Asymptotic Stationarity Properties of Out-of-Sample Forecast Errors of Misspecified RegARIMA Models

David F. Findley, Bureau of the Census

1. Introduction

Suppose we have observations $Y_t, 1 \leq t \leq T$ of a time series of the form

$$Y_t = \alpha \xi_t + y_t, \tag{1}$$

where ξ_t is a sequence of nonstochastic column vectors and y_t is a mean zero process whose autocovariance structure is stationary or asymptotically stationary in a sense to be defined. With monthly or quarterly economic data for example, the regressor sequence ξ_t might describe holiday effects (Bell and Hillmer, 1983) and trading day effects (Findley, Monsell, Bell, Otto, and Chen, 1998) as well as localized effects such as a shift of the level of the series or other intervention effect (Box and Tiao, 1975). Such data are candidates for regARMA modeling. The modeler considers a regressor ξ_t^M that might not coincide with ξ_t, and proceeds as though, for some α^M to be estimated, the residual process $y_t^M = Y_t - \alpha^M \xi_t^M$ obeys a zero-mean, invertible ARMA (r, s) model, although this may not be correct even when the choice of regressor is. For model selection in this situation, Findley (1990, 1991) suggests graphical diagnostics that can show whether one of the model choices provides persistently better h-step-ahead forecasts $Y_{t+h|t}^M$ of data Y_{t+h}, $t_0 \leq t \leq T-h$, for some relevant $h \geq 1$, where t_0 is large enough for parameter estimation. Findley et al. (1998) emphasize comparisons of "out-of-sample" forecasts obtained when the model coefficients used to calculate $Y_{t+h|t}^M$ are estimated from $Y_s, 1 \leq s \leq t$. The diagnostics for such comparisons that are implemented in the X-12-ARIMA time series modeling and seasonal adjustment program discussed in this reference often suggest that the accumulating squared errors $\sum_{t=t_0}^{\tau} \left(Y_{t+h} - Y_{t+h|t}^M\right)^2$ increase roughly linearly in τ, or, more concretely, that the average squared out-of-sample forecast errors converge as $\tau \to \infty$, to a finite limit, even for models that are far from correct. In this article, under very weak assumptions on y_t and practically general assumptions on ξ_t given in Section 2, we establish the convergence with probability one (w.p.1) of

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

to a limiting value that shows the large-sample effects of any misspecification of the regressor or of the asymptotic second moment properties of y_t, see (31) below. Convergence over a set of realizations of the time series having probability one is the natural type of convergence to consider because it is forecasts of the observed realization of Y_t that are of interest. The analysis is made delicate by the fact that each $Y_{t+h|t}^M$, and hence each term of the sum of squared errors, is determined by a different estimate of the model parameters. Our focus on out-of-sample forecast errors was partly stimulated by Rissanen (1986). In later Sections, to simplify the expressions, we use $t_0 = 1$ in sums and T in place of $T-h$ in denominators.

2. Basic Assumptions and Some Consequences

We consider regressor sequences ξ_t, $t \geq 1$ of the form

$$\xi_t = \begin{bmatrix} x_t \\ X_t \end{bmatrix}, \tag{2}$$

where, with $'$ denoting transpose, the vectors x_t define transient effects satisfying

$$\sum_{t=1}^{\infty} t^{1/2} \left(x_t' x_t\right)^{1/2} < \infty, \tag{3}$$

and are such that no individual regressor sequence $(x_{it})_{t \geq 1}$ is a linear combination of the rest, $(x_{jt})_{t \geq 1}$, $j \neq i$, or, equivalently, $\det\left[\sum_{t=1}^{\infty} x_t x_t'\right] > 0$. The vectors X_t are such that

$$\Gamma_k \equiv \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-k} X_{t+k} X_t' \text{ exits for } k \geq 0, \tag{4}$$

with Γ_0 (finite and) positive definite,$\Gamma_0 > 0$.

From (4), it follows that there exists a nondecreasing positive semidefinite matrix valued function $G_X(\lambda), -\pi \leq \lambda \leq \pi$ such that $\Gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} dG_X(\lambda)$, see Hannan (1970, p. 76ff.), where also (4) is verified for periodic regressors and therefore for regressors used to model trading day and holiday effects.

Example 1 For $X_t = (-1)^t$, one has $\Gamma_k = (-1)^k$, and $G_X(\lambda)$ can be defined as

$$G_X(\lambda) = \begin{cases} 0, & -\pi \le \lambda < \pi \\ 1, & \lambda = \pi \end{cases}. \quad (5)$$

The random variables y_t, $t \ge 1$ in (1) should be asymptotically stationary in the sense of Parzen (1962), i.e. the sample lagged second moments converge *w.p.1* to finite limits,

$$\gamma_k \equiv w.p.1\text{-}\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} y_{t+k} y_t, \; k \ge 0. \quad (6)$$

We call γ_k the asymptotic lag k second moment of y_t. Let $G_y(\lambda), -\pi \le \lambda \le \pi$ denote a nondecreasing function such that $\gamma_k = \int_{-\pi} \pi e^{-ik\lambda} dG_y(\lambda)$. $G_y(\lambda)$ is called the asymptotic spectral distribution of y_t. The property (6) follows from a variety of sets of conditions, see Theorem IV.3.6 of Hannan (1970, p. 210) and Section 3 of Findley, Pötscher and Wei (2001) (hereafter FPW (2001)).

Our other requirement is that the series y_t and X_t be asymptotically orthogonal in the sense that, for each $k \ge 0$,

$$w.p.1\text{-}\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} y_{t+k} X_t' = 0. \quad (7)$$

Because $\Gamma_0 > 0$, this condition with $k = 0$ is necessary and sufficient for the strong consistency of the least squares estimates of the coefficients of X_t, see the derivation of (11) below. For weakly stationary y_t, there are a variety of conditions that yield (7), see Hannan (1978) and Section 3 of FPW (2001).

2.1 Some consequences

Because $w.p.1\text{-}\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} y_t^2$ exists and is finite, we have

$$w.p.1\text{-}\lim_{t\to\infty} \frac{y_t}{t^{1/2}} = 0, \quad (8)$$

see Remark 2.1 of FPW (2001), and similarly for $X_t/t^{1/2}$.

Consequently, $\sum_{t=1}^{\infty} y_{t+k} x_t' = \sum_{t=1}^{\infty} \left\{(t+k)^{-1/2} y_{t+k}\right\} \left\{(t+k)^{1/2} x_t'\right\}$ converges with probability one, by (8) and (3). Similarly, if $\|\cdot\|$ denotes a matrix norm, we have ($w.p.1$)

$$\sum_{t=1}^{\infty} \|y_{t+k} x_t'\| < \infty, \; \sum_{t=1}^{\infty} \|x_{t+k} X_t'\| < \infty. \quad (9)$$

A consequence of these properties is the convergence of the ordinary least squares (OLS) estimate of α in (1) from data $Y_t, 1 \le t \le T$ which is defined by

$$\hat{\alpha}_T \equiv \sum_{t=1}^{T} Y_t \xi_t' \left[\sum_{t=1}^{T} \xi_t \xi_t' \right]^{-1}, \quad (10)$$

The limiting value is given by

$$w.p.1\text{-}\lim_{T\to\infty} \hat{\alpha}_T = \alpha + \left[\sum_{t=1}^{\infty} y_t x_t' \left[\sum_{t=1}^{\infty} x_t x_t' \right]^{-1} \; 0 \right] \quad (11)$$

where 0 denotes a zero row vector with the dimension of X_t. So this estimator is strongly consistent for X_t, but asymptotically biased for x_t.

3. ARMA Forecasting of Asymptotically Stationary y_t^*

3.1 ARMA Parameterizations

Let y_t^* be an asymptotically stationary time series with asymptotic second moments γ_k^* and asymptotic spectral distribution $G_{y^*}(\lambda)$. Suppose y_t^* is being modeled as an invertible ARMA(r, s) model for given $r, s \ge 0$ 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

$$a(z) \ne 0 \ne c(z), \; |z| \le 1. \quad (12)$$

Initially, we do not require $a_r \ne 0$ or $c_s \ne 0$. Define $\theta(z) \equiv c(z)/a(z)$ and $\tilde{\theta}(z) \equiv c(z)/a(z) = \theta(z)^{-1}$. The coefficients of the power series expansions $\theta(z) = \sum_{j=0}^{\infty} \theta_j z^j$ and $\tilde{\theta}(z) = \sum_{j=0}^{\infty} \tilde{\theta}_j z^j$ are the model's autoregressive representation coefficients and moving average representation coefficients, respectively. Let $\Theta^{r,s}$ denote the set of all such autoregressive representation coefficient sequences $\theta \equiv (1, \theta_1, \ldots)$. For a given $\theta \in \Theta^{r,s}$, let $a\theta(z) \equiv 1 + a_{\theta,1} z + \cdots + a_{\theta,r} z^r$ and $a\theta(z) \equiv 1 + c_{\theta,1} z + \cdots + c_{\theta,s} z^s$ denote polynomials with no zeros in $\{|z| \le 1\}$ such that $\theta(z) \equiv c\theta(z)/a\theta(z)$. These polynomials are uniquely defined only if $\theta \notin \Theta^{r-1,s-1}$ in which case either $a_{\theta,r} \ne 0$ or $c_{\theta,s} \ne 0$, and $a\theta(z)$ and $c\theta(z)$ have no zeros in common. Thus, the subset $\Theta_{\max}^{r,s} \equiv \Theta^{r,s} \setminus \Theta^{r-1,s-1}$ is the set of uniquely identified models in $\Theta^{r,s}$. Note that $\Theta_{\max}^{r,s} = \Theta^{r,s}$ if $r = 0$ or $s = 0$.

Appendix A of Pötscher (1991) explains in detail why the moving average representation coefficient sequence $\tilde{\theta} \equiv \left(1, \tilde{\theta}_1, \ldots\right)$ is a natural parameter for ARMA models which avoids discontinuities that arise at $a(z)$ and $c(z)$ with common zero the associated pathological behavior of m.l.e.'s of ARMA(r,s). For invertible models, the autoregressive representation coefficients $\theta \equiv (1, \theta_1, \ldots)$ can play same role, because of the coordinatewise continuity of the transformations $\theta \leftrightarrow \tilde{\theta}$, see FPW (2002).

3.2 Basic Forecast Formulas

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

$$\pi(h,\theta)(B) \equiv B^{-h}\left(\tilde{\theta}(B) - \tilde{\theta}_{h-1}(B)\right)\theta(B), \tag{13}$$

(see Hannan 1970, p. 147). That is,

$$y_{t+h|t}^*(\theta) = \pi(h,\theta)(B)\, y_t^* = \sum_{j=0}^{\infty} \pi_j(h,\theta)\, y_{t-j}^*. \tag{14}$$

For this forecast function, the associated forecast error filter is

$$\eta(h,\theta)(B) \equiv \tilde{\theta}_{h-1}(B)\,\theta(B), \tag{15}$$

i.e.

$$y_{t+h}^* - y_{t+h|t}^*(\theta) = \eta(h,\theta)(B)\, y_{t+h}^*$$

More generally, for any series y_t^* with stationary second moments $Ey_{t+k}^* y_t^* = \gamma_k^*$, the sum in (14) converges in mean square and defines the θ-model's h-step ahead forecast of y_{t+h}^*, whose mean square error has the formula

$$\int_{-\pi}^{\pi} \left|\eta(h,\theta)\left(e^{i\lambda}\right)\right|^2 dG_{y^*}(\lambda).$$

Given an asymptotically stationary series y_t^* that is available only for $t \geq 1$ and is not necessarily second-moment stationary, for any $\theta \in \Theta^{r,s}$ we can use a truncated version (14) to define the θ-model's predictors,

$$y_{t+h|t}^*(\theta) \equiv \sum_{j=0}^{t-1} \pi_j(h,\theta)\, y_{t-j}^*. \tag{16}$$

3.3 Åström's Recursion Formula for $y_{t+h|t}^*(\theta)$

For ARMA processes, Åström (1970) established a useful recursion for the h-step-ahead forecasts (14) 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 MA polynomial $c(z)$ and the AR polynomial $a(z)$, which we assume have degrees s and r, respectively. With $\tilde{\theta}_{h-1}(z) = \sum_{j=0}^{h-1} \tilde{\theta}_j z^j$, this algorithm yields

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

where

$$g_h(z) = z^{-h}\left\{\tilde{\theta}(z) - \tilde{\theta}_{h-1}(z)\right\} a(z) \tag{17}$$

is a polynomial of degree at most $q = \max\{r-1, s-h\}$. Because the forecast filter (13) can be expressed as $\pi(h,\theta)(B) = g_h(B)\, c(B)^{-1}$, the predictor sequence $y_{t+h|t}^*(\theta)$ defined by (14) satisfies the difference equation

$$c(B)\, y_{t+h|t}^*(\theta) = g_h(B)\, y_t^*, \tag{18}$$

which is *Åström's Recursion Formula.* Let $\overline{\Theta}^{r,s}$ denote the superset of $\Theta^{r,s}$ obtained by weakening (12) to $a(z) \neq 0 \neq c(z),\ |z| < 1$.

Proposition 1 *Let $\theta \in \overline{\Theta}^{r,s}$ be given and also a sequence $y_t^*, t \geq 1$. Define $y_t^* = 0$ for $-q+1 \leq t \leq 0$. Then the truncated predictors (16) whose coefficient are defined by (13) are the solution of (18) for $t \geq 1$ determined by the initial conditions $y_{t+h|t}^*(\theta) = 0, -s+1 \leq t \leq 0$.*

This recursion property is the key to obtaining results for out-of-sample forecast errors. One can modify the proof of Lemma 5 of Lai and Ying (1991), which is concerned with predictors defined by an Åström Recursion Formula with time-varying but converging coefficients, to establish

Theorem 2 *Let $y_t^*, t \geq 1$ be an asymptotically stationary sequence. Suppose a sequence θ^t, $t \geq 1$ of random variates with values in $\overline{\Theta}^{r,s}$ is given such that $\theta^t \rightarrow_{w.p.1} \theta^\infty \in \Theta_{\max}^{r,s}$. Then, for each $h \geq 1$. the truncated predictors $y_{t+h|t}^*(\theta^t)$ have the property that*

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(y_{t+h|t}^*(\theta^t) - y_{t+h|t}^*(\theta^\infty)\right)^2 \rightarrow_{w.p.1} 0.$$

Now from Proposition 2.1 and Theorem 2.1 of FPW (2001) we can obtain

Theorem 3 *Let θ^t, $t \geq 1$ be a sequence in $\overline{\Theta}^{r,s}$ such that*

$$\theta^t \rightarrow_{w.p.1} \theta^\infty \in \Theta^{r,s}_{\max}. \tag{19}$$

Then the out-of-sample forecast errors $y^*_{t+h} - y^*_{t+h|t}(\theta^t)$, $t \geq 1$ are jointly asymptotically stationarity. In particular

$$w.p.1\text{-}\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-h} \left(y^*_{t+h} - y^*_{t+h|t}\left(\theta^t\right) \right)^2 = \int_{-\pi}^{\pi} \left| \eta\left(h, \theta^\infty\right)\left(e^{i\lambda}\right) \right|^2 dG_{y^*}(\lambda). \tag{20}$$

Example 2. Consider the situation in which a first order autoregressive model is fit to y^*_t. With $\theta = (1, -\phi, 0, 0, \ldots)$, suppose the Yule-Walker estimates of ϕ are used, $\phi^t \equiv \sum_{s=1}^{t-1} y^*_{t+1} y^*_t / \sum_{s=1}^{t} \left(y^*_t\right)^2$, $t \geq 2$. Then $\phi^t \rightarrow_{w.p.1} \rho^*_1 \equiv \gamma^*_1/\gamma^*_0$. From (15) and (20), we obtain

$$w.p.1\text{-}\lim_{t\to\infty} \frac{1}{T-h} \sum_{t=1}^{T-h} \left(y^*_{t+h} - \left(\phi^t\right)^h y^*_t \right)^2$$
$$= \int_{-\pi}^{\pi} \left| 1 - \left(\rho^*_1\right)^h e^{ih\lambda} \right|^2 dG_{y^*}(\lambda)$$
$$= \gamma^*_0 \left\{ \left(1 - \left(\rho^*_h\right)^2\right) + \left(\rho^*_h - \left(\rho^*_1\right)^h\right)^2 \right\}, \tag{21}$$

with $\rho^*_h \equiv \gamma^*_h/\gamma^*_0$, which is the minimizer of $\int_{-\pi}^{\pi} \left|1 - \rho e^{ih\lambda}\right|^2 dG_{y^*}(\lambda)$.

Although (20) and results that follow do not require $\theta^t = \theta^t\left(y^*_1, \ldots, y^*_t\right)$, this is the most natural type of random θ^t to occur with forecast errors of the form $y^*_{t+h} - y^*_{t+h|t}\left(\theta^t\right)$. For simplicity, we shall call all forecast errors of this form out-of-sample forecast errors (OSFE's) For in-sample forecast errors obtained with parameter estimates θ^T, (20) was obtained in a more general setting in FPW (2002), but the case of regressor misspecification was not considered.

4. Estimating the Coefficients of a Misspecified Regressor

Let x_t and X_t be partitioned into modeled and not-modeled vectors,

$$x_t = \begin{bmatrix} x^M \\ x^N_t \end{bmatrix}, \; X_t = \begin{bmatrix} X^M \\ X^N_t \end{bmatrix}, \tag{22}$$

with the corresponding partition of α in (1) being $\alpha = \left[a^M a^N A^M A^N\right]$ and those of Γ_k and $G_X(\lambda)$ being

$$\Gamma_k = \begin{bmatrix} \Gamma^{MM}_k & \Gamma^{MN}_k \\ \Gamma^{NM}_k & \Gamma^{NN}_k \end{bmatrix} \tag{23}$$

and

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

respectively. Then, with $\xi^{M\prime}_t \equiv \left[x^{M\prime}_t X^{M\prime}_t\right]'$ and $\alpha^M \equiv \left[a^M A^M\right]$, we have

$$Y_t = \alpha^M \xi^M_t + y^M_t \tag{24}$$

with

$$y^M_t = a^N x^N_t + A^N X^N_t + y_t. \tag{25}$$

It follows from (6), (4), (7) and (9) that y^M_t is asymptotically stationary with

$$\gamma^M_k = A^N \Gamma^{NN}_k A^{N\prime} + \gamma_k = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_{y^M}(\lambda),$$

with $G_{y^M}(\lambda) \equiv A^N G^{NN}(\lambda) A^{N\prime} + G_y(\lambda)$. We are assuming that $A^N \neq 0$ (otherwise there would be no misspecification). Therefore, in contrast to (7)

$$w.p.1\text{-}\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T-k} y^M_{t+k} X^{M\prime}_t = w.p.1\text{-}\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T-k} A^N X^N_{t+k} X^{M\prime}_t = A^N \Gamma^{NM}_k \tag{26}$$

will generally be non-zero for some k unless the sequences X^N_t and X^N_t are asymptotically orthogonal.

5. Joint Asymptotic Stationarity of OSFE's from a Misspecified regARMA Model.

Suppose we are given $\xi^{M\prime}_t \equiv \left[x^{M\prime}_t \; X^{M\prime}_t\right]$ and data $Y_t, 1 \leq t \leq T$. The OLS estimator of α^M in (25) is

$$\hat{\alpha}^M_T \equiv \sum_{t=1}^{T} Y_t \xi^{M\prime}_t \left[\sum_{t=1}^{T} \xi^M_t \xi^{M\prime}_t \right]^{-1}. \tag{27}$$

With Γ^{NM}_0 and Γ^{MM}_0 as in (23), define $C^{NM} \equiv \Gamma^{NM}_0 \left(\Gamma^{MM}_0\right)^{-1}$. Due to (26), instead of (11), we have

$$w.p.1\text{-}\lim_{T\to\infty} \hat{\alpha}^M_T = \alpha^M \tag{28}$$

$$+\left[\sum_{t=1}^{\infty} y_t^M x_t^{M\prime}\left[\sum_{t=1}^{\infty} x_t^M x_t^{M\prime}\right]^{-1}+C \quad A^N C^{NM}\right]$$

where

$$C \equiv A^N C^{NM} \sum_{t=1}^{\infty} X_t^M x_t^{M\prime}\left(\sum_{t=1}^{\infty} x_t^M x_t^{M\prime}\right)^{-1}.$$

Let θ^t, $t \geq 1$ be random variables with values in $\overline{\Theta}^s$ such that

$$\theta^t \rightarrow_{w.p.1} \theta^{M,\infty} \in \Theta_{\max}^{r,s} \tag{29}$$

holds. We consider the forecast functions

$$Y_{t+h|t}^M\left(\theta^t\right) \equiv \alpha_t^M \xi_{t+h}^M + y_{t+h|t}^M\left(\theta^t\right) \tag{30}$$

with

$$y_{t+h|t}^M\left(\theta^t\right) \equiv \sum_{j=0}^{t-1} \pi_j\left(h, \theta^t\right)\left(Y_t - \alpha_t^M \xi_{t-j}^M\right).$$

Theorem 4 *When (29) holds, the OSFE sequences* $Y_{t+h} - Y_{t+|t}^M\left(\theta^t\right)$, $t \geq 1$ *for* $h = 1, 2, \ldots$ *are jointly asymptotically stationary with*

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(Y_{t+h} - Y_{t+h|t}^M\left(\theta^t\right)\right)^2$$

$$\rightarrow_{w.p.1} \int_{-\pi}^{\pi}\left|\eta\left(h, \theta^{\infty,M}\right)\left(e^{i\lambda}\right)\right| dG_y(\lambda)$$

$$+B^N\left[\int_{-\pi}^{\pi}\left|\eta\left(h, \theta^{\infty,M}\right)\left(e^{i\lambda}\right)\right|^2 dG_X(\lambda)\right] B^{N\prime} \tag{31}$$

where, with I *denoting the identity matrix of order* $\dim X_t^N$,

$$B^N \equiv A^N\left[-C^{NM} \ I\right]. \tag{32}$$

The first expression on the right of (31) can be interpreted as the limiting mean squared error of the forecast of y_{t+h} and the second as that of the forecast of $\alpha\xi_{t+h}$.

Set

$$\sigma_{hh}(\theta) \equiv \int_{-\pi}^{\pi}\left|\eta(h, \theta)\left(e^{i\lambda}\right)\right|^2 dG_y(\lambda), \tag{33}$$

and note that if X_t^M and X_t^N are asymptotically orthogonal, then $C^{NM} = 0$.

Example 3. Consider the fitting of an AR(1) model for the regression error and doing one-step-ahead forecasting ($h = 1$). For a model with the correct regressor X_t, we have, from (21) with $y_t^* = y_t$ that

$$\sigma_{11}\left(\theta^{\infty}\right) = \gamma_0\left(1 - \rho_1^2\right).$$

For a misspecified regressor with $X_t^N = (-1)^t$, (21) with $y_t^* = y_t^M$ and $\gamma_k^M = \left(A^N\right)^2(-1)^k + \gamma_k$ yield

$$\sigma_{11}\left(\theta^{M,\infty}\right) = \int_{-\pi}\pi\left|1 - \rho_1^M e^{i\lambda}\right|^2 dG_y(\lambda)$$

$$= \int_{-\pi}\pi\left|1 - \rho_1 e^{i\lambda} + \left(\rho_1 - \rho_1^M\right) e^{i\lambda}\right|^2 dG_y(\lambda)$$

$$= \sigma_{11}\left(\theta^{\infty}\right) + \gamma_0\left(\rho_1 - \rho_1^M\right)^2,$$

where

$$\rho_1^M = \frac{\gamma_1 - \left(A^N\right)^2}{\gamma_0 + \left(A^N\right)^2}$$

Suppose that the nonconstant entries of X_t^M have been deseasonalized, as in Soukup and Findley (2000). Then $X_t^N = (-1)^t$ is orthogonal to X_t^M. As $\eta\left(1, \theta^{M,\infty}\right)\left(e^{i\lambda}\right) = 1 - \rho_1^M e^{i\lambda}$, the final term on the right in (31) has the value $\left(A^N\right)^2\left(1 + \rho_1^M\right)^2$. Thus, the increase in asymptotic mean square one-step forecast error resulting from the regressor misspecification is

$$\gamma_0\left(\rho_1 - \rho_1^M\right)^2 + \left(A^N\right)^2\left(1 + \rho_1^M\right)^2.$$

We note that if the AR(1) model incorrect in the sense that $\rho_k \neq \rho_1^k$ for some $k > 1$, then models different from the AR(1) could have an asymptotic mean square one-step forecast error smaller than $\sigma_{11}\left(\theta^{\infty}\right)$.

6. Extensions

Results for the truncated predictors $y_{t+h|t}^*(\theta)$ often serve as stepping stones toward results for the more commonly used *finite past* predictors defined by

$$\tilde{y}_{t+h|t}^*(\theta) \equiv \sum_{j=0}^{t-1} \pi_{t,j}(h, \theta)\, y_{t-j}^*,$$

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

$$\begin{aligned} & \left[\pi_{t,j}(h, \theta)\right]_{0 \leq j \leq t-1}\left[\rho_{j-k}(\theta)\right]_{0 \leq j,k \leq t-1} \\ = \ & \left[\rho_{h+k}(\theta)\right]_{1 \leq k \leq t}, \end{aligned}$$

with $\rho_k(\theta) = \sum_{j=0}^{\infty} \theta_{j+k}\theta_j / \sum_{j=0}^{\infty} \theta_j^2$, $k \geq 0$. All of the preceding asymptotic stationarity assertions and limit formulas can be established for the situation in which these predictors are used instead of (16), also to define Generalized Least Squares estimates of α^M to be used in place of OLS estimate used above. This is because under the assumptions of Lemma 2, we have, for $\varepsilon > 0$ and θ^t

$$\sum_{j=0}^{t-1} (1+\varepsilon)^j \sup_{\theta^t \in \theta^{r,s}_{\max}} \left| \pi_{t,j}\left(h, \theta^t\right) - \pi_j\left(h, \theta^t\right) \right| = O_{w.p.1}(1)$$

for every $h \geq 1$, see (5.5) and Theorem 2.1 (a2) of FPW (2002).

It can be shown that a limiting second moment formula analogous to (31) holds if the condition on the regressors (4) is weakened to allow polynomial regressors in X_t^M when the entries of X_t^M and X_t^N regressors are either constant or are periodic with mean zero. In this situation, we require the transitory regressors to satisfy $\sum_{t=1}^{\infty} (1+\varepsilon)^j \left(x_t' x_t\right)^{1/2} < \infty$ for some $\varepsilon > 0$, as the intervention regressors of Box and Tiao (1975) do. For within-sample forecast errors, we need no such restrictions: the analogue of (31) holds for all regressors X_t^M satisfying what Hannan (1970) calls Grenander's conditions, see Findley (2001).

Finally, the ARMA models can be replaced by ARIMA models as in FPW (2002). Details are available in the technical report Findley (2001).

Disclaimer. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than reviews given to official Census Bureau publications. Any opinions expressed are those of the author and may not reflect Census Bureau policy.

REFERENCES

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

Bell, W. R. and Hillmer, S. C. (1983). "Modelling Time Series with Calendar Variation," *Journal of the American Statistical Association,* 526–534.

Box, G. E. P. and Tiao, G. C. (1975). "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association* 70, 70–79.

Findley, D. F. (1990). "Making difficult model comparisons," Statistical Research Division Research Report RR-90/11, Census Bureau.//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 In-Sample and Out-of-Sample Forecast Errors of Misspecified RegARIMA Models," Statistical Research Division Research Reports in Statistics No. 2001-02, Census Bureau. //www.census.gov/srd/papers/pdf/rrs2001-2.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," *Journal of Business and Economic Statistics* 16, 127–177 (with discussion).

Findley, D. F., Pötscher, B. M., Wei, C. Z. (2001). "Uniform Convergence of Sample Second Moments of Time Series Arrays," *The Annals of Statistics* 29, 815–838.

Findley, D. F., Pötscher, B. M., Wei, C. Z. (2002). "Modeling of Time Series Arrays by Multistep Prediction or Likelihood Methods," *Journal of Econometrics* (to appear)

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

Hannan, E. J. (1978). "Rates of Convergence for Time Series Regression," *Advances in Applied Probability* 10, 740–743.

Lai, T. L. and Ying, Z. (1991). "Recursive Identification and Adaptive Prediction in Linear Stochastic Systems," *SIAM Journal of Control and Optimization* 29, 1061–1090.

Parzen. E. (1962). "Spectral Analysis of Asymptotically Stationary Time Series," *Bulletin of the International Statistical Institute,* 39, 87–103.

Pötscher, B. M. (1991). "Noninvertibility and Pseudo-Maximum Likelihood Estimation of Misspecified ARMA Models," *Econometric Theory* 7, 435–449.

Rissanen, J.R. (1986). "Order Estimation by Accumulated Prediction Errors," *Journal of Applied Probability,* 23A, 55–61.