t$, and any ARMA model θ^*, we consider two types of GLS estimators of A^M, one based on the truncated past forecasts,

$$A_t^M(\theta^*) = \sum_{s=1}^{t}\left(Y_s - Y_{s|s-1}(\theta^*)\right)\left(X_s^M - X_{s|s-1}^M(\theta^*)\right)' \\ \times\left(\sum_{s=1}^{t}\left(X_s^M - X_{s|s-1}^M(\theta^*)\right)\left(X_s^M - X_{s|s-1}^M(\theta^*)\right)'\right)^{-1} \tag{43}$$

as in Pierce (1971), and the second based on the finite-past forecasts,

$$\hat{A}_t^M(\theta^*) = \sum_{s=1}^{t}\left(Y_s - \hat{Y}_{s|s-1}(\theta^*)\right)\left(X_s^M - \hat{X}_{s|s-1}^M(\theta^*)\right)' \\ \times\left(\sum_{s=1}^{t}\left(X_s^M - \hat{X}_{s|s-1}^M(\theta^*)\right)\left(X_s^M - \hat{X}_{s|s-1}^M(\theta^*)\right)'\right)^{-1}, \tag{44}$$

as in Amemiya (1973). Note that both reduce to the OLS estimator (40) when θ^* is the parameter for white noise, i.e., $\theta_j^* = 0,\ j \geq 1$. In these formulas and elsewhere, a generalized inverse is to be understood whenever the inverse matrix fails to exist. When $\theta^* \in \Theta_{\xi,\varepsilon}$ with $\xi, \varepsilon > 0$, then (30) insures that the inverses exist for all sufficiently large t, by virtue of Theorem 4 with $h = 1$, $\theta^t = \theta^*$, $V_t = X_t^M$, and (45) below.

4.3. Limiting properties of GLS estimates

Partition $\Gamma_k^X(\theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda}|\theta^*(e^{i\lambda})|^2 dG_X(\lambda)$ analogously to (29) as

$$\Gamma_k^X(\theta^*) = \begin{bmatrix}\Gamma_k^{MM}(\theta^*) & \Gamma_k^{MN}(\theta^*)\\ \Gamma_k^{NM}(\theta^*) & \Gamma_k^{NN}(\theta^*)\end{bmatrix},$$

with $\Gamma_k^{MM}(\theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda} \left|\theta^*\left(e^{i\lambda}\right)\right|^2 dG_{MM}(\lambda)$, etc. For any $\theta^* \in \Theta_{\xi,\varepsilon}$ with $\xi > 0$, we have

$$\Gamma_0^{MM}(\theta^*) \geq \min_{|\lambda|\leq\pi} \int_{-\pi}^{\pi} \left|\theta^*\left(e^{i\lambda}\right)\right|^2 \Gamma_0^{MM} \geq m(\xi)\,\Gamma_0^{MM} > 0 \tag{45}$$

with $m(\xi) = 2^{-s}(1+\xi)^{-r}\xi^r$ (see (78) below), so we can define

$$C^{NM}(\theta^*) = \Gamma_0^{NM}(\theta^*)\,\Gamma_0^{MM}(\theta^*)^{-1}. \tag{46}$$

The following results are an immediate consequence of (45) and of Proposition 11.1 and (the proof of) Theorem 5.2 of Findley (2003).

Theorem 5. *Let a convergent sequence $\theta^{*,t}$ in some $\Theta_{\xi,\varepsilon}$ with $\xi, \varepsilon > 0$ be given, with $\theta^{*,t} \to \theta^*$. Under the assumptions of Section* 3, *we have*

$$\lim_{t\to\infty} \left(A_t^M\left(\theta^{*,t}\right) - A^M\right) t^{-1/2} D_{M,t}^{-1} = A^N C^{NM}(\theta^*). \tag{47}$$

Likewise $(\hat{A}_t^M(\theta^{*,t}) - A^M)t^{-1/2}D_{M,t}^{-1}$ converges to $A^N C^{NM}(\theta^*)$. When X_t^M and X_t^N are asymptotically orthogonal, $A^N C^{NM}(\theta^*) = 0$.

Remark 2. As the OLS estimation case with $\theta^* = (1, 0, 0, \ldots)$ shows, the ARMA model used or estimated to obtain the estimates of A^M and the associated h-step-ahead out-of-sample forecasts $\hat{A}_t^M(\theta^*)X_{t+h}^M$ of the mean function values AX_{t+h} can differ from the ARMA model used for the disturbance series y_t^M in (35), see (48) below. Another such situation occurs when $h > 1$ and the ARMA model or parameter estimate used to forecast the regression residual series $Y_t - \hat{A}_t^M(\theta^{*,t})X_{t+h}^M$ is chosen to minimize average squared h-step-ahead forecast errors, whereas the model or parameter estimate used to obtain $\hat{A}_t^M(\theta^{*,t})$ is chosen to be optimal for one-step-ahead forecasting or likelihood maximization. For these reasons, hereafter we use asterisked symbols, θ^* or $\theta^{*,t}$, to designate models used for regression coefficient estimation and symbols without asterisk, θ or θ^t, to designate models for y_t^M.

5. Asymptotic Stationarity of Forecast Errors from a Misspecified regARMA Model with GLS Estimates

5.1. Forecasting with GLS estimates of A^M

We analyze errors of forecasts obtained with GLS estimates of A^M. To cover the situations of Remark 2 in addition to usual case of maximum likelihood estimation of all parameters, we allow the ARMA parameter estimates used for GLS estimates to be different from those used to estimate the covariance structure of the regression residuals.

For $h \geq 1$, $1-h \leq t \leq T$ and any $\theta, \theta^* \in \Theta^{r,s}$, consider the forecast functions

$$Y^M_{t+h|t}(\theta,\theta^*) = A^M_t(\theta^*)\, X^M_{t+h} + y^M_{t+h|t}(\theta,\theta^*), \tag{48}$$

$$y^M_{t+h|t}(\theta,\theta^*) = \begin{cases} \sum_{j=0}^{t-1} \pi_j(h,\theta)\left(Y_{t-j} - A^M_t(\theta^*)\, X^M_{t-j}\right), & 1 \leq t \leq T \\ 0, & 1-h \leq t \leq 0. \end{cases}$$

The forecast errors have the decomposition

$$\begin{aligned} Y_{t+h} - Y^M_{t+h|t}(\theta,\theta^*) = &\left\{y_{t+h} - y_{t+h|t}(\theta)\right\} + \left(A^M - A^M_t(\theta^*)\right)\left\{X^M_{t+h} - X^M_{t+h|t}(\theta)\right\} \\ &+ A^N\left\{X^N_{t+h} - X^N_{t+h|t}(\theta)\right\}. \end{aligned} \tag{49}$$

For finite-past forecasting using $\hat{A}^M_t(\theta^*)$ rather than $A^M_t(\theta^*)$, we define $\hat{Y}^M_{t+h|t}(\theta,\theta^*) = \hat{A}^M_t(\theta^*)\, X^M_{t+h} + \hat{y}^M_{t+h|t}(\theta,\theta^*)$ with

$$\hat{y}^M_{t+h|t}(\theta,\theta^*) = \begin{cases} \sum_{j=0}^{t-1} \pi_{t,j}(h,\theta)\left(Y_t - \hat{A}^M_t(\theta^*)\, X^M_{t-j}\right), & 1 \leq t \leq T, \\ 0, & 1-h \leq t \leq 0. \end{cases}$$

The forecast errors have a decomposition analogous to (49).

Let $\theta^{*,t}$, $t \geq 1$, and θ^t, $t \geq 1$, be convergent sequences contained in some set $\Theta_{\xi,\varepsilon}$ with $\xi, \varepsilon > 0$, having limits θ^* and θ respectively, with $\theta \in \Theta^{r,s}_{\max}$. Then with

$$\begin{aligned} \beta_t(\theta^*) &= \left[1 \left(A^M - A^M_t(\theta^*)\right) t^{-1/2} D^{-1}_{M,t}\, A^N\right], \\ \hat{\beta}_t(\theta^*) &= \left[1 \left(A^M - \hat{A}^M_t(\theta^*)\right) t^{-1/2} D^{-1}_{M,t}\, A^N\right], \end{aligned} \tag{50}$$

and, with $\mathring{U}_t$ as in (39), the observable forecast errors are given by

$$\begin{aligned} Y_t - Y^M_{t|t-h}\left(\theta^t, \theta^{t,*}\right) &= \beta_{t-h}\left(\theta^{t,*}\right)\left\{\mathring{U}_t - \mathring{U}_{t|t-h}\left(\theta^t\right)\right\}, \quad 1 \leq t \leq T, \\ Y_t - \hat{Y}^M_{t|t-h}\left(\theta^t, \theta^{t,*}\right) &= \hat{\beta}_{t-h}\left(\theta^{t,*}\right)\left\{\mathring{U}_t - \hat{\mathring{U}}_{t|t-h}\left(\theta^t\right)\right\}, \quad 1 \leq t \leq T. \end{aligned} \tag{51}$$

Define $\beta(\theta^*) = \left[1 \; -A^N C^{NM}(\theta^*)\, A^N\right]$. Under the respective assumptions of Theorem 5, we have

$$\lim_{t\to\infty} \beta_t\left(\theta^{t,*}\right) = \lim_{t\to\infty} \hat{\beta}_t\left(\theta^{t,*}\right) = \beta(\theta^*). \tag{52}$$

From these observations, Theorem 4 and Lemma 11 of the Appendix immediately yield the following theorem, in preparation for which we define

$$B^{NM}(\theta^*) = A^N\left[-C^{NM}(\theta^*)\;\; I_N\right], \tag{53}$$

$$\mathring{G}_{M,\theta^*}(\lambda) = G_y(\lambda) + B^{NM}(\theta^*)\,\mathring{G}_X(\lambda)\,B^{NM}(\theta^*)', \tag{54}$$

and, for any $\theta, \theta^* \in \Theta^{r,s}$,

$$\mathring{\Gamma}_k^M(h,l,\theta,\theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda}\,\eta(h,\theta)\left(e^{i\lambda}\right)\eta(l,\theta)\left(e^{-i\lambda}\right)d\mathring{G}_{M,\theta^*}(\lambda). \tag{55}$$

Theorem 6. *Let convergent sequences $\theta^{*,t}$, θ^t contained in some $\Theta_{\xi,\varepsilon}$ with $\xi,\varepsilon > 0$ be given, with $\theta^{*,t} \to \theta^*$, $\theta^t \to \theta$ and $\theta \in \Theta^{r,s}_{\max}$. Under the Assumptions of Section 3, the out-of-sample forecast errors $Y_t - Y^M_{t|t-h}(\theta^t, \theta^{*,t})$, $t \geq 1$, $h \geq 1$ are jointly A.S. in the sense that, for any $h, l \geq 1$ and $k \geq 0$,*

$$\frac{1}{T}\sum_{t=1}^{T-k}\left(Y_{t+k} - Y^M_{t+k|t+k-h}\left(\theta^{t+k},\theta^{*,t+k}\right)\right)\left(Y_t - Y^M_{t|t-l}\left(\theta^t,\theta^{*,t}\right)\right) \to \mathring{\Gamma}_k^M(h,l,\theta,\theta^*) \tag{56}$$

as $T \to \infty$. Likewise the series $Y_t - \hat{Y}^M_{t|t-h}(\theta^t,\theta^{*,t})$, $t \geq 1$, $h \geq 1$ are jointly A.S. with the same asymptotic second moments as $Y_t - Y^M_{t|t-h}(\theta^t,\theta^{*,t})$, $t \geq 1$, $h \geq 1$.

This result reveals that the asymptotic sample second moments of the out-of-sample forecast errors considered are the same as those of the θ-model forecast errors of the A.S. series

$$y_t^M(\theta^*) = y_t + A^N\left(X_t^N - C^{NM}(\theta^*)\,\mathring{X}_t^M\right),\ 1 \leq t \leq T, \tag{57}$$

since $\mathring{G}_{M,\theta^*}(\lambda)$ is the spectral distribution function of this series. For the mean squared forecast error case $k = 0$ and $l = h$ in (56), with $\sigma_{hh}(\theta) = \int_{-\pi}^{\pi}|\eta(h,\theta)(e^{i\lambda})|^2 dG_y(\lambda)$, (54) yields

$$\mathring{\Gamma}_0^M(h,h,\theta,\theta^*) = \sigma_{hh}(\theta) + B^{NM}(\theta^*)\left[\int_{-\pi}^{\pi}\left|\eta(h,\theta)\left(e^{i\lambda}\right)\right|^2 d\mathring{G}_X(\lambda)\right]B^{NM}(\theta^*)'. \tag{58}$$

By specializing the argument used to establish Theorem 6, $\sigma_{hh}(\theta)$ is seen to be the asymptotic average squared error of the h-step-ahead forecast of Y_t when X_t is known. Similarly, using Theorem 5 and Lemma 11, the second quantity on the right in (58) can be shown to be the limit of the average of the squares of the h-step-ahead forecast errors of $AX_t - A_t^M(\theta^*)\,X_t^M = \{(A^M - A_t^M)t^{-1/2}D^{-1}_{M,t}\}\mathring{X}_t^M + A^N X_t^N$, $1 \leq t \leq T$.

Remark 3. Note that $\mathring{\Gamma}_k^M(h,l,\theta,\theta^*)$ of (55) is, in general, different from the function obtained by replacing X_t with $\mathring{X}_t$ in the quantities defining

$$\Gamma_k^M(h,l,\theta,\theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda}\eta(h,\theta)\left(e^{i\lambda}\right)\eta(l,\theta)\left(e^{-i\lambda}\right)dG_{M,\theta^*}(\lambda), \tag{59}$$

where $G_{M,\theta^*}(\lambda) = G_y(\lambda) + B^{NM}(\theta^*) G_X(\lambda) B^{NM}(\theta^*)'$. Such replacement would yield $\mathring{C}^{NM}(\theta^*) = \mathring{\Gamma}_0^{NM}(\theta^*) \mathring{\Gamma}_0^{MM}(\theta^*)^{-1}$ rather than $C^{NM}(\theta^*)$ in (53)−(55).

6. Asymptotic Orthogonality and Optimality of GLS

In Findley (2003), an optimality property of GLS relative to other regression estimates for one-step-ahead within-sample forecasting was derived from properties of the special case $\Gamma_0^M(1,1,\theta,\theta^*)$ of (59). The corresponding property for out-of-sample forecast errors will follow automatically when $\mathring{\Gamma}_0^M(1,1,\theta,\theta^*) = \Gamma_0^M(1,1,\theta,\theta^*)$ for all θ, θ^*, see Theorem 8 below. The following result obtains a more general property, $\mathring{\Gamma}_k^M(h,l,\theta,\theta^*) = \Gamma_k^M(h,l,\theta,\theta^*)$ for all k, in a situation that will be shown to commonly occur. Its proof is in the Appendix.

Proposition 7. *Suppose* $X_t^M = [X_t^{M1'}\ X_t^{M2'}]'$, *with* X_t^{M2} *A.S. and with* X_t^{M1} *asymptotically orthogonal to both* X_t^{M2} *and* X_t^N *in the sense that* $\Gamma_k^{M1M2} = 0$ *and* $\Gamma_k^{NM2} = 0$ *for all* k. *Then for all* $\theta, \theta^* \in \Theta^{r,s}$,

$$\mathring{\Gamma}_k^M(h,l,\theta,\theta^*) = \Gamma_k^{M2}(h,l,\theta,\theta^*) = \Gamma_k^M(h,l,\theta,\theta^*) \tag{60}$$

holds for all k, *where* $\Gamma_k^{M2}(h,l,\theta,\theta^*)$ *is the analogue of* (59) *for the A.S. regressor* $X2_t = [X_t^{M2'}\ X_t^{N'}]'$.

To illustrate the scope of this result, we consider examples of asymptotic orthogonality.

6.1. Asymptotically orthogonal regressors

Let $G_{ij}(\lambda)$ denote the (i,j)-entry of $G_X(\lambda)$. The regressors defined by the i-th and j-th coordinates of X_t are asymptotically orthogonal if and only if $G_{ij}(\lambda)$ is constant on $[-\pi,\pi]$, or, equivalently, if differences $\Delta G_{ij} = G_{ij}(\lambda'') - G_{ij}(\lambda')$ with $-\pi \le \lambda' < \lambda'' \le \pi$ are zero. Because the positive semidefiniteness of $\Delta G_X(\lambda)$ yields $(\Delta G_{ij})^2 \le \Delta G_{ii} \Delta G_{jj}$, this happens whenever $G_{ii}(\lambda)$ is constant except at a sequence of frequencies λ_k where a jump occurs, $G_{ii}(\lambda_k+) - G_{ii}(\lambda_k-) > 0$, while G_{jj} is continuous at these frequencies, $G_{jj}(\lambda_k+) - G_{jj}(\lambda_k-) = 0$.

Grenander and Rosenblatt (1984, Chap.7) shows that a regressor of the form

$$c_0 + \sum_{k=1}^{H} (c_k \cos\lambda_k t + d_k \sin\lambda_k t) + c_{H+1}(-1)^t, \tag{61}$$

has a spectral distribution function with jumps at each frequency λ_k for which $c_k^2 + d_k^2 \neq 0$, and also at the frequency 0, resp. π, if $c_0 \neq 0$, resp. $c_{H+1} \neq 0$. Elsewhere, its spectral distribution function is constant. It is also shown that the same conclusions apply to a regressor of this form multiplied by a polynomial in t. It follows that two sinusoids (or products thereof with polynomials) are asymptotically orthogonal if and only if they have no common frequency components. Thus, for example, polynomials in t are asymptotically orthogonal to periodic regressors with mean zero ($c_0 = 0$), and deseasonalized regressors of the sort used to model trading day and holiday effects (see Findley, Monsell, Bell, Otto and Chen (1998) and Findley and Soukup (2000)) are asymptotically orthogonal to seasonal regressors. Further, an A.S. or stationary regressor whose asymptotic second moments are determined by a spectral density has an absolutely continuous asymptotic spectral measure. Therefore, such a regressor is asymptotically orthogonal to polynomials and to regressors of the form (61), and also to products of such regressors. More general examples, with $H = \infty$ in (61) for instance, can be extracted from Chap. 7 of Grenander and Rosenblatt (1984).

Now we consider the usual case in which some coordinate $X_{i,t}^M$ of X_t^M is constant, $X_{i,t}^M = 1$ for all t. In this case, with no loss of generality, X_t^N can be required to be asymptotically orthogonal to this constant regressor, i.e., to have $\bar{X}^N = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T X_t^N$ be equal to 0, because the effect of replacing X_t^N by $X_t^N - \bar{X}^N$ is balanced by changing A_i^M to $A_i^M + A^N\bar{X}^N$. In practice, there may be no cost in making the technically stronger requirement that $G_{NN}(\lambda)$ be continuous at $\lambda = 0$, from which it follows that X_t^N is asymptotically orthogonal to any polynomial regressors in X_t^M, because the asymptotic spectral measures of polynomial regressors increase only at $\lambda = 0$; recall Subsection 3.1. Suppose the subvector X_t^{M1} of X_t^M consisting of regressors that are S.A.S. but not A.S. contains only regressors that are positive integer powers of t, and suppose also that the subvector X_t^{M2} of X_t^M of A.S. regressors of X_t^M has been defined in such a way that each *nonconstant* regressor $X_{j,t}^{M2}$ of X_t^{M2} has mean zero, $\lim_{T\to\infty} T^{-1}\sum_{t=1}^T X_{j,t}^{M2} = 0$ (e.g., so that $c_0 = 0$ in nonconstant regressors of the form (61)). Then X_t^{M1} is asymptotically orthogonal to both X_t^{M2} and X_t^N, fulfilling the assumptions of Proposition 7. The formula (60) shows that the S.A.S. regressors X_t^{M1} have no influence on the asymptotic forecast error second moments.

6.2. An optimality property of GLS

When $\mathring{\Gamma}_0^M(1,1,\theta,\theta^*)$ coincides with $\Gamma_0^M(1,1,\theta,\theta^*)$, under (60) for instance, the properties of the latter function established in Corollary 6.3 of Findley (2003) immediately yield an optimality property of GLS for out-of-sample forecasting.

Theorem 8. *Suppose* $\mathring{\Gamma}_0^M(1,1,\theta,\theta^*) = \Gamma_0^M(1,1,\theta,\theta^*)$ *holds for all* θ, θ^* *in some* $\Theta_{\xi,\varepsilon}$ *with* $\xi,\varepsilon > 0$. *If* $\bar{\theta}$ *denotes a minimizer of* $\mathring{\Gamma}_0^M(1,1,\theta,\theta)$ *on a compact subset* Θ *of* $\Theta_{\xi,\varepsilon}$, *we have*

$$\mathring{\Gamma}_0^M(1,1,\bar{\theta},\bar{\theta}) = \min_{\theta,\theta^*\in\Theta} \mathring{\Gamma}_0^M(1,1,\theta,\theta^*). \tag{62}$$

More precisely, for a fixed $\theta^* \in \Theta$*, let* $\bar{\theta}^*$ *denote a minimizer of* $\mathring{\Gamma}_0^M(1,1,\theta,\theta^*)$ *over* Θ*. Then*

$$\mathring{\Gamma}_0^M\left(1,1,\bar{\theta},\bar{\theta}\right) \leq \mathring{\Gamma}_0^M\left(1,1,\bar{\theta}^*,\theta^*\right), \tag{63}$$

with strict inequality holding for some values of A^N *if and only if either* $\mathring{\Gamma}_0^M(1,1,\bar{\theta},\bar{\theta}) < \mathring{\Gamma}_0^M\left(1,1,\bar{\theta}^*,\bar{\theta}^*\right)$ *or*

$$C^{NM}\left(\theta^*\right) \neq C^{NM}\left(\bar{\theta}^*\right) \tag{64}$$

holds, for $C^{NM}(\theta^*)$ *as in* (46)*. When strict inequality obtains in* (63)*, then*

$$C^{NM}\left(\theta^*\right) \neq C^{NM}\left(\bar{\theta}\right). \tag{65}$$

The case of (63) of general interest concerns $\theta^* = (1,0,0,\ldots)$ when, in (48), $A_t^M(\theta^*) = A_t^M$, the OLS estimator. Then (62) yields an out-of-sample analogue of the optimality property of GLS estimation relative to OLS for one-step-ahead within-sample forecasting discussed in Findley (2003). In this reference, it is also shown that Gaussian likelihood maximization leads to estimates θ^t that converge to the set of minimizers of $\Gamma_0^M(1,1,\theta,\theta) = \mathring{\Gamma}_0^M(1,1,\theta,\theta)$. Combining these results, we have the following optimality property of GLS: for model families as in Theorem 8, and realizations yielding parameter estimates that converge to a minimizer of $\mathring{\Gamma}_0^M(1,1,\theta,\theta)$, *OLS estimation is never better than GLS estimation for one-step-ahead out-of-sample forecasting asymptotically. When* X_t^M *is misspecified and not asymptotically orthogonal to* X_t^N*, OLS is typically worse.*

The superiority of GLS over OLS can be observed with time series of typical lengths. Among an unsystematically chosen set of eight monthly U.S. Imports series of length at most 156 months modeled by seasonal ARIMA models with trading day effect regressors as described in Findley, Monsell, Bell, Otto and Chen (1998), Kellie Wills determined that the average squared out-of-sample forecast error from the eighty-fifth month to the end of the series was smaller with GLS trading day coefficient estimates for six of the series (for $h = 1, 12$) and smaller with OLS estimates for two of the series (for $h = 1, 12$).

6.3. AR(1) Models: $h = 1$ and $\dim X_t^{M2} = \dim X_t^N = 1$

For the situation of Proposition 7, we now present some illustrative formulas related to the minimum asymptotic average squared forecast errors $\mathring{\Gamma}_0^M\left(1,1,\bar{\theta},\bar{\theta}\right)$ and $\mathring{\Gamma}_0^M\left(1,1,\bar{\theta}^*,\theta^*\right)$ based on (60) for the situation in which $\dim X_t^{M2} = \dim X_t^N = 1$ and a first-order autoregressive model, with $\theta = \theta(\phi) = (1,-\phi,0,0,\ldots)$, is used for the regression error series y_t^M in (34). Thus

$$\Gamma_0^X(\theta) = \int_{-\pi}^{\pi} \left|1-\phi e^{i\lambda}\right|^2 dG_X(\lambda) = \left(1+\phi^2\right)\Gamma_0^X - \phi\left(\Gamma_1^X + \Gamma_{-1}^X\right) \tag{66}$$

and, from (54) and (60),

$$\mathring{\Gamma}_0^M(1,1,\theta,\theta^*) = \int_{-\pi}^{\pi}\left|1-\phi e^{i\lambda}\right|^2 dG_y(\lambda) + B^{NM}(\theta^*)\int_{-\pi}^{\pi}\left|1-\phi e^{i\lambda}\right|^2 dG_{X2}(\lambda)\, B^{NM}(\theta^*)'. \quad (67)$$

Setting $\theta^* = \theta$ and making use of (66), we obtain

$$\mathring{\Gamma}_0^M(1,1,\theta,\theta) = \left(1+\phi^2\right)\left\{\gamma_0^y + \left(A^N\right)^2\Gamma_0^{NN}\right\} - 2\phi\left\{\gamma_1^y + \left(A^N\right)^2\Gamma_1^{NN}\right\} - \left(A^N\right)^2\frac{\left\{(1+\phi^2)\,\Gamma_0^{NM2} - \phi\left(\Gamma_1^{NM2}+\Gamma_{-1}^{NM2}\right)\right\}^2}{(1+\phi^2)\,\Gamma_0^{M2M2} - 2\phi\Gamma_1^{M2M2}}.$$

Thus, $\mathring{\Gamma}_0^M(1,1,\theta,\theta)$ is a rational function of ϕ whose minimizing value $\bar{\phi}$ is a zero of a polynomial in ϕ of degree five in general.

For this reason, to demonstrate strict inequality in (63), we proceed indirectly and show for the OLS choice $\theta^* = (1,0,0,\dots)$, i.e., for $\phi^* = 0$, that very generally (64) holds when one or more of Γ_0^{NM2}, Γ_1^{NM2}, and Γ_{-1}^{NM2} is nonzero. We begin by noting from (66) that

$$C^{NM2}(\theta(\phi)) = \frac{(1+\phi^2)\,\Gamma_0^{NM2} - \phi\left(\Gamma_1^{NM2}+\Gamma_{-1}^{NM2}\right)}{(1+\phi^2)\,\Gamma_0^{M2M2} - 2\phi\Gamma_1^{M2M2}}$$

is strictly monotonic for $-1 < \phi < 1$ quite generally, since

$$\frac{d}{d\phi}C^{NM2}(\theta(\phi)) = \left(1-\phi^2\right)\frac{\left\{2\Gamma_0^{NM2}\Gamma_1^{M2M2} - \left(\Gamma_1^{NM2}+\Gamma_{-1}^{NM2}\right)\Gamma_0^{M2M2}\right\}}{\left\{(1+\phi^2)\,\Gamma_0^{M2M2} - 2\phi\Gamma_1^{M2M2}\right\}^2},$$

is nonzero with constant sign over $(-1,1)$ whenever

$$2\Gamma_0^{NM2}\Gamma_1^{M2M2} - \left(\Gamma_1^{NM2}+\Gamma_{-1}^{NM2}\right)\Gamma_0^{M2M2} \neq 0. \quad (68)$$

Because $\phi^* = 0$, the minimizer $\bar{\theta}^* = \left(1,-\bar{\phi}^*,0,0,\dots\right)$ of $\mathring{\Gamma}_0^M(1,1,\theta,\theta^*)$ is determined by the value $\bar{\phi}^*$ minimizing (67) with $B^{NM2}(\theta^*) = A^N\left[-C^{NM2}\ 1\right]$ and $C^{NM2} = \Gamma_0^{NM2}/\Gamma_0^{M2M2}$, see (41). This is the property of the lag one asymptotic autocorrelation of (57) with $X_t^M = X_t^{M2}$, because of (54). Thus

$$\bar{\phi}^* = \frac{\gamma_1^y + \left(A^N\right)^2\left\{\Gamma_1^{NN} + \left(C^{NM2}\right)^2\Gamma_1^{M2M2} - C^{NM2}\left(\Gamma_1^{NM2}+\Gamma_{-1}^{NM2}\right)\right\}}{\gamma_0^y + (A^N)^2\left\{\Gamma_0^{NN} - (C^{NM2})^2\,\Gamma_0^{M2M2}\right\}}. \quad (69)$$

Except possibly at a single value of $(A^N)^2$, this optimal $\bar{\phi}^*$ will be non-zero, i.e., will be such that $\bar{\theta}^* \neq \theta^*$ when either γ_1^y or $\Delta^{NM2} = \Gamma_1^{NN} + (C^{NM2})^2\Gamma_1^{M2M2} -$

$C^{NM2}(\Gamma_1^{NM2}+\Gamma_{-1}^{NM2})$ is nonzero, which will generally be the case. The property $\theta^* \neq \bar{\theta}^*$ yields (64) under (68), due to (84). For more details and an example in which $\Delta^{NM2} \neq 0$ is verified, see Subsection 7.3 of Findley (2003).

7. Joint Asymptotic Stationarity of Forecast Errors from Misspecified regARIMA Models

Suppose $Y_t = AX_t + y_t$ in (1) arises by applying a "differencing" transformation $\delta(B) = \sum_{j=0}^{d-1} \delta_j B^j$ to nonstationary data W_t of the form

$$W_t = AZ_t + w_t \;\; (t \geq -d+1), \tag{70}$$

with $X_t = \delta(B) Z_t$ and $y_t = \delta(B) w_t$. Then the decomposition $X_t = [X_t^{M\prime}\; X_t^{N\prime}]'$ is obtained from a decomposition $Z_t = [Z_t^{M\prime}\; Z_t^{N\prime}]'$ with $X_t^M = \delta(B) Z_t^M$, etc.

With $\delta_0 = 1$ and $\delta(z) = \sum_{j=0}^{d} \delta_j z^j$, define $\tilde{\delta}_0 = 1$ and $\tilde{\delta}_j = -\sum_{i=0}^{j-1} \tilde{\delta}_i \delta_{j-i}$, $j = 1, 2, \ldots$. Bell (1984, p.650) shows it is not difficult to verify from $Y_t = \delta(B) W_t$ that, for any $h \geq 1$ and $t \geq 0$, there exist coefficients $c_{j,h}$ depending only on $\delta_1, \ldots, \delta_d$ and h, such that

$$W_{t+h} = \sum_{j=0}^{h-1} \tilde{\delta}_j Y_{t+h-j} + \sum_{j=0}^{d-1} c_{j,h} W_{t-j}, \;\; (t \geq 0).$$

Therefore, when $t \geq 0$, given forecasts of Y_{t+h-j}, $0 \leq j \leq h-1$, say $Y_{t+h-j|t}^M(\theta, \theta^*, T)$, $0 \leq j \leq h-1$, defined as in (48), we have a corresponding forecast of W_{t+h},

$$W_{t+h|t}^M(\theta, \theta^*, T) = \sum_{j=0}^{h-1} \tilde{\delta}_j Y_{t+h-j|t}^M(\theta, \theta^*, T) + \sum_{j=0}^{d-1} c_{j,h} W_{t-j}. \tag{71}$$

The observable forecast errors at times $t \geq h$ thus have the decomposition

$$W_t - W_{t|t-h}^M(\theta, \theta^*, T) = \sum_{j=0}^{h-1} \tilde{\delta}_j \left(Y_{t-j} - Y_{t-j|t-h}^M(\theta, \theta^*, T) \right). \tag{72}$$

Consequently, these forecast errors inherit joint and uniform asymptotic stationarity from the corresponding properties of the $Y_t - Y_{t-j|t-h}^M(\theta, \theta^*, T)$. An immediate consequence of Theorem 6 is

Theorem 9. *Let convergent sequences $\theta^{*,t}$, θ^t contained in some $\Theta_{\xi,\varepsilon}$ with $\xi, \varepsilon > 0$ be given, with $\theta^{*,t} \to \theta^*$, $\theta^t \to \theta$ and $\theta \in \Theta_{\max}^{r,s}$. Under the Assumptions of Section 3, the out-of-sample forecast errors $W_t - W_{t|t-h}^M(\theta^t, \theta^{*,t}), t \geq 1, h \geq 1$ are jointly A.S. In particular,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(W_t - W_{t|t-h}^M\left(\theta^t, \theta^{*,t}\right)\right)^2 = \mathring{\Gamma}_0^{M,\delta}(h, h, \theta, \theta^*) \tag{73}$$

with

$$\mathring{\Gamma}_0^{M,\delta}(h,h,\theta,\theta^*) = \int_{-\pi}^{\pi} \Big| \sum_{j=0}^{h-1} e^{ij\lambda} \tilde{\delta}_j \eta(h-j,\theta)\left(e^{i\lambda}\right) \Big|^2 d\mathring{G}_{M,\theta^*}(\lambda).$$

Likewise the forecast error series $W_t - \hat{W}^M_{t|t-h}\left(\theta^t, \theta^{*,t}\right), t \geq 1, h \geq 1$ *are jointly A.S. with the same asymptotic second moments as* $W_t - W^M_{t|t-h}\left(\theta^t, \theta^{*,t}\right), t \geq 1, h \geq 1$.

$\mathring{\Gamma}_0^{M,\delta}(h,h,\theta,\theta^*)$ has a decomposition analogous to (58). For $h = 1$, we have $\mathring{\Gamma}_0^{M,\delta}(1,1,\theta,\theta^*) = \mathring{\Gamma}_0^{M}(1,1,\theta,\theta^*)$, so the optimality results for GLS obtained in Theorem 8 apply in the regARIMA modeling case as well.

Remark 4. Theorems 5, 6 and 9 extend to sets $\Theta_{0,\varepsilon}$ with $\varepsilon > 0$ providing $\Gamma_0^{MM}(\theta^*) > 0$ holds for the limit θ^* because of the continuity of $\Gamma_0^{MM}(\theta)$ on these sets.

8. Extensions

8.1. Tests

We have presented convergence results for sample means of squared forecast errors of regARIMA models under simple, nearly minimal assumptions. A natural next step would seem to be to undertake, under additional assumptions, the derivation of distributional results for statistics that test whether the asymptotic mean squared errors from competing models are the same or not. In fact West (1996), using moment and mixing assumptions about high-level quantities related to the statistics, has derived tests for an interesting variety of scenarios involving out-of-sample forecast errors and multivariate models. However, the applications presented in the technical report Findley (1990) of a similar test statistic, derived (incompletely) for the case of within-sample average squared forecast errors in the report (with some additional details in Findley (1991) and Findley and Wei (1993)), show that it is not straightforward to use such statistics effectively because they are not robust against outliers or nonstationarities in the fourth moments of the data, and such problems have to be identified and dealt with appropriately prior to testing. Rivers and Vuong (2002) provide a rigorous theory for model selection tests like those of Findley (1990). Their assumptions are also high-level, but they verify them for a simple, interesting example.

8.2. Transitory regressors

The widely used intervention regressors of Box and Tiao (1975) converge to 0 exponentially rapidly with increasing t. They are not covered by the results presented above because their natural scaling sequence is the constant 1 for (30), which causes (27) to fail. However, the arguments in Section 9 of Findley (2003) show that the addition to X_t in (1) of such regressors has no impact on the asymptotic second moments (55).

Acknowledgement

Don Poskitt encouraged us to extend our results for A.S. regressors, summarized in Findley (2001), to cover polynomials. We are grateful to Tucker McElroy and a referee for careful readings of the paper and suggestions. Any views expressed are the author's and not necessarily those of the U.S. Census Bureau.

Appendix: Proofs and Additional Lemmas

A.1. Proof of Proposition 1.

Let $v_t, t \geq 1$ be a coordinate sequence of $V_t, t \geq 1$. For $T \geq 1$, define the polynomial $v^T(z) = \sum_{t=1}^{T} v_t z^t$. Due to (15), $c(z)^{-1} = \sum_{j=1}^{\infty} \tilde{c}_j z^j$ is convergent for $|z| < 1$. From (5) and (12), $c(z)\,\pi(h,\theta)(z)\,v^T(z) = c(z)\left\{g_h(z)\,c(z)^{-1} v^T(z)\right\} = g_h(z)\,v^T(z)$. For $1 \leq t \leq T$, the coefficient of z^t on the l.h.s. is $\sum_{j=0}^{\min\{t-1,s\}} c_j$ $v_{t+h-j|t-j}(\theta)$, and on the r.h.s. it is $\sum_{j=0}^{\min\{t-1,q\}} g_{h,j} v_{t-j}$. Using the Proposition's initializations, we can express the coincidence of these coefficients as $\sum_{j=0}^{s} c_j v_{t+h-j|t-j}(\theta) = \sum_{j=0}^{q} g_{h,j} v_{t-j}$. Because T is arbitrary, this verifies the Proposition.

A.2. Further properties of $\Theta_{\xi,\varepsilon}$

Lemma 10. (a) *When $\varepsilon > 0$, then for any $0 \leq \varepsilon_- < \varepsilon$ and any $\xi \geq 0$,*

$$\sup_{\theta \in \Theta_{\xi,\varepsilon}} \sum_{j=1}^{\infty} (1+\varepsilon_-)^j |\theta_j| < \infty, \tag{74}$$

$$\lim_{J\to\infty} \sup_{\theta \in \Theta_{\xi,\varepsilon}} \sum_{j=J}^{\infty} (1+\varepsilon_-)^j |\theta_j| = 0. \tag{75}$$

Similarly, when $\xi > 0$ then for every $0 \leq \xi_- < \xi$ and any $\varepsilon \geq 0$, the sums $\sum_{j=1}^{\infty} (1+\xi_-)^j \left|\tilde{\theta}_j\right|$ converge uniformly on $\Theta_{\xi,\varepsilon}$.

(b) *For any $\varepsilon \geq 0$, the uniform convergence of $\sum_{j=1}^{\infty} (1+\varepsilon)^j |\theta_j|$ on a subset $\Theta \subseteq \bar{\Theta}^{r,s}$ implies the uniform convergence of the corresponding sums $\sum_{j=1}^{\infty} (1+\varepsilon)^j |\eta_j(h,\theta)|$ on Θ for the forecast error filters (7), for each $h \geq 1$.*

(c) *If Θ is a subset of $\Theta^{r,s}$ on which the sums $\sum_{j=1}^{\infty} (1+\varepsilon)^j |\theta_j|$ and $\sum_{j=1}^{\infty} (1+\varepsilon)^j$ $|\tilde{\theta}_j|$ converge uniformly for some $\varepsilon > 0$, then there exist $t_\Theta \geq 1$ and $K_\Theta > 0$*

such that for all $t \geq t_\Theta$,

$$\sup_{\theta\in\Theta}\sum_{j=0}^{t-1}|\eta_{t,j}(h,\theta)-\eta_j(h,\theta)|=\sup_{\theta\in\Theta}\sum_{j=0}^{t-h-1}|\pi_{t-h,j}(h,\theta)-\pi_j(h,\theta)|\leq K_\Theta\,(1+\varepsilon)^{-t}. \tag{76}$$

Hence, given any A.S. series V_t, $t \geq 1$, *for the forecast functions* $V_{t+h|t}(\theta)$ *and* $\hat{V}_{t+h|t}(\theta)$ *defined as in* (9) *and* (10), *we have*

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-h-1}\sup_{\theta\in\Theta}\left(\hat{V}_{t+h|t}(\theta)-V_{t+h|t}(\theta)\right)'\left(\hat{V}_{t+h|t}(\theta)-V_{t+h|t}(\theta)\right)=0. \tag{77}$$

Proof. We first consider the uniform convergence assertions (a) and (b). For $\theta \in \Theta_{\xi,\varepsilon}$, $\theta(z)$ is a ratio of products of factors of the form $1-\nu z$ with $|\nu| \leq (1+\varepsilon)^{-1}$ or $(1+\xi)^{-1}$. If, say, $|\nu| \leq (1+\varepsilon)^{-1}$, then for $0 \leq \varepsilon_- < \varepsilon$ we have

$$(\varepsilon-\varepsilon_-)(1+\varepsilon)^{-1} \leq |1-\nu z| \leq 2+\varepsilon_- \tag{78}$$

for $|z| \leq 1+\varepsilon_-$. It follows readily that there exists a $K(\varepsilon_-) < \infty$ such that $\sup_{\theta\in\Theta_{\xi,\varepsilon}}|\theta(z)| \leq K(\varepsilon_-)$, $|z| \leq 1+\varepsilon_-$. Applying Cauchy's inequality (Hille (1959, p.202)), for the j-th coordinates of each $\theta\in\Theta$ we obtain $\sup_{\theta\in\Theta_{\xi,\varepsilon}}|\theta_j| \leq K(\varepsilon_-)(1+\varepsilon_-)^{-j}$ $(j\geq 0)$. Hence, given such an ε_-, by choosing ε_0 so that $\varepsilon_- < \varepsilon_0 < \varepsilon$, we obtain

$$\begin{aligned}\sup_{\theta\in\Theta_{\xi,\varepsilon}}\sum_{j=J}^{\infty}(1+\varepsilon_-)^j|\theta_j| &= \sup_{\theta\in\Theta_{\xi,\varepsilon}}\sum_{j=J}^{\infty}\left(\frac{1+\varepsilon_-}{1+\varepsilon_0}\right)^j(1+\varepsilon_0)^j|\theta_j| \\ &\leq K(\varepsilon_0)\sum_{j=J}^{\infty}\left(\frac{1+\varepsilon_-}{1+\varepsilon_0}\right)^j\end{aligned}$$

for all $J \geq 0$, from which (74) and (75) follow. The analogous argument applies to the reciprocal functions $\tilde{\theta}(z)$.

The assumptions of (b) are that $D_J = \sup_{\theta\in\Theta}\sum_{j=J}^{\infty}(1+\varepsilon)^j|\theta_j|$, $J \geq 0$ converges to zero and that $D_0 < \infty$. It follows by induction from (4) that $\sup_{\theta\in\Theta,0\leq j\leq J}(1+\varepsilon)^j\left|\tilde{\theta}_j\right| \leq D_0^J$. From this inequality and (7), for a given $h \geq 1$ and each $j \geq 0$,

$$\begin{aligned}\left|(1+\varepsilon)^j\eta_j(h,\theta)\right| &= \left|(1+\varepsilon)^j\sum_{i=0}^{\min(j,h-1)}\tilde{\theta}_i\theta_{j-i}\right| \\ = \Big|\sum_{i=0}^{\min(j,h-1)}(1+\varepsilon)^i\tilde{\theta}_i(1+\varepsilon)^{j-i}\theta_{j-i}\Big| &\leq D_0^{h-1}\sum_{i=\max(0,j-h+1)}^{j}(1+\varepsilon)^i|\theta_i|.\end{aligned}$$

Therefore, for all $J \geq 0$, $\sup_{\theta\in\Theta} \sum_{j=J}^{\infty} (1+\varepsilon)^j |\eta_j(h,\theta)| \leq hD_0^{h-1} D_{\max(0,J-h+1)}$, from which uniform convergence follows because $\lim_{J\to\infty} D_{\max(0,J-h+1)} = 0$.

Applying (b), we obtain (76) of (c) immediately from the Baxter inequality (3.5) of FPW (2004). From this bound, (77) follows from (c) of Proposition 5.2 of FPW (2004) and (a2) of Theorem 2.1 of FPW (2001).

A.3. Proof of Proposition 3

Whenever $\theta^t \to \theta$, it is clear from (4) that $\tilde{\theta}^t \to \tilde{\theta}$ and hence from (5) that $\pi(h,\theta) = (\pi_0(h,\theta), \pi_1(h,\theta), \ldots)$ satisfies

$$\pi\left(h,\theta^t\right) \to \pi(h,\theta). \tag{79}$$

Consequently, for any fixed $u \geq 1$ and any sequence $V_t, t \geq 1$, the truncated forecasts $V_{u+h|u}(\theta^t) = \sum_{j=0}^{u-1} \pi_j(h,\theta^t) V_{u+h-j}$ have the property

$$\lim_{t\to\infty} V_{u+h|u}\left(\theta^t\right) = V_{u+h|u}(\theta). \tag{80}$$

The analogous result holds for the finite-past forecasts, because the $\rho_k(\theta)$ are continuous and the eigenvalues of the matrix $[\rho_{j-k}(\theta)]_{0\leq j,k\leq t-1}$ in (11) are bounded away from zero due to (74)$-$(75), see Proposition 3.1 and (3.12) of FPW (2004).

We now extend the proof of Lemma 5 of Lai and Ying (1990) to cover our situation in which, unlike the situation they consider, the initial values we use to calculate $v_{t+h|t}(\theta^t)$ via (14), namely $v_{m-1-j+h|m-1-j}(\theta^t)$, $0 \leq j \leq s-1$, with m as in Lemma 2, depend on t. Applying (14) to both $v_{t+h|t}(\theta^t)$ and $v_{t+h|t}(\theta)$ and subtracting, we obtain

$$\begin{aligned} &c_t(B)\left\{v_{t+h|t}\left(\theta^t\right) - v_{t+h|t}(\theta)\right\} \\ &= c_t(B)\, v_{t+h|t}\left(\theta^t\right) - c(B)\, v_{t+h|t}(\theta) - \{c_t(B) - c(B)\}\, v_{t+h|t}(\theta) \\ &= \{g_{h,t}(B) - g_h(B)\}\, v_t - \{c_t(B) - c(B)\}\, v_{t+h|t}(\theta) = \phi_t. \end{aligned} \tag{81}$$

With ε and $m > s$ as in Lemma 2, it follows from the proof of part (i) of Lai and Ying's Lemma that there is a $K > 0$ such that, with $\rho = (1+\varepsilon)^{-1}$, we have, for $t \geq m$,

$$\begin{aligned} &\left|v_{t+h|t}\left(\theta^t\right) - v_{t+h|t}(\theta)\right| \\ &\leq K\Big\{\sum_{i=0}^{t-m} \rho^i |\phi_{t-i}| + \rho^{t-m} \sum_{j=0}^{s-1} \left|v_{m-1-j+h|m-1-j}\left(\theta^t\right) - v_{m-1-j+h|m-1-j}(\theta)\right|\Big\}. \end{aligned} \tag{82}$$

From (80), the sequences $v_{u+h|u}(\theta^t) - v_{u+h|u}(\theta)$, $1 \leq u \leq m-1$, converge to 0, and so are bounded:

$$\sup_{t\geq 1, 1\leq u\leq m-1} \left|v_{u+h|u}\left(\theta^t\right) - v_{u+h|u}(\theta)\right| < \infty. \tag{83}$$

Hence, from (82) and the Cauchy-Schwarz inequality, there is a $K' > 0$ such that for $T \geq m$,

$$\sum_{t=m}^{T}\left(v_{t+h|t}\left(\theta^{t}\right)-v_{t+h|t}\left(\theta\right)\right)^{2} \leq K'\Big\{1+\sum_{t=m}^{T}\phi_{t}^{2}\Big\}.$$

Because of this inequality and (83), in order to verify (19) it suffices to verify $\lim_{T\to\infty} T^{-1}\sum_{t=m}^{T}\phi_t^2 = 0$. From ϕ_t's defining formula (81), this follows from (16), $\lim_{T\to\infty} T^{-1}\sum_{t=m}^{T} v_t^2 = \gamma_0^v < \infty$, and $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\{v_{t+h|t}(\theta)\}^2 = \int_{-\pi}^{\pi}|\pi(h,\theta)(e^{i\lambda})|^2 dG_v(\lambda) < \infty$, the latter being a consequence of $\sum_{j=1}^{\infty}|\pi_j(h,\theta)| < \infty$ and Theorem 2.1 of FPW (2001).

A.4. Proof of Proposition 7

Under the assumptions on X_t^{M1}, Proposition 11.1 of Findley (2003) shows that

$$\begin{aligned} C^{NM}\left(\theta^{*}\right) &= \left[0\;\Gamma_0^{NM2}\left(\theta^{*}\right)\right]\begin{bmatrix}\left[\Gamma_0^{M1M1}\left(\theta^{*}\right)\right]^{-1} & 0\\ 0 & \left[\Gamma_0^{M2M2}\left(\theta^{*}\right)\right]^{-1}\end{bmatrix}\\ &= \left[0\;C^{NM2}\left(\theta^{*}\right)\right], \end{aligned} \tag{84}$$

from which (60) follows easily because $X2_t$ is A.S.

A.5. Lemma 11 and its proof

The following Lemma is used to establish Theorem 6.

Lemma 11. (a) *Let $V_t, t \geq 1$ and $\hat{V}_t, t \geq 1$ denote vector sequences of the same dimension such that, for the scaling matrices $D_{V,T} = \mathrm{diag}(d_{1,T}^{-1},\ldots,d_{\dim V,T}^{-1})$, $T \geq 1$,*

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left(\hat{V}_t - V_t\right)' D_{V,T}^2\left(\hat{V}_t - V_t\right) = 0 \tag{85}$$

holds. Then for every $k \geq 0$, $D_{V,T}\sum_{t=1}^{T-k}\left\{\hat{V}_{t+k}\hat{V}_t' - V_{t+k}V_t'\right\}D_{V,T} \to 0$. Thus, if $V_t, t \geq 1$ is S.A.S. with diagonal scaling matrices $D_{V,T}$, then so is $\hat{V}_t, t \geq 1$, and both sequences have the same asymptotic second moment matrices.

(b) *For any an S.A.S. sequence $U_t, t \geq 1$, whose scaling matrices satisfy $D_{U,T} \searrow 0$, and any convergent sequence of matrices $\beta_t \to \beta$ of order $\dim U$, we have*

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left(\beta_t U_t - \beta U_t\right)' D_{U,T}^2\left(\beta_t U_t - \beta U_t\right) = 0. \tag{86}$$

Hence $\beta_t U_t, t \geq 1$ is S.A.S. with scaling matrices $D_{U,T}$ and asymptotic second moment sequence $\beta\Gamma_k^U\beta', k \geq 0$.

The assertion of (a) is equivalent to the validity of $d_{u,T}^{-1} d_{v,T}^{-1} \sum_{t=1}^{T-k} \{\hat{u}_{t+k}\hat{v}_t - u_{t+k}v_t\} \to 0$ for every combination of coordinate entries u_t and v_t of V_t (the same coordinates for every t), where $\hat{u}_t$ and $\hat{v}_t$ are the corresponding entries of $\hat{V}_t$. This follows from a straightforward modification of the argument on pp. 830–831 of FPW (2001).

For (b), it is enough to consider the case $\dim U_t = 1$, and since $\beta_t U_t - \beta U_t = (\beta_t - \beta) U_t$, we can further assume $\beta = 0$. Note that, for any $1 \leq T_0 < T$,

$$D_{U,T}^2 \sum_{t=1}^{T} \beta_t^2 U_t^2 \leq \left(\sup_{t \leq T_0} \beta_t^2 \right) D_{U,T}^2 \sum_{t=1}^{T_0} U_t^2 + \left(\sup_{t \geq T_0+1} \beta_t^2 \right) D_{U,T}^2 \sum_{t=T_0+1}^{T} U_t^2. \quad (87)$$

Because $\sup_{T \geq 1} D_{U,T}^2 \sum_{t=T_0+1}^{T} U_t^2 \leq \sup_{T \geq 1} D_{U,T}^2 \sum_{t=1}^{T} U_t^2 < \infty$, (since $D_{U,T}^2 \sum_{t=1}^{T} U_t^2$ converges), and $\beta_t^2 \to 0$, a standard argument, see Lemma A.1 of FPW (2001), shows that the r.h.s. of (87) tends to 0, which yields (86). The remaining assertions follow from (a).

References

Amemiya, T. (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica* **41**, 723-732.

Anderson, T. W. and Takemura, A. (1986). Why do noninvertible estimated moving averages occur? *J. Time Series Anal.* **7**, 235-254.

Åström, K. J. (1970). *Introduction to Stochastic Control Theory.* Academic Press, New York.

Bell, W. R. (1984). Signal extraction for nonstationary time series. *Ann. Statist.* **12**, 646-664.

Bell, W. R. and Hillmer, S. C. (1983). Modelling time series with calendar variation. *J. Amer. Statist. Assoc.* **78**, 526-534.

Box, G. E. P. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Assoc.* **70**, 70-79.

Brillinger, D. R. (1975). *Time Series Analysis: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

Chen, H.-F. (2002). *Stochastic Approximation and Its Applications*. Kluwer, Dordrecht.

Dawid, A. P. (1984). Statistical theory: the prequential approach (with discussion). *J. Royal Statist. Soc. Ser. A* **147**, 278-292.

Findley, D. F. (1990). Making difficult model comparisons. Statistical Research Division Research Report RR-90/11, Census Bureau.
http://www.census.gov/srd/papers/pdf/rr90-11.pdf

Findley, D. F. (1991). Convergence of finite multistep predictors from incorrect models and its role in model selection. *Note di Matematica* **XI**, 145-155.

Findley, D. F. (2001). Asymptotic stationarity properties of out-of-sample forecast errors of misspecified RegARIMA models. Proceedings of the Business and Economic Statistics Section of the American Statistical Association, CD-ROM. Also available from
http://www.census.gov/ts/papers/osfeasap.pdf

Findley, D. F. (2002). Convergence of a Robbins-Monro algorithm for recursive parameter estimation with non-monotonic weights and multiple zeros (Revised). Statistical Research Division Research Report Series, Statistics #2001-02, Census Bureau.
http://www.census.gov/srd/papers/pdf/rr2001-02.pdf

Findley, D. F. (2003). Properties of forecast errors and estimates of misspecified RegARIMA and intermediate memory models and the Optimality of GLS for one-step-ahead forecasting. Statistical Research Division Research Report Series, Statistics #2003-02, Census Bureau. http://www.census.gov/srd/papers/pdf/rrs2003-06.pdf

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program (with discussion). *J. Bus. Econ. Statist.* **16**, 127-177.

Findley, D. F., Pötscher, B. M. and Wei, C. Z. (2001). Uniform convergence of sample second moments of time series arrays. *Ann. Statist.* **29**, 815-838.

Findley, D. F., Pötscher, B. M. and Wei, C. Z. (2004). Modeling of time series arrays by multistep prediction or likelihood methods. *J. Econometrics* **118**, 151-187.

Findley, D. F. and R. J. Soukup (2000). Modeling and Model Selection for Moving Holidays. 2000 Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 102-107. American Statistical Association, Alexandria. http://www.census.gov/ts/papers/asa00_eas.pdf

Findley, D. F. and Wei, C.-Z. (1993). Moment bounds for deriving time series CLT's and model selection procedures. *Statist. Sinica* **3**, 453-480.

Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25**, 252-272.

Grenander U. and Rosenblatt, M. (1984). *Time Series.* 2nd Edition. Chelsea, New York.

Hannan, E. J. (1970). *Multiple Time Series.* Wiley, New York.

Hille, E. (1959). *Analytic Function Theory*, Vol. I. Ginn, Boston.

Hjorth, U. (1994). *Computer Intensive Statistical Methods, Validation, Model Selection and Bootstrap.* Chapman and Hall, London.

Ing, C.-K. (2004). Selecting optimal multistep predictor for autoregressive processes of unknown order. *Ann. Statist.* **32**, 693-722.

Lai, T. L. and Ying, Z. (1991). Recursive identification and adaptive prediction in linear stochastic systems. *SIAM J. Contr. Optim.* **29**, 1061-1090.

Newton, H. J. and Pagano, M. (1983). The finite memory prediction of covariance stationary time series. *SIAM J. Sci. Statist. Comput.* **4**, 330-339.

Pierce, D. A. (1971). Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika* **58**, 299-312.

Pötscher, B. M. (1987). Convergence results for maximum likelihood type estimators in multivariable ARMA models. *J. Multivariate Anal.* **21**, 29-52.

Pötscher, B. M. (1991). Noninvertibility and pseudo-maximum likelihood estimation of misspecified ARMA models. *Econometric Theory* **7**, 435-449. Corrections: *Econometric Theory* **10**, 811.

Rivers, B. D. and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. *Econometrics J.* **5**, 1-35.

Rissanen, J. R. (1986). Order estimation by accumulated prediction errors. *J. Appl. Probab.* **23** A, 55-61.

Wei, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20**, 1-42.

West, K. D. (1996). Asymptotic inference about prediction ability. *Econometrica* **64**, 1067-1084.

U.S. Bureau of the Census, Statistical Research Division, Washington, DC 20233-9100, U.S.A.

E-mail: david.f.findley@census.gov

(Received December 2003; accepted October 2004)