RESEARCH REPORT SERIES
(*Statistics #2003-06)*

Properties of Forecast Errors and Estimates of Misspecified RegARIMA and Intermediate Memory Models and the Optimality of GLS for One-Step-Ahead Forecasting

David F. Findley

Methodology and Standards Directorate
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: September 26, 2003

Properties of Forecast Errors and Estimates of Misspecified RegARIMA and Intermediate Memory Models and the Optimality of GLS for One-Step-Ahead Forecasting.

David F. Findley, Bureau of the Census

July 24, 2003

Abstract

We consider the modeling of time series that have an asymptotically stationary autocovariance structure and a mean function of linear regression form in which the regression vector satisfies a weakened version of Grenander's conditions, one that allows transitory regression variables of the sort used for outlier and intervention effect modeling. Neither the model's regression vector sequence nor its parametric family of invertible, short/intermediate-memory autocovariances need be correct. Convergence of both likelihood maximizing and squared-forecast-error minimizing parameter estimates are established as a consequence of uniform strong laws for sample second moments of forecast errors. Both OLS and GLS estimates of the mean function are considered. We show that GLS has an optimal one-step-ahead forecasting property relative to OLS when the model omits a regression variable of the true mean function that is asymptotically correlated with a modeled regression variable. Some inherent ambiguity in the concept of bias for regression coefficient estimators in this situation is discussed.

1. Introduction

Models for most economic indicator series and many other time series require a time-varying mean function as well as an autocovariance structure specification. Suppose that, after making any needed variance stabilizing transformations (such as taking logarithms and then differencings), one has observations $Y_t, 1 \leq t \leq T$ of a time series of the form

$$Y_t = \alpha\xi_t + y_t, \tag{1.1}$$

where ξ_t is a sequence of column vectors that we shall usually regard as nonstochastic (but see Remark 1), and y_t is a process whose autocovariance structure is only required to be asymptotically stationary in a sense to be defined. With monthly or quarterly economic data for example, the regressor sequence ξ_t might describe moving 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 be able to produce $\alpha\xi_t$ for all t, due to omissions, approximations, over-simplifications, etc., and proceeds as though, for a coefficient vector α^M to be estimated, the residual process $y_t^M = Y_t - \alpha^M \xi_t^M$ has the autocovariance sequence of an autoregressive moving average (ARMA) model, or some alternative parametric model such as the exponential model of (Bloomfield, 1973), although this autocovariance assumption might be incorrect,.

Given a family of covariance stationary models for y_t^M with parameter (or index) set Θ, for each $\theta \in \Theta$, let $(\theta_j)_{j\geq 0}$ denote the θ-model's one-step-ahead linear forecast *error* ("autoregressive representation") coefficient sequence. Thus $\theta_0 = 1$ and $-\sum_{j=1}^{\infty} \theta_j y_{t-j}^M$ is the model's linear forecast of y_t^M from $y_s^M, -\infty < s \leq t-1$. With observations $Y_t, 1 \leq t \leq T$, if we set

$$Y_t\left[\theta\right] = \sum_{j=0}^{t-1} \theta_j Y_{t-j},\ \xi_t^M\left[\theta\right] = \sum_{j=0}^{t-1} \theta_j \xi_{t-j}^M,\ (t \geq 1)\,,$$

then a Generalized Least Squares (GLS) estimate of α^M for the θ-model can be defined by

$$\alpha_T^M\left(\theta\right) = \sum_{t=1}^{T} Y_t\left[\theta\right] \xi_t^M\left[\theta\right]' \left(\sum_{t=1}^{T} \xi_t^M\left[\theta\right] \xi_t^M\left[\theta\right]'\right)^{-1},$$

where $'$ denotes transpose, see Pierce (1975), for example. With $\alpha_T^M\left(\theta\right)\xi_t^M$ providing a candidate model for the mean function, θ can be estimated by conditional or unconditional Gaussian maximum likelihood estimation. For the conditional estimates, with which we start for simplicity, for given $1 \leq t \leq T-1$, one defines the θ-model's forecast of Y_{t+1} from $Y_s, 1 \leq s \leq t$ to be

$$Y_{t+1|t}^M\left(\theta, T\right) = \alpha_T^M\left(\theta\right)\xi_{t+1}^M + \sum_{j=0}^{t-1}\left(-\theta_j\right)\left(Y_{t-j} - \alpha_T^M\left(\theta\right)\xi_{t-j}^M\right),$$

and estimates θ with a minimizer θ^T of

$$\frac{1}{T}\sum_{t=1}^{T-1}\left(Y_{t+1} - Y_{t+1|t}^M\left(\theta, T\right)\right)^2 \tag{1.2}$$

over Θ.

Starting from the assumption that the sample moments $T^{-1}\sum_{t=1}^{T-k} y_{t+k}y_t$ converge for all $k \geq 0$ (almost surely (a.s.) or in probability (i.p.)) as $T \to \infty$, our first step toward proving the convergence of the sequence $\theta^T, T \geq 1$ is to show that (1.2) converges uniformly to a continuous limit over all compact sets Θ of invertible time series models whose autocovariance sequences $\gamma_k\left(\theta\right)$ satisfy $\sum_k \left|\gamma_k\left(\theta\right)\right| < \infty$ (short memory models). For any $h \geq 1$, such uniform convergence is established for the h-step-ahead forecast error analogue of (1.2) and also for all lagged sample second moments of forecast errors in Theorem 7.1 (and in Theorem 8.1 for models that require differencing, e.g. regARIMA models). The mode of convergence obtained, a.s. or i.p., is always the mode of convergence of the sample moments $T^{-1}\sum_{t=1}^{T-k} y_{t+k}y_t$, $k \geq 0$. These results generalize the time series special case of Theorem 4.1 of Findley, Pötscher and Wei (2003) (hereafter FPW 2003) which only covers situations in which either there is no mean function, i.e. $\alpha = 0$ in (1.1), or the correct regressor ξ_t is known and is estimated by Ordinary Least Squares (OLS). In the present article, the regressor ξ_t is assumed to satisfy a slightly weakened version of the well-known conditions of Grenander (1954), so that not only periodic functions e.g. those used to model trading day and moving holiday effects, and polynomials are encompassed, but also the widely used intervention regressors of Box and Tiao (1975). These intervention regressors are transitory in the sense that they decay to zero so rapidly that their coefficients cannot be consistently estimated even when ξ_t is fully known. The model's regressor ξ_t^M is taken to be a subvector of ξ_t. The remaining coordinates of ξ_t can be taken to be those of any vector sequence ξ_t^N whose coordinate variables compensate for the inadequacies of ξ_t^M, so that $\alpha\xi_t = \alpha^M \xi_t^M + \alpha^N \xi_t^N$ for subvectors α^M and α^N of α. When ξ_t^N contains nontransitory regressors that are not asymptotically orthogonal to the sequence ξ_t^M, we describe in Corollary 7.3 how GLS estimates of α^M generally result in smaller asymptotic average squared one-step-ahead forecast error than OLS estimates, thereby establishing an optimality property of GLS. Subsection 7.3 provides instances of a number of our general formulas for the special case of first order autoregressive modeling.

Our basic data and regressor assumptions are given in Section 3, together with Proposition 3.1, which establishes asymptotic orthogonality property needed of transitory regressors. Section 4 provides specifics about ξ_t^M and ξ_t^N and the OLS estimate of α^M, and, in Subsection 4.2, some consequences of the ambiguity in the definition of ξ_t^N. The invertible, short memory autocovariance models for y_t that we consider and their infinite-past forecast functions are discussed in Section 5, where Theorem 6.2 describes uniform limiting properties of the bias $\alpha_T^M(\theta) - \alpha^M$ over compact Θ.

The results of Sections 4–8 apply to "conditional" variates defined by truncated sums like $\sum_{j=0}^{t-1} \theta_j Y_{t-j}$ that are defined as though the infinite-past prediction error filter $(\theta_j)_{j\geq 0}$ were being applied to data subject to the "condition" that $Y_s = 0$ when $s < 0$. Analogous results involving the models' time-varying "finite-past" prediction and prediction error filters are obtained in Theorem 9.2 under slightly stronger assumptions on the model set Θ. Theorem 10.1 describes convergence properties of estimators of θ obtained by maximizing Gaussian likelihood functions or by minimizing sums of squared h-step-ahead forecast errors, using OLS, GLS, or an h-step-ahead forecasting generalization of GLS to estimate the coefficient of ξ_t^M.

Most proofs are presented in Appendix B and utilize a Proposition obtained mainly from Findley, Pötscher and Wei (2001) (hereafter FPW 2001) and two related Lemmas. These auxiliary results are collected in Appendix A.

2. Joint Scalable Asymptotic Stationarity

Under the data assumptions made in the next Section, ξ_t and y_t in (1.1) together form a multivariate sequence with the asymptotic stationarity property we now define. Let $V_t,\ t \geq 1$ be an real-valued column vector sequence, some of whose entries might be stochastic with others nonstochastic, and let I_V denote the identity matrix whose order is the dimension $\dim V$ of V_t. This sequence is said to be scalably asymptotically stationary (S.A.S.) if there exists a decreasing *scaling* sequence $D_{V,1} \geq D_{V,2} \geq \ldots$ of positive definite diagonal matrices $D_{V,T} =$

$diag\left(d_{1,T}^{-1}, \ldots, d_{\dim V,T}^{-1}\right)$ satisfying

$$\lim_{T\to\infty} D_{V,T+k}^{-1} D_{V,T} = I_V \ \ (k = 0, 1, \ldots) \tag{2.1}$$

(coordinatewise convergence) and such that for all $k \geq 0$ the scaled sample second moments $D_{V,T} \sum_{t=1}^{T-k} V_{t+k} V_t' D_{V,T}$ converge as $T \to \infty$, almost surely or in probability, i.e. the limits

$$\Gamma_k^V = \lim_{T\to\infty} D_{V,T} \sum_{t=1}^{T-k} V_{t+k} V_t' D_{V,T} \ a.s.\ [i.p.] \tag{2.2}$$

exist (finitely), the mode of convergence, a.s. or i.p., being the same for all k. For example, for an entry of V_t, of the form t^n, $n \geq 0$, the corresponding diagonal entry of $D_{V,T}$ can be taken to be the reciprocal of $T^{n+\frac{1}{2}}$ (assuming $\Gamma_0^V > 0$).

Under (2.1)-(2.2), negatively lagged scaled sample second moments also converge: for $k > 0$,

$$\Gamma_{-k}^V = \lim_{T\to\infty} D_{V,T} \sum_{t=k+1}^{T} V_{t-k} V_t' D_{V,T} = \left(\Gamma_k^V\right)' \ a.s.\ [i.p.].$$

and the matrix sequence Γ_k^V, $k = 0, \pm 1, \ldots$ is positive semidefinite. (Here and subsequently, the mode of convergence, a.s. or i.p., is to be taken as the mode that applies in (2.2). Due to this property, there is a nondecreasing, positive semidefinite matrix valued function $G_V(\lambda)$ such that

$$\Gamma_k^V = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_V(\lambda),$$

see Grenander (1954) or Chapter II of Hannan (1970). The Γ_k^V are the *asymptotic second moment matrices* of the sequence V_t, and $G_V(\lambda)$ is its *asymptotic spectral distribution matrix.* We say that the entries of V_t are jointly S.A.S. We use the term *asymptotically stationary* (A.S.) when $D_{V,T} = T^{-1/2} I_V$, i.e., when the sample second moments $T^{-1}\sum_{t=1}^{T-k} V_{t+k} V_t'$ converge a.s. [i.p.].

The properties (2.1)–(2.2) yield

$$\lim_{T\to\infty} D_{V,T} V_{T-j} = 0 \ a.s.\ [i.p.],\ \ j \geq 0. \tag{2.3}$$

For example,

$$\begin{aligned} & D_{V,T} V_T' V_T D_{V,T} \\ = \ & D_{V,T} \sum_{t=1}^{T} V_t' V_t D_{V,T} - \left(D_{V,T} D_{V,T-1}^{-1}\right) D_{V,T-1} \sum_{t=1}^{T-1} V_{t+k} V_t D_{V,T-1} \left(D_{V,T-1}^{-1} D_{V,T}\right) \\ \to \ & \Gamma_0^V - \Gamma_0^V = 0 \ a.s.\ [i.p.]. \end{aligned}$$

If

$$\lim_{T\to\infty} D_{V,T} = 0, \tag{2.4}$$

holds, then also $\lim_{T\to\infty} D_{V,T} V_{1+j} = 0 \ a.s.\ [i.p.]$ for all $j \geq 0$.

2.1. Array reformulation

To connect with the results of FPW (2001, 2003) and to facilitate interpretation, we reformulate the basic properties of interest as follows. Under (2.1)–(2.2), the *array*

$$V_t(T) = T^{1/2} D_{V,T} V, \ 1 \leq t \leq T, T = 1, 2, \ldots \tag{2.5}$$

is asymptotically stationary in the sense that the limits

$$\Gamma_k^V = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} V_{t+k}(T) V_t(T)' \ a.s.\ [i.p.] \ (k = 0, 1, \ldots) \tag{2.6}$$

exist (with Γ_k^V as in (2.2)), and it also has the property

$$\lim_{T\to\infty} T^{-1/2} V_{T-j}(T) = 0 \ a.s.\ [i.p.], \text{ for fixed } j \geq 0. \tag{2.7}$$

When (2.4) holds, then we further have

$$\lim_{T\to\infty} T^{-1/2} V_{1+j}(T) = 0 \ a.s.\ [i.p.]\ , j \geq 0. \tag{2.8}$$

Following FPW (2001, 2003), we call (2.7) and (2.8) *negligibility* properties.

Two A.S. arrays $V_t(T), W_t(T),\ 1 \leq t \leq T,\ T = 1, 2, \ldots$ are said to be *asymptotically orthogonal* if

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} V_{t+k}(T) W_t(T)' = 0 \ a.s.\ [i.p.] \ (k = 0, 1, \ldots),$$

the mode of convergence being that of (2.6).

3. Data and Regressor Assumptions

For data of the form (1.1), we require y_t to be A.S., i.e. the limits

$$\gamma_k^y = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} y_{t+k} y_t \ a.s.\ [i.p.] \tag{3.1}$$

exist for $k = 0, 1, \ldots$. The associated asymptotic spectral distribution is denoted $G_y(\lambda)$.

The regressor sequence ξ_t, $t \geq 1$ in (1.1) is required to be S.A.S. and to have two additional properties. First, the two series y_t and ξ_t must be asymptotically orthogonal,

$$\lim_{T\to\infty} T^{-\frac{1}{2}} \sum_{t=k+1}^{T-k} y_t \xi'_{t\pm k} D_{\xi,T} = 0 \ a.s.\ [i.p.], \ (k = 0, 1, \ldots). \tag{3.2}$$

(In statements like this, the applicable mode of convergence is the mode assumed in (3.1)). Second, for

$$\Gamma_k^\xi = \lim_{T\to\infty} D_{\xi,T} \sum_{t=1}^{T-k} \xi_{t+k} \xi'_t D_{\xi,T} \ (k = 0, 1, \ldots), \tag{3.3}$$

we require

$$\Gamma_0^{\xi} > 0. \tag{3.4}$$

Note that if ξ_t contains a coordinate that is constant, e.g. equal to 1 for all t, then the corresponding scaling factor in $D_{\xi,T}$ can be taken to be $T^{-1/2}$, and (3.2) yields $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} y_t = 0$ a.s. [i.p.]. In this sense, y_t in (1.1) can be thought of as an asymptotically mean zero process.

If ξ_t contains a coordinate whose scaling sequence does not decrease to zero, then its scaling scaling sequence can be taken to be a positive constant, for example 1.0. Thus we can arrange the coordinates of ξ_t so that

$$\xi_t = \begin{bmatrix} X_t \\ x_t \end{bmatrix}, \tag{3.5}$$

where

$$D_{\xi,T} = diag \begin{bmatrix} D_{X,T} & 0 \\ 0 & I_x \end{bmatrix},$$

with I_x denoting the identity matrix of order $\dim x_t$ and with $D_{X,T}$ decreasing to 0,

$$D_{X,T} \searrow 0. \tag{3.6}$$

Our final basic regressor assumption is

$$\sum_{t=1}^{\infty} (x_t' x_t)^{1/2} < \infty, \tag{3.7}$$

($'$ denotes transpose) to guarantee that X_t and x_t are asymptotically orthogonal,

$$\lim_{T\to\infty} \sum_{t=1+k}^{T} x_t X_{t\pm k}' D_{X,T} = 0 \quad (k = 0, 1, \ldots), \tag{3.8}$$

see Proposition 3.1.below which also shows that asymptotic orthogonality with y_t,

$$\lim_{T\to\infty} T^{-\frac{1}{2}} \sum_{t=1+k}^{T} x_t y_{t\pm k} = 0 \quad (k = 0, 1, \ldots),$$

follow from (3.7) and the A.S. property of y_t. Hereafter, we refer to (3.1)–(3.4) and (3.7) as the assumptions of this subsection.

Any sequence x_t satisfying (3.7) (or the weaker condition $\sum_{t=1}^{\infty} x_t' x_t < \infty$) is S.A.S. with $D_{x,T} = I_x$ for $T \geq 1$ and $\Gamma_k^x = \sum_{t=1}^{\infty} x_t x_{t+k}'$ for $k \geq 0$. By virtue of (3.8), the S.A.S. property of ξ_t reduces to the S.A.S. property of X_t. Following Grenander (1954), Hannan (1970, pp. 78–79) verifies the latter for regressors X_t whose components X_{it} are cosinusoids, $\cos \lambda t$ and $\sin \lambda t$, with $0 < \lambda \leq \pi$ or linear combinations thereof, e.g. periodic functions, or are polynomials. Anderson (1971, pp. 581–582) does this also for products of polynomials and cosinusoids.

The property (3.2) holds quite generally. For example, under (3.4) it is equivalent to

$$\lim_{T\to\infty} T^{-1} \sum_{t=k+1}^{T-k} y_t \xi_{t\pm k}' \left(\sum_{t=k+1}^{T-k} \xi_{t\pm k} \xi_{t\pm k}' \right)^{-1} \sum_{t=k+1}^{T-k} \xi_{t\pm k} y_t = 0, \ a.s. \ [i.p.] \ (k = 0, 1, \ldots), \tag{3.9}$$

because the l.h.s. of (3.9) is equal to

$$\left(\frac{1}{T^{\frac{1}{2}}}\sum_{t=k+1}^{T-k} y_t \xi'_{t\pm k} D_{\xi,T}\right)\left(D_{\xi,T}\sum_{t=k+1}^{T-k} \xi_{t\pm k}\xi'_{t\pm k} D_{\xi,T}\right)^{-1}\left(\frac{1}{T^{\frac{1}{2}}}\sum_{t=k+1}^{T-k} y_t \xi'_{t\pm k} D_{\xi,T}\right)'.$$

Section 3 of FPW (2001) shows that (3.9) holds a.s. (for arbitrary ξ_t if Moore-Penrose inverses are used) when y_t is a weakly stationary linear processes, $y_t = \sum b_j \varepsilon_{t-j}$, with independent white noise process ε_t such that $\sup_t E\left|\varepsilon_t\right|^r < \infty$, for some $r > 2$ when the spectral density of y_t is bounded, or for some $r > 4$ when the spectral density is unbounded but square integrable.

Remark 1. All of the results of the paper continue to hold for stochastic regressors ξ_t satisfying (3.2) if convergence in (3.3) holds a.s. and convergence statements involving the sample second moment matrices of regressors are interpreted throughout the paper as holding a.s. However, the interpretation changes. In the stochastic regressor case, $\alpha\xi_t$ is no longer interpreted as the mean function of Y_t, and the functions described as log-likelihood functions below could more properly be called log-quasilikehihood functions. For simplicity, outside this remark, we shall only refer to the case of nonstochastic regressors (some of which can be realizations of stochastic processes).

3.1. On transitory regressors

The assumption (3.7) implies that x_t models transitory effects whose coefficients cannot be consistently estimated. ($\sum_{t=1}^{\infty} x'_t x_t = \infty$ is a necessary condition for consistent estimation, see Lai and Wei, 1984, whose argument for the OLS case extends to the GLS cases we consider by virtue of the inequalities of Lemma 12.3 and Proposition 9.1 below.) The usual examples of x_t, namely additive outlier regressors, level-shift regressors, and the other intervention regressors of Box and Tiao (1975), decay at least exponentially to zero and therefore satisfy the much stronger summability condition $\sum_{t=1}^{\infty}(1+\varepsilon)^t\left(x'_t x_t\right)^{1/2} < \infty$ for some $\varepsilon > 0$.

The next result, whose proof is in Appendix B, shows that (3.2) and (3.8) are automatically satisfied under (3.6) and (3.7). For later use, we give a result that applies uniformly to families of sequences. For a column vector v, define $\|v\| = \left(v'v\right)^{1/2}$, and, for a matrix M whose column dimension is the dimension of v, define $\|M\| = \sup_{\|v\|=1}\|Mv\| = \lambda_{\max}^{1/2}\left(M'M\right)$. We use $\lambda_{\max}(\cdot)$ resp. $\lambda_{\min}(\cdot)$ to denote the maximal resp. minimal eigenvalue.

Proposition 3.1. *For a given index set H, suppose the family of sequences $x_t(\eta), t \geq 1,\ \eta \in H$ satisfies*

$$\sup_{\eta\in H}\sum_{t=1}^{\infty}\left\|x_t(\eta)\right\| < \infty. \tag{3.10}$$

Let $V_t(\zeta), t \geq 1,\ \zeta \in Z$ be a family of asymptotically S.A.S. sequences with a common scaling matrix sequence $D_{V,T}$ that satisfies (2.4). Then under

$$\sup_{t\geq 1,\zeta\in Z}\left\|D_{V,t}V_t(\zeta)\right\| < \infty \text{ a.s. [i.p.]} \tag{3.11}$$

and

$$\lim_{t\to\infty}\sup_{\zeta\in Z}\left\|D_{V,t}V_t(\zeta)\right\| = 0 \text{ a.s. [i.p.]}, \tag{3.12}$$

it holds that

$$\lim_{T\to\infty} \sup_{\eta\in H,\zeta\in Z} \sum_{t=1+k}^{T} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,T} \right\| = 0 \text{ } a.s. \text{ } [i.p.] \text{ } (k = 0, 1, \ldots). \tag{3.13}$$

Consequently, the families of sequences $x_t(\eta)$ *and* $V'_{t\pm k}(\zeta)$ *are uniformly asymptotically orthogonal,* $\sup_{\eta\in H,\zeta\in Z} \left\| \sum_{t=1+k}^{T} x_t(\eta) V'_{t\pm k}(\zeta) D_{V,T} \right\| \to 0$ *a.s.* [*i.p.*] $(k = 0, 1, \ldots)$.

Remark 2. Our later results will show that the transitory regressors x_t make no contribution to the limiting second moments of out-of-sample forecast errors, a result that is neither surprising nor completely obvious. We have included them in our analysis because of their importance. Such regressors (primarily for additive outliers and level shifts and exponentially decaying intervention variables) occur in the majority of regARIMA models fit to thousands of time series by statistical offices and central banks around the world.

3.2. OLS estimation of α

We now consider convergence properties of the ordinary least squares (OLS) estimator

$$\alpha_T = \sum_{t=1}^{T} Y_t \xi'_t \left[\sum_{t=1}^{T} \xi_t \xi'_t \right]^{-1}. \tag{3.14}$$

of α in (1.1). Because

$$\alpha_T - \alpha = \sum_{t=1}^{T} y_t \xi'_t \left[\sum_{t=1}^{T} \xi_t \xi'_t \right]^{-1} = \sum_{t=1}^{T} y_t \xi'_t D_{\xi,T} \left[D_{\xi,T} \sum_{t=1}^{T} \xi_t \xi'_t D_{\xi,T} \right]^{-1} D_{\xi,T},$$

under (3.4) we have the equivalence of (3.2) and

$$T^{-1/2} (\alpha_T - \alpha) D_{\xi,T}^{-1} \to O \text{ } a.s. \text{ } [i.p.]. \tag{3.15}$$

Remark 3. Let $\alpha = \begin{bmatrix} A & a \end{bmatrix}$ and $\alpha_T = \begin{bmatrix} A_T & a_T \end{bmatrix}$ be the partitions of α and α_T corresponding to the partition (3.5) of ξ_t. It follows from (3.15) that A_T is a consistent estimator of A when the entries of the matrices $T^{\frac{1}{2}} D_{X,T}$, $T \geq 1$ are bounded, i.e.,

$$T^{\frac{1}{2}} D_{X,T} \leq K I_X \tag{3.16}$$

holds for some $K > 0$, as Hannan (1970) observed for the case in which $\xi_t = X_t$, i.e. no transitory regressors occur. For example, if all regressors of X_t are periodic, then $T^{1/2} D_{X,T} = I_X$. More generally, (3.16) is satisfied when X_t consists of periodic functions, polynomials, and their products. We do not require (3.16) or consistency for our main results.

3.3. Array reformulation of the basic properties

The following reformulation of the properties required of ξ_t will enable us to make use of the results of FPW (2001, 2003). Together, (3.8) and (3.2) yield that the array

$$V_t(T) = \begin{bmatrix} y_t \\ T^{1/2} D_{X,T} X_t \\ T^{1/2} x_t \end{bmatrix}, \; 1 \le t \le T, T = 1, 2, \ldots, \tag{3.17}$$

has the following A.S. property:

$$\Gamma_k^V = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} V_{t+k}(T) V_t(T)'$$

$$= \begin{bmatrix} \gamma_k^y & 0 & 0 \\ 0 & \Gamma_k^X & 0 \\ 0 & 0 & \Gamma_k^x \end{bmatrix}, \; a.s. \; [i.p.] \quad (k = 0, 1, \ldots), \tag{3.18}$$

with $\Gamma_k^x = \sum_{t=1}^{\infty} x_{t+k} x_t'$, and $\Gamma_k^X = \lim_{T\to\infty} D_{X,T} \sum_{t=1+k}^{T} X_{t+k} X_t' D_{X,T}$ for each $k \ge 0$ and $\Gamma_0^x > 0$ and $\Gamma_0^X > 0$. The negligibility property (2.3) becomes

$$\lim_{T\to\infty} T^{-1/2} V_{T-j}(T) = 0 \; a.s. \; [i.p.], j \ge 0.$$

Due to (3.6), the subvector array,

$$U_t(T) = \begin{bmatrix} y_t \\ T^{1/2} D_{X,T} X_t \end{bmatrix}, \; 1 \le t \le T, T = 1, 2, \ldots, \tag{3.19}$$

whose asymptotic second moments are $\Gamma_k^U = diag\left(\gamma_k^y, \Gamma_k\right)$, has the additional negligibility property

$$\lim_{T\to\infty} T^{-1/2} U_{1+j}(T) = 0 \; a.s. \; [i.p.], j \ge 0, \tag{3.20}$$

because the scaling matrices $diag\left(T^{-1/2}, D_{X,T}\right)$ of $\begin{bmatrix} y_t & X_t' \end{bmatrix}'$ decrease to 0.

The asymptotic spectral distribution matrices of $V_t(T)$ and $U_t(T)$ thus have block diagonal forms $G_V(\lambda) = diag\left(G_y(\lambda), G_X(\lambda), G_x(\lambda)\right)$ and $G_U(\lambda) = diag\left(G_y(\lambda), G_X(\lambda)\right)$, respectively.

4. Regressor Misspecification and Some Consequences

Because the regression vector $\xi_t^M = \begin{bmatrix} X_t^{M\prime} & x_t^{M\prime} \end{bmatrix}'$ used by the modeler is subvector of ξ_t, we can arrange for the components x_t and X_t in (3.5) to be partitioned as

$$x_t = \begin{bmatrix} x_t^M \\ x_t^N \end{bmatrix}, \; X_t = \begin{bmatrix} X^M \\ X_t^N \end{bmatrix}, \tag{4.1}$$

where the superscript N designates the regressors not included in the model. Let the corresponding partition of α in (1.1) be $\alpha = \begin{bmatrix} A^M & A^N & a^M & a^N \end{bmatrix}$ and those of $D_{X,T}$, Γ_k^X and $G_X(\lambda)$ be

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

$$\Gamma_k^X = \begin{bmatrix} \Gamma_k^{MM} & \Gamma_k^{MN} \\ \Gamma_k^{NM} & \Gamma_k^{NN} \end{bmatrix}, \tag{4.2}$$

and

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

respectively. Setting $\xi_t^M = \left[X_t^{M\prime} x_t^{M\prime}\right]'$, $\xi_t^N = \left[X_t^{N\prime} x_t^{N\prime}\right]'$, $\alpha^M = \left[A^M a^M\right]$, $\alpha^N = \left[A^N a^N\right]$ and

$$y_t^M = \alpha^N \xi_t^N + y_t = a^N x_t^N + A^N X_t^N + y_t, \tag{4.3}$$

the data decomposition being modeled is

$$Y_t = \alpha^M \xi_t^M + y_t^M. \tag{4.4}$$

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

$$D_{N,T} = T^{-1/2} I_N, \tag{4.5}$$

with I_N being the identity matrix of appropriate dimension. Omitted regressor variables of larger order would give rise to model residuals that would become infinite in magnitude as T increases and would therefore be recognized as not A.S. with large enough samples. Under (4.5), it follows from (3.18) that y_t^M is A.S. For each $k \geq 0$,

$$\begin{aligned} \gamma_k^M &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-k} y_{t+k}^M y_t^M \ a.s.\ [i.p.] \\ &= A^N \Gamma_k^{NN} A^{N\prime} + \gamma_k^y = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_{y^M}(\lambda), \end{aligned}$$

with $G_{y^M}(\lambda) = A^N G^{NN}(\lambda) A^{N\prime} + G_y(\lambda)$. (We note that by setting $A^N = 0$ in this limiting formula and those of the propositions and theorems below, the formulas that apply when the correct regressor is used in the model, $\xi_t^M = \xi_t$, can be obtained.)

If $A^N \neq 0$, then y_t^M and ξ_t^M will not usually be asymptotically orthogonal. Specifically,

$$\begin{aligned} &\lim_{T\to\infty} T^{-1/2} \sum_{t=1}^{T-k} y_{t+k}^M X_t^{M\prime} D_{M,T} \\ = &\lim_{T\to\infty} T^{-1/2} \sum_{t=1}^{T-k} A^N X_{t+k}^N X_t^{M\prime} D_{M,T} = A^N \Gamma_k^{NM} \ a.s.\ [i.p.] \end{aligned} \tag{4.6}$$

will generally be non-zero for some k unless the sequences X_t^M and X_t^N are asymptotically orthogonal.

Only slight modifications would be needed to (4.6) and the related limiting formulas obtained below if we replaced (4.5) with the weaker condition $\lim_{T\to\infty} T^{1/2} D_{N,T} = \tilde{D}_N$ with $\tilde{D}_N$ finite and possibly singular, in order to cover the situation in which, for example, X_t^N includes a regression variable of the form t^q, $-.5 \leq q < 0$.

4.1. OLS estimation of α^M

Given data $Y_t, 1 \leq t \leq T$, the OLS estimator of α^M for the model (4.4) is

$$\alpha_T^M = \sum_{t=1}^{T} Y_t \xi_t^{M\prime} \left[\sum_{t=1}^{T} \xi_t^M \xi_t^{M\prime} \right]^{-1}. \tag{4.7}$$

With $I_{x,M}$ denoting the identity matrix of order $\dim x_t$, define

$$D_{\xi,M,T} = \begin{bmatrix} I_{x,M} & 0 \\ 0 & D_{M,T} \end{bmatrix}.$$

Let $\alpha_T^M = \begin{bmatrix} A_T^M & a_T^M \end{bmatrix}$ be the partition of α_T^M corresponding to the partition $\alpha^M = \begin{bmatrix} A^M & a^M \end{bmatrix}$. Clearly

$$\alpha_T^M - \alpha^M = \sum_{t=1}^{T} y_t \xi_t^{M\prime} \left[\sum_{t=1}^{T} \xi_t^M \xi_t^{M\prime} \right]^{-1} + \alpha^N \sum_{t=1}^{T} \xi_t^N \xi_t^{M\prime} \left[\sum_{t=1}^{T} \xi_t^M \xi_t^{M\prime} \right]^{-1}.$$

So if we define

$$C^{NM} = \Gamma_0^{NM} \left(\Gamma_0^{MM} \right)^{-1}, \tag{4.8}$$

it follows from (3.2), (4.5), and (4.6) that

$$\left(\alpha_T^M - \alpha^M \right) T^{-1/2} D_{\xi,M,T}^{-1} \to \begin{bmatrix} A^N C^{NM} & 0 \end{bmatrix} \ a.s.\ [i.p.], \tag{4.9}$$

i.e., $T^{-1/2} \left(A_T^M - A^M \right) D_{M,T}^{-1} \to A^N C^{NM}$ a.s. [i.p.], and $T^{-1/2} \left(a_T^M - a^M \right) \to O$ a.s. [i.p.].

4.2. Ambiguity concerning X_t^N and asymptotic "bias"

With $\xi_t^M(T) = T^{1/2} D_{M,T} \xi_t^M$, $1 \leq t \leq T$, the result (4.9) can be interpreted as showing that the OLS estimate $\alpha_T^M T^{-1/2} D_{M,T}^{-1}$ of the coefficient $\alpha^M T^{-1/2} D_{M,T}^{-1}$ in the relation

$$Y_t = \left\{ \alpha^M T^{-1/2} D_{M,T}^{-1} \right\} \xi_t^M(T) + y_t^M,\ 1 \leq t \leq T$$

has the asymptotic bias $\begin{bmatrix} A^N C^{NM} & 0 \end{bmatrix}$. If $A^N C^{NM} \neq 0$, this bias has the desirable consequence that $\alpha_T^M \xi_t^M$ is, asymptotically, a better predictor than $\alpha^M \xi_t^M$ of $\alpha \xi_t$, see $\Theta^*7.2$ below, where a more general result applicable to GLS estimates is presented. More often than not, considerations of simplicity or parsimony give rise to regressors in ξ_t^M that are clearly incomplete approximations to the effects being modeled. In particular, if a fixed definition of the regressor sequence X_t^N is maintained, one can have $\Gamma_0^{NM} \neq 0$, and therefore $C^{NM} \neq 0$. However, when $D_{M,T} = T^{-1/2}$, if ambiguity in the definition of X_t^N (and therefore in the definitions of ξ_t^N and ξ_t) is accepted, then the subvector A_T^M of α_T^M that provides the OLS coefficient estimates of X_t^M can be assumed to be consistent (asymptotically *unbiased*) with no loss of generality: Indeed, if $A^N C^{NM} \neq 0$, the identity

$$\begin{aligned} AX_t &= A^M X_t^M + A^N X_t^N \\ &= \left(A^M + A^N C^{NM} \right) X_t^M + A^N \left(X_t^N - C^{NM} X_t^M \right) \\ &= \breve{A}^M X_t^M + A^N \breve{X}_t^N \end{aligned}$$

shows that X_t^N can replaced by $\breve{X}_t^N = X_t^N - C^{NM} X_t^M$, which results in $\breve{A}^M = A^M + A^N C^{NM}$, the limit of A_T^M, as the coefficient to be estimated. (Because A_T^M does not depend on the definition of X_t^N, neither does its limit $\breve{A}^M$.) The subvector A^M of A provides unambiguously "natural" coefficients for the entries of X_t^M from the asymptotic perspective only when the sequence X_t^N is asymptotically orthogonal to X_t^M.

In Section 7.2, it will be made clear that optimal one-step ahead forecasting of Y_t does not require an optimal estimate of $\alpha\xi_t$ or AX_t. Instead it requires a GLS estimate of the coefficient vector of X_t^M whose limiting value when $D_{M,T} = T^{-1/2}$ differs from $\breve{A}^M$ except in special situations.

A formula for the asymptotic bias of estimators of the transitory regression coefficient vector a^M, including the limit of $a_T^M - a^M$, is obtained in Remark 7 of Section 9 for a broad class of transitory regressors.

4.3. Examples of Asymptotically Orthogonal Regressors

Let $G^{ij}(\lambda)$ denote the (i,j)-entry of $G^{XX}(\lambda)$. The 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, all differences $\Delta G^{ij} = G^{ij}(\lambda'') - G^{ij}(\lambda')$ with $-\pi \le \lambda' < \lambda'' \le \pi$ have the value zero. Because the positive semidefinitenenss of $\Delta G^{XX}(\lambda)$ 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, $G^{ii}(\lambda_k+) - G^{ii}(\lambda_k-) > 0$, and G^{jj} is continuous at these frequencies, $G^{jj}(\lambda_k+) - G^{jj}(\lambda_k-) = 0$.

Anderson (1971, p. 581–582) shows that a regressor of the form

$$c_0 + \sum_{k=1}^{H} (c_k \cos\lambda_k t + d_k \sin\lambda_k t) + c_{H+1}(-1)^{H+1}, \tag{4.10}$$

for example, a periodic regressor, has a spectral distribution function that has jumps at each frequency λ_k, assuming $c_k^2 + d_k^2 \ne 0$, and also at the frequency 0, resp. π, if $c_0 \ne 0$, resp. $c_{H+1} \ne 0$. 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 regressors of the types mentioned are asymptotically orthogonal if and only if they have no common frequency components. In particular, polynomials in t are asymptotically orthogonal to periodic regressors with mean zero ($c_0 = 0$). For the same reason, deseasonalized regressors of the sort used to model trading day and holiday effects (see Findley et al. (1998) and Findley and Soukup (2001)) are asymptotically orthogonal to seasonal regressors. Similarly, a regressor that is a realization of a second moment stationary time series whose second moments are determined by a spectral density (i.e. whose spectral measure is absolutely continuous) is asymptotically orthogonal to polynomials, to regressors of the form (4.10), and to their products.

In the misspecified regressor situation, if some coordinate $X_{i,t}^M$ of X_t^M is constant, e.g. $X_{i,t}^M = 1$ for all t, then X_t^N can be assumed to be asymptotically orthogonal to this constant regressor, i.e. to have $\bar{X}^N = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} X_t^N$ a.s. [i.p.] 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 $A_i^M + A^N \bar{X}^N$. As a consequence, one can usually assume that $G^{NN}(\lambda)$ is 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 $\lambda = 0$, as we indicated above.

5. Invertible, Short-Memory Modeling and Forecasting for y_t and y_t^M

Because of their asymptotic stationarity, it is natural to assume that y_t and y_t^M will be modeled as weakly stationary time series. As in FPW (2003), the models we consider are those that have spectral densities of the form

$$f_{\theta,\sigma}(\lambda) = \frac{\sigma^2}{2\pi}\left|1 + \sum_{j=1}^{\infty}\theta_j e^{ij\lambda}\right|^{-2}, \tag{5.1}$$

with $\sigma > 0$ and with the real-valued coefficient sequence $\theta = (1, \theta_1, \theta_2, \dots)$ belonging to the set

$$\Theta_{is} = \left\{\theta : \sum_{j=0}^{\infty}|\theta_j| < \infty, \text{ and } \theta(z) \neq 0 \text{ if } |z| \leq 1\right\}, \tag{5.2}$$

where $\theta(z)$ denotes the coefficient generating function, $\sum_{j=0}^{\infty}\theta_j z^j$, and $\theta_0 = 1$. Invertible ARMA models and the exponential models of Bloomfield (1973) are the familiar examples of model families with spectral densities of the form (5.1) with $\theta \in \Theta_{is}$.

It follows from (5.1) and (5.2) that each $f_{\theta,\sigma}(\lambda)$ is strictly positive and is continuous as a function of λ (in fact, jointly continuous in θ, σ, λ when distance in Θ_{is} is measured with the l^1-norm $\|\theta\|_1 = \sum_{j=0}^{\infty}|\theta_j|$). The subscripts i and s are used to indicate that the θ in Θ_{is} "parameterize" the autoregressive representations of all *invertible*, *short-memory* time series models. (By definition, these are the models with a strictly positive spectral density and an absolutely summable autocovariance sequence. Theorem 3.8.4 of Brillinger (1975, p.78) shows that each such model has a spectral density of the form (5.1) with $\theta \in \Theta_{is}$ and $\sigma > 0$. This class includes what Brockwell and Davis (1991, Section 13.2) call (invertible) *intermediate-memory* models, meaning models whose lag k autocovariance is of order $|k|^{-\alpha}$ with $1 < \alpha < 2$) While the *models* considered are short-memory, the data assumptions (Section 3.1) permit long-memory behavior, $\sum_k |\gamma_k^y| = \infty$. The parameter θ determines the autocorrelation structure specified by the model, and σ determines the scale.

To express the various forecast functions of the model defined by a given $\theta \in \Theta_{is}$, we utilize the associated sequence $\tilde{\theta} = (1, \tilde{\theta}_1, \tilde{\theta}_2, \dots)$ of moving average (or innovations) representation coefficients whose entries are defined by the power series relation $1 + \sum_{j=1}^{\infty}\tilde{\theta}_j z^j = (1 + \sum_{j=1}^{\infty}\theta_j z^j)^{-1}$. With $\tilde{\theta}_0 = 1$, the $\tilde{\theta}_j,\ j \geq 1$, can be calculated recursively,

$$\tilde{\theta}_j = -\sum_{i=1}^{j}\theta_i\tilde{\theta}_{j-i}\,,\ j = 1, 2, \dots\ . \tag{5.3}$$

For $\theta \in \Theta_{is}$, a well-known result of Wiener affirms that $\left\|\tilde{\theta}\right\|_1 = \sum_{j=0}^{\infty}\left|\tilde{\theta}_j\right| < \infty$. Moreover, the mapping $\theta \mapsto \tilde{\theta}$ is a one-to-one mapping of Θ_{is} onto itself that is its own inverse, and it is $\|\cdot\|_1$-continuous, see 10.12 and the proof of 11.6 of Rudin (1973).

5.1. The θ-model's forecast and forecast error filters

Let y_t^θ denote a zero mean, weakly stationary time series with spectral density (5.1) for some $\theta \in \Theta_{is}$ and $\sigma > 0$. Then for any $h \geq 1$, the optimal linear h-step-ahead predictor of y_{t+h}^θ from the *infinite* past $y_s^\theta, -\infty < s \leq t$, has the form $\sum_{j=0}^{\infty} \pi_j(h,\theta) y_{t-j}^\theta$ with coefficients $\pi_j(h,\theta), j = 0, 1, \ldots$ that minimize $\int_{-\pi}^{\pi} \left| 1 - \sum_{j=0}^{\infty} \pi_j e^{i(h+j)\lambda} \right|^2 f_{\theta,\sigma}(\lambda)\, d\lambda$. The minimizing coefficients can be calculated from the identity

$$\sum_{j=0}^{\infty} \pi_j(h,\theta) z^{j+h} = \sum_{i=h}^{\infty} \tilde{\theta}_i z^i (\sum_{i=0}^{\infty} \theta_i z^i) = 1 - (\sum_{i=0}^{h-1} \tilde{\theta}_i z^i)(\sum_{i=0}^{\infty} \theta_i z^i), \tag{5.4}$$

see, for example, Theorem III.2.6 of Hannan (1970 p. 147). Thus, for $j \geq 0$, $\pi_j(h,\theta) = -\sum_{i=0}^{h-1} \tilde{\theta}_i \theta_{j+h-i}$. Consequently, the prediction filter $\pi(h,\theta) = (\pi_j(h,\theta))_{j\geq 0}$ is absolutely summable, $\|\pi(h,\theta)\|_1 \leq 1 + \|\theta\|_1 \sum_{j=0}^{h-1} \left|\tilde{\theta}_j\right| < \infty$. The associated prediction error has the formula $\sum_{j=0}^{\infty} \eta_j(h,\theta) y_{t+h-j}^\theta$, with absolutely summable coefficients given by

$$\eta_j(h,\theta) = \begin{cases} 1 & , \ j = 0 \\ 0 & , \ 1 \leq j \leq h-1 \\ -\pi_{j-h}(h,\theta) = \sum_{i=0}^{h-1} \tilde{\theta}_i \theta_{j-i} & , \ j \geq h \end{cases} .$$

In view of (5.4), the corresponding generating function $\eta(h,\theta)(z) = \sum_{j=0}^{\infty} \eta_j(h,\theta) z^j$ is given by

$$\eta(h,\theta)(z) = \theta(z) \sum_{i=0}^{h-1} \tilde{\theta}_i \, z^i. \tag{5.5}$$

Of course, $\eta(1,\theta)(z) = \theta(z)$.

In certain incorrect model situations, it can happen that the linear predictor that is optimal in the sense of minimizing mean squared one-step-ahead forecast error based on an infinite past has a prediction error filter θ with the property that $\theta(z)$ has a zero of magnitude one. Such a θ belongs to

$$\bar{\Theta}_{is} = \left\{ \theta : \ \sum_{j=0}^{\infty} |\theta_j| < \infty, \text{ and } \theta(z) \neq 0 \text{ if } |z| < 1 \right\}, \tag{5.6}$$

but not to Θ_{is}, see Appendix A of Pötscher (1991), which shows how such θ arise from the large sample limit properties of standard parameter estimates. Each $\theta \in \bar{\Theta}_{is}$ defines a one-step-ahead infinite-past (prediction and) prediction error filter that is applicable to any weakly stationary time series. We shall regard each $\theta \in \bar{\Theta}_{is}$ as a model for y_t^M in (4.3) that can provide a generalized least squares estimate of α^M. For these models, in the next Sections and Subsections we derive limiting properties of average squared errors of "truncated" (or "conditional") forecasts of the sort we now define. Later we shall establish limiting properties of parameter estimates determined by minimizing sample mean squared forecast error.

Suppose finitely many observations are available of a possibly multivariate time series V_t, $1 \leq t \leq T$ or time series array $V_t(T)$, $1 \leq t \leq T, T = 1, 2, \ldots$. Given any filter $\phi = (\phi_0, \phi_1, \ldots)$, we define

$$V_t[\phi](T) = \begin{cases} \sum_{j=0}^{t-1} \phi_j V_{t-j,T}, & 1 \leq t \leq T, \\ 0 & t < 0. \end{cases} . \tag{5.7}$$

Because the filter coefficients ϕ_j are scalars, when $V_t(T)$ is defined by (2.5), then

$$V_t[\phi](T) = T^{1/2} D_{V,T} V_t[\phi], \ 1 \leq t \leq T.$$

For forecasting, with any $\theta \in \Theta_{is}$ and finite data span, $V_t(T), 1 \leq t \leq T$, we define the θ-model's *truncated* infinite-past forecast functions h-step-ahead forecasts to be $V_t[\pi(h,\theta)](T), 1-h \leq t \leq T$. Then the observable forecast errors are given by $V_t(T) - V_{t-h}[\pi(h,\theta)](T) = V_t[\eta(h,\theta)](T), 1 \leq t \leq T$. The time-varying "finite-past" forecast and forecast error functions defined each $\theta \in \Theta_{is}$ will be considered in Section 9.

To understand the role of $\bar{\Theta}_{is}$ it is helpful to know the following fact, which is established in Appendix B.

Lemma 5.1. *If a sequence θ^T, $T = 1, 2, \ldots$ in $\bar{\Theta}_{is}$ converges to some θ in absolute sum norm, $\left\|\theta^T - \theta\right\|_1 \to 0$, then $\theta(z)$ has no zeros in $\{|z| < 1\}$, i.e. $\theta \in \bar{\Theta}_{is}$. Further, every $\theta \in \bar{\Theta}_{is}$ not in Θ_{is} is the limit in this sense of a sequence θ^T, $T = 1, 2, \ldots$ in Θ_{is}. Thus $\bar{\Theta}_{is}$ is the $\|\cdot\|_1$-closure of Θ_{is}.*

6. GLS estimation of α and α^M

In the situation in which y_t in (1.1) is a Gaussian process that is defined for all $-\infty < t < \infty$, the one-step-ahead forecast error process $\eta(1,\theta)(B) y_t = \theta(B) y_t$ is uncorrelated and therefore i.i.d. when the θ-model's spectral density function correctly describes the autocorrelation structure of y_t. Then, application of $\theta(B)$ to both sides of (1.1) produces data for which OLS estimation of α is efficient. To implement this idea with finite data $Y_t, 1 \leq t \leq T$, one can, as in Pierce (1971), use the truncated infinite-past one-step-ahead forecast error filters to define Generalized Least Squares (GLS) estimates for any θ-model of interest. In the notation of (5.7), wherewith $Y_t[\theta] = \sum_{j=0}^{t-1} \theta_j Y_{t-j}$ and $\xi_t[\theta] = \sum_{j=0}^{t-1} \theta_j \xi_{t-j}$, this GLS estimator of α in (1.1) is

$$\alpha_T(\theta) = \sum_{t=1}^{T} Y_t[\theta] \, \xi_t[\theta]' \left(\sum_{t=1}^{T} \xi_t[\theta] \, \xi_t[\theta]' \right)^{-1}. \tag{6.1}$$

Note that $\alpha_T(\theta)$ reduces to the OLS estimator (3.14) when θ is the parameter for white noise, $\theta = (1, 0, 0, \ldots)$. In (6.1) and elsewhere, a generalized inverse is to be understood whenever an inverse matrix fails to exist. For $\theta \in \Theta_{is}$, this can only happen for a fixed finite number of T values, due to (3.3)–(3.4) and (e) of Proposition 12.1 in Appendix A below.

An alternative GLS estimator will be discussed in Section 9.

6.1. The $\|\cdot\|_1$-compact subsets of $\bar{\Theta}_{is}$

Henceforth, we consider only model families $\Theta \subseteq \bar{\Theta}_{is}$ whose absolute coordinate sums $\sum_{j=0}^{\infty} |\theta_j|$ converge uniformly on Θ, i.e., the pair of conditions

$$\begin{aligned} \sup_{\theta \in \Theta} \sum_{j=0}^{\infty} |\theta_j| &< \infty, \\ \lim_{j_0 \to \infty} \sup_{\theta \in \Theta} \sum_{j=j_0}^{\infty} |\theta_j| &= 0. \end{aligned} \tag{6.2}$$

holds. This property characterizes the relatively $\|\cdot\|_1$-compact subsets of $\bar{\Theta}_{is}$, meaning the subsets with the property that every sequence θ^T, $T = 1, 2, \ldots$ in Θ has a convergent subsequence $\theta^{T'}$ such that $\left\|\theta^{T'} - \theta\right\|_1 \to 0$, for some θ with $\|\theta\|_1 < \infty$, see Theorem IV.8.9 and IV.8.3 of Dunford and Schwartz (1957). Under (6.2), coordinatewise convergence of a sequence θ^T in Θ, i.e. $\theta_j^T \to \theta_j$ for all $j \geq 0$, is equivalent to convergence in mean absolute sum norm, $\left\|\theta^T - \theta\right\|_1 \to 0$. So for these Θ, compactness in the sense of coordinatewise convergence is equivalent to $\|\cdot\|_1$-compactness.

Because

$$\max_{-\pi \leq \lambda \leq \pi} \left|\theta^T\left(e^{i\lambda}\right) - \theta\left(e^{i\lambda}\right)\right| \leq \left\|\theta^T - \theta\right\|_1, \tag{6.3}$$

it is clear that for any $\Theta \subseteq \Theta_{is}$, in order to guarantee that the compact set $\bar{\Theta}$ consisting of (Θ and) all such limits θ is a subset of Θ_{is}, it suffices to require the obviously necessary condition,

$$m_\Theta = \inf_{\substack{-\pi \leq \lambda \leq \pi \\ \theta \in \Theta}} \left|\sum_{j=0}^{\infty} \theta_j e^{ij\lambda}\right| > 0, \tag{6.4}$$

in addition to (6.2).

In Appendix B we derive the following alternative condition to (6.4):

Lemma 6.1. *If $\Theta \subseteq \Theta_{is}$ is such that (6.2) holds, then (6.4) holds if and only if $\sum_{j=0}^{\infty} \left|\tilde{\theta}_j\right|$ converges uniformly on $\tilde{\Theta} = \left\{\tilde{\theta} : \theta \in \Theta\right\}$.*

Remark 4 . When Θ defines a family of invertible ARMA (r, s) models, (6.2) is equivalent to the requirement that the zeros of the moving average polynomials belong to $\{|z| \geq 1 + \varepsilon\}$ for some $\varepsilon > 0$ (in which case $\sum_{j=0}^{\infty}(1 + \varepsilon_0)^j |\theta_j|$ converges uniformly on Θ for any $0 \leq \varepsilon_0 < \varepsilon$). Then (6.4) also holds if and only if the same is true of the zeros of the autoregressive polynomials, for some possibly different $\varepsilon > 0$. Model parameterization by θ avoids problems that afflict parameterization by means of ARMA coefficients at coefficient values that yield AR and MA polynomials with a common zero: see the Appendix of Pötscher (1991) for an elementary discussion of alternative parameterizations of ARMA models and their properties. Convergence of a sequence of ARMA coefficients implies coordinatewise convergence to the model's θ-parameters. The converse implication holds when the degrees of the AR and MA polynomials remain constant through the limit and the limiting AR and MA polynomials have no common zero.

6.2. A uniform convergence property

Partition $\Gamma_0^X(\theta) = \int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG_X(\lambda)$ analogously to (4.2), i.e.

$$\Gamma_0^X(\theta) = \begin{bmatrix} \Gamma_0^{MM}(\theta) & \Gamma_0^{MN}(\theta) \\ \Gamma_0^{NM}(\theta) & \Gamma_0^{NN}(\theta) \end{bmatrix},$$

with $\Gamma_0^{MM}(\theta) = \int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG^{MM}(\lambda)$, etc. For any $\theta \in \bar{\Theta}_{is}$, define

$$C^{NM}(\theta) = \Gamma_0^{NM}(\theta)\,\Gamma_0^{MM}(\theta)^{-1}. \tag{6.5}$$

(If $\theta \in \Theta_{is}$, then $\Gamma_0^X(\theta)$ is nonsingular and therefore also $\Gamma_0^{NM}(\theta)$, see Proposition 12.1(c). If $\theta(z)$ has a unit root, i.e. if $\theta \in \bar{\Theta}_{is} \setminus \Theta_{is}$, then $\Gamma_0^{MM}(\theta)$ can be singular, but we exclude nonsingularity below.) For the GLS estimators of α^M defined by

$$\alpha_T^M(\theta) = \sum_{t=1}^{T} Y_t[\theta]\,\xi_t^M[\theta]' \left(\sum_{t=1}^{T} \xi_t^M[\theta]\,\xi_t^M[\theta]'\right)^{-1}, \tag{6.6}$$

here is a uniform generalization of (4.9) whose proof in Appendix B follows from a uniform generalization of (3.18) obtained from Proposition 12.1 in Appendix A.

Theorem 6.2. *Suppose the assumptions of Section 3and (4.5) apply. Let $\Theta^* \subseteq \bar{\Theta}_{is}$ be such that (6.2) and*

$$\inf_{\theta^*\in\Theta^*} \lambda_{\min}\left(\Gamma_0^{MM}(\theta^*)\right) > 0 \tag{6.7}$$

hold. Then

$$\sup_{\theta^*\in\Theta^*} \left\| \left(\alpha_T^M(\theta^*) - \alpha^M\right) T^{-1/2} D_{\xi,M,T}^{-1} - \begin{bmatrix} A^N C^{NM}(\theta^*) & 0 \end{bmatrix} \right\| \to 0 \text{ a.s. [i.p.]}. \tag{6.8}$$

Also, the matrix function $C^{NM}(\theta^*)$ is continuous on Θ^* as well as bounded,

$$\sup_{\theta^*\in\Theta^*} \left\|C^{NM}(\theta^*)\right\| < \infty. \tag{6.9}$$

When $\Theta^* \subseteq \Theta_{is}$, the condition (6.7) is a consequence of (6.4).

7. Uniform Asymptotic Stationarity of Forecast Errors from a Misspecified Model with GLS Estimates of α^M

7.1. Forecasting with GLS estimates of α^M

We analyze errors of forecasts obtained with GLS estimates of α^M. If the OLS estimates of α^M are used instead, one need only replace any instances of $C^{NM}(\theta^*)$ with C^{NM} in the asymptotic formulas below to obtain the OLS results. For $h \geq 1$ and $1-h \leq t \leq T$ and any $\theta, \theta^* \in \Theta_{is}$, we consider the forecast functions

$$Y^M_{t+h|t}(\theta, \theta^*, T) = \alpha^M_T(\theta^*)\, \xi^M_{t+h} + y^M_{t+h|t}(\theta, \theta^*, T) \tag{7.1}$$

with

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

Thus

$$Y^M_{t+h|t}(\theta, \theta^*, T) = Y_t[\pi(h,\theta)] + \alpha^M_T(\theta^*)\, \xi^M_{t+h}[\eta(h,\theta)], \tag{7.2}$$

so that

$$\begin{aligned} Y_{t+h} - Y^M_{t+h|t}(\theta, \theta^*, T) &= Y_{t+h}[\eta(h,\theta)] - \alpha^M_T(\theta^*)\, \xi^M_{t+h}[\eta(h,\theta)] \\ &= y_{t+h}[\eta(h,\theta)] \\ &\quad + \left\{\alpha \xi_{t+h}[\eta(h,\theta)] - \alpha^M_T(\theta^*)\, \xi^M_{t+h}[\eta(h,\theta)]\right\}, \end{aligned} \tag{7.3}$$

where filtered quantities are truncated as in (5.7), e.g. $\xi^M_{t+h}[\eta(h,\theta)] = \sum_{j=0}^{t+h-1} \eta_l(h,\theta)\, \xi^M_{t+h-j}$, even when values X^M_{t+h-j} are known for $t+h-j < 0$. Partition the array (3.17) as

$$V_t(T) = \begin{bmatrix} y_t \\ T^{\frac{1}{2}} D_{M,T} X^M_t \\ X^N_t \\ T^{1/2} x^M_t \\ T^{1/2} x^N_t \end{bmatrix},\ 1 \leq t \leq T, \tag{7.4}$$

and partition $\alpha^M_T(\theta^*)$ as $\begin{bmatrix} A^M_T(\theta^*) & a^M_T(\theta^*) \end{bmatrix}$, corresponding to the partition $\begin{bmatrix} A^M & a^M \end{bmatrix}$ of α^M. Then with

$$\beta_T(\theta^*) = \left[1\ \left(A^M - A^M_T(\theta^*)\right) T^{-1/2} D^{-1}_{M,T}\ A^N\ T^{-1/2}\left(a^M - a^M_T(\theta^*)\right)\ T^{-1/2}\alpha^N\right], \tag{7.5}$$

the observable forecast errors are given by

$$Y_t - Y^M_{t|t-h}(\theta, \theta^*, T) = \beta_T(\theta^*)\, V_t[\eta(h,\theta)](T),\ 1 \leq t \leq T. \tag{7.6}$$

Defining $\beta(\theta^*) = \left[1\ -A^N C^{NM}(\theta^*)\ A^N\ 0\ 0\right]$, under the assumptions of Theorem 6.2, we have

$$\sup_{\theta^* \in \Theta^*} \|\beta(\theta^*)\| < \infty, \quad \sup_{\theta^* \in \Theta^*} \|\beta_T(\theta^*) - \beta(\theta^*)\| \to 0\ a.s.\ [i.p.]. \tag{7.7}$$

From these observations, Proposition 12.1 and Lemma 12.2 of Appendix A immediately yield the following theorem, showing that, uniformly on relative compact parameter sets, the limiting sample second moments of the forecast errors (7.3) are the same as those of the θ-model forecast errors of the A.S. array

$$y^M_t(\theta^*, T) = y_t + A^N\left(X^N_t - C^{NM}(\theta^*)\, T^{\frac{1}{2}} D_{M,T} X^M_t\right),\ 1 \leq t \leq T. \tag{7.8}$$

This interpretation arises from the fact that, with

$$B^N(\theta^*) = A^N\left[-C^{NM}(\theta^*)\ I_N\right], \tag{7.9}$$

the asymptotic spectral distribution function of this array is

$$G_{M,\theta^*}(\lambda) = G_y(\lambda) + B^N(\theta^*)\, G_X(\lambda)\, B^N(\theta^*)'. \tag{7.10}$$

For any $\Theta, \Theta^* \subseteq \bar{\Theta}_{is}$, as usual, $\Theta \times \Theta^*$ denotes the set $\{(\theta, \theta^*) : \theta \in \Theta, \theta^* \in \Theta^*\}$, and convergence on this set means coordinatewise convergence.

Theorem 7.1. *Suppose the assumptions of Theorem 6.2 hold for the model set $\Theta^* \subseteq \bar{\Theta}_{is}$. Then for any $\Theta \subseteq \bar{\Theta}_{is}$ for which (6.2) holds and any $h \geq 1$, the forecast error arrays $Y_t - Y^M_{t|t-h}(\theta, \theta^*, T), 1 \leq t \leq T$ are jointly continuous and jointly uniformly asymptotically stationary on $\Theta \times \Theta^*$ in the sense that, for any $h, l \geq 1$ and $k \geq 0$,*

$$\frac{1}{T}\sum_{t=1}^{T-k}\left(Y_{t+k} - Y^M_{t+k|t+k-h}(\theta, \theta^*, T)\right)\left(Y_t - Y^M_{t|t-l}(\theta, \theta^*, T)\right) \to \Gamma^M_k(h, l, \theta, \theta^*) \tag{7.11}$$

holds uniformly a.s. [i.p.] over $\Theta \times \Theta^*$, with

$$\Gamma^M_k(h, l, \theta, \theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda}\eta(h, \theta)\left(e^{i\lambda}\right)\eta(l, \theta)\left(e^{-i\lambda}\right) dG_{M,\theta^*}(\lambda). \tag{7.12}$$

The functions $\Gamma^M_k(h, l, \theta, \theta^*)$ are jointly continuous and bounded on $\Theta \times \Theta^*$.

We note the following consequence.

Corollary 7.2. *Suppose the assumptions of Section 3and (4.5) apply. Let $\Theta, \Theta^* \subseteq \bar{\Theta}_{is}$ be such that (6.2) holds. Let θ^T, $T \geq 1$ and $\theta^{*,T}$, $T \geq 1$ be random sequences in Θ and Θ^* respectively that are convergent a.s. [i.p.] to limits, θ^∞ and $\theta^{*,\infty}$ resp. with*

$$\lambda_{\min}\left(\Gamma^{MM}_0(\theta^{*,\infty})\right) > 0. \tag{7.13}$$

Then for any $k \geq 0$ and $h, l \geq 1$,

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

$$\to \Gamma^M_k(h, l, \theta^\infty, \theta^{*,\infty}) \text{ a.s. [i.p.]}.$$

In Theorem 10.1 of Section 10 below, we establish that standard estimation procedures can produce convergent sequences of parameter estimates of the sort assumed in this corollary.

Theorem 7.1 shows that the quantities $\Gamma^M_0(h, h, \theta, \theta^*)$ are of special interest because they describe limiting average squared forecast errors. With

$$\sigma_{hh}(\theta) = \int_{-\pi}^{\pi}\left|\eta(h, \theta)\left(e^{i\lambda}\right)\right|^2 dG_y(\lambda), \tag{7.14}$$

(7.10) yields the decomposition

$$\Gamma_0^M(h,h,\theta,\theta^*) = \sigma_{hh}(\theta) + B^N(\theta^*)\left[\int_{-\pi}^{\pi}\left|\eta(h,\theta)\left(e^{i\lambda}\right)\right|^2 dG_X(\lambda)\right] B^N(\theta^*)'. \tag{7.15}$$

By specializing the argument used to establish Theorem 7.1, $\sigma_{hh}(\theta)$ is seen to be the limiting average squared error of the h-step-ahead forecast of Y_t when X_t is known (up to coordinates with nonzero coefficients in α). Similarly, the second quantity on the right in (7.15) is seen to be the limit of the average of the squares of h-step-ahead forecast errors of the mean function and its its non-transitory component:

$$\begin{aligned}
& \lim_{T\to\infty}\frac{1}{T}\sum_{T=1}^{T-1}\left(\alpha\xi_{t+h}[\eta(h,\theta)] - \alpha_T^M(\theta^*)\,\xi_{t+h}^M[\eta(h,\theta)]\right)^2 \\
= \quad & \lim_{T\to\infty}\frac{1}{T}\sum_{T=1}^{T-1}\left(AX_{t+h}[\eta(h,\theta)] - A_T^M(\theta^*)\,X_{t+h}^M[\eta(h,\theta)]\right)^2 \\
= \quad & B^N(\theta^*)\left[\int_{-\pi}^{\pi}\left|\eta(h,\theta)\left(e^{i\lambda}\right)\right|^2 dG_X(\lambda)\right] B^N(\theta^*)' \;\; a.s.\ [i.p.].
\end{aligned} \tag{7.16}$$

We turn next to the case $h=1$ because of its role in GLS estimation.

Remark 5. It is easy to see that any subvector with the property that it is asymptotically orthogonal both to the remaining regressors of X_t^M and to X_t^N has no influence of the values of the asymptotic forecast error second moments (7.12). For example, the results of Subsection 4.3 show that when X_t^M has polynomial regressors $t^n,\ n\geq 0$, including the constant regressor $(n=0)$, and when the remaining regressors of X_t^M and those of X_t^N consist of regressors of the form (4.10) with $c_0=0$, then the subvector of X_t^M of all regressors of the form $t^n,\ n\geq 0$ has this property.

7.2. Advantages of GLS estimates when $h=1$

Because $\eta(1,\theta)\left(e^{i\lambda}\right) = \theta\left(e^{i\lambda}\right)$ when $h=1$, the limit in (7.16) for this case is

$$A^N\begin{bmatrix} -C^{NM}(\theta^*) & I_N\end{bmatrix}\Gamma_0^X(\theta)\begin{bmatrix} -C^{NM}(\theta^*) & I_N\end{bmatrix}' A^{N\prime}. \tag{7.17}$$

The matrix $C^{NM}(\theta)$ is the unique value of C minimizing $\begin{bmatrix} -C & I_N\end{bmatrix}\Gamma_0^X(\theta)\begin{bmatrix} -C & I_N\end{bmatrix}'$ in the ordering of symmetric matrices because

$$\begin{aligned}
\begin{bmatrix} -C & I_N\end{bmatrix}\Gamma_0^X(\theta)\begin{bmatrix} -C & I_N\end{bmatrix}' \quad = & \\
& \begin{bmatrix} -C^{NM}(\theta) & I_N\end{bmatrix}\Gamma_0^X(\theta)\begin{bmatrix} -C^{NM}(\theta) & I_N\end{bmatrix} \\
& + \left(C^{NM}(\theta) - C\right)\Gamma_0^{MM}(\theta)\left(C^{NM}(\theta) - C\right)'.
\end{aligned} \tag{7.18}$$

In particular,

$$\begin{bmatrix} -C^{NM}(\theta) & I_N\end{bmatrix}\Gamma_0^X(\theta)\begin{bmatrix} -C^{NM}(\theta) & I_N\end{bmatrix}' \leq \begin{bmatrix} -C & I_N\end{bmatrix}\Gamma_0^X(\theta)\begin{bmatrix} -C & I_N\end{bmatrix}', \tag{7.19}$$

with inequality holding for any $C \neq C^{NM}(\theta)$. For example, when $C^{NM}(\theta) \neq 0$, it follows from setting $C = 0$ in (7.19) that (7.16) is less than or equal to

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T} \sum_{T=1}^{T-1} \left(\alpha\xi_{t+1}[\theta] - \alpha^M \xi_{t+1}^M [\theta]\right)^2 &= \lim_{T\to\infty} \frac{1}{T} \sum_{T=1}^{T-1} \left(\alpha^N \xi_{t+1}^N [\theta]\right)^2 \\ &= A^N \Gamma_0^{NN}(\theta) A^{N\prime}\end{aligned}$$

with strict inequality holding for some values of A^N. That is, asymptotically, the "bias" of $\alpha_T^M(\theta)$ of as an estimator of α^M results in $\alpha_T^M(\theta)\xi_{t+1}^M[\theta]$ being a better estimator of $\alpha\xi_{t+1}[\theta]$ on average than $\alpha^M \xi_{t+1}^M[\theta]$, due to the contribution of the linear approximation $A^N C^{NM}(\theta) T^{\frac{1}{2}} D_{M,T} X_t^M$ of $A^N \xi_{t+1}^N[\theta]$.

GLS, done in conjunction with one-step-ahead squared forecast error minimization over a compact set, results in minimization of (1.2), an expression that coincides with the l.h.s. of (7.11) for $\theta^* = \theta, k = 0, h = l = 1$. It follows from Theorem 7.1 that this procedure, which is known as conditional maximum likelihood estimation, minimizes $\Gamma_0^M(1,1,\theta,\theta)$ asymptotically, as does unconditional (exact) maximum likelihood estimation, which is discussed in later sections. A specific result is formulated in Theorem 10.1. If a different choice of θ^* is used, say $\theta^* = (1,0,0,\dots)$ to OLS estimates, then this Theorem shows that these estimation procedures lead to minimization of $\Gamma_0^M(1,1,\theta,\theta^*)$. Returning to (7.17), note that if $C^{NM}(\theta^*) \neq C^{NM}(\theta)$, then (7.19) with $C = C^{NM}(\theta^*)$ yields

$$B^N(\theta)\Gamma_0^X(\theta)B^N(\theta)' < B^N(\theta^*)\Gamma_0^X(\theta)B^N(\theta^*)' \tag{7.20}$$

for some values of A^N. Our next result, whose proof is given in Appendix B, is formulated to accommodate the fact that minimizers of $\Gamma_0^M(1,1,\theta,\theta)$ and $\Gamma_0^M(1,1,\theta,\theta^*)$ with respect to θ need not be unique when the model class Θ is incorrect in the sense that it cannot model the asymptotic autocovariance sequence γ_k^y, i.e. when there is no $\theta \in \Theta$ such that $\gamma_k^y = \sigma^2\gamma_k(\theta)$ holds for all $k \geq 0$ and some σ^2, where

$$\gamma_k(\theta) = \int_{-\pi}^{\pi} \frac{e^{-ik\lambda}}{2\pi} \left|\sum_{j=0}^{\infty} \theta_j e^{ij\lambda}\right|^{-2} d\theta,\ k \geq 0. \tag{7.21}$$

Corollary 7.3. *Let $\Theta \subset \bar{\Theta}_{is}$ be a compact set over which $\sum_{j=0}^{\infty}|\theta_j|$ converges uniformly and $\inf_{\theta\in\Theta}\lambda_{\min}\left(\Gamma_0^{MM}(\theta)\right) > 0$ holds. Let $\bar{\theta}$ denote a minimizer of $\Gamma_0^M(1,1,\theta,\theta)$ over Θ. Then $\Gamma_0^M\left(1,1,\bar{\theta},\bar{\theta}\right) = \min_{\theta,\theta^*\in\Theta}\Gamma_0^M(1,1,\theta,\theta^*)$. More precisely, for each $\theta^* \in \Theta$, let $\bar{\theta}^*$ denote a minimizer of $\Gamma_0^M(1,1,\theta,\theta^*)$ over Θ. Then*

$$\Gamma_0^M\left(1,1,\bar{\theta},\bar{\theta}\right) \leq \Gamma_0^M\left(1,1,\bar{\theta}^*,\theta^*\right), \tag{7.22}$$

with strict inequality holding when $\Gamma_0^M\left(1,1,\bar{\theta},\bar{\theta}\right) < \Gamma_0^M\left(1,1,\bar{\theta}^*,\bar{\theta}^*\right)$ or when

$$C^{NM}(\theta^*) \neq C^{NM}\left(\bar{\theta}^*\right), \tag{7.23}$$

and A^N is such that (7.20) holds for $\theta = \bar{\theta}^*$.

The interesting choice of θ^* in the Corollary is $\theta^* = (1, 0, 0, \dots)$, which yields $\alpha_T^M(\theta^*) = \alpha_T^M$, the OLS estimate (3.14). (Model sets for the autocovariance structure of y_t usually include the white noise model $(1, 0, 0, \dots)$ as a degenerate case). With this choice of θ^* and with $\bar{\theta}$ and $\bar{\theta}^*$ as in the Corollary, Theorem 10.1 below will show that (conditional or exact) maximum likelihood estimation over Θ leads to limiting average squared forecast errors with the value $\Gamma_0^M\left(1, 1, \bar{\theta}, \bar{\theta}\right)$ if GLS is used, but with the value $\Gamma_0^M\left(1, 1, \bar{\theta}^*, \theta^*\right)$ if OLS is used. Thus the Corollary yields the following optimality property of GLS: In conjunction with maximum likelihood estimation of θ and in the asymptotic sense being considered, *OLS estimation is never better than GLS estimation for one-step-ahead forecasting. When the nontransitory component X_t of the mean function is misspecified, OLS is typically worse.*

Indeed, it seems likely that the Corollary's conditions, $\Gamma_0^M\left(1, 1, \bar{\theta}, \bar{\theta}\right) < \Gamma_0^M\left(1, 1, \bar{\theta}^*, \bar{\theta}^*\right)$ and (7.23), both hold except in quite special situations such as when $\Gamma_0^M(1, 1, \theta, \theta^*)$ does not depend on θ^*, which happens for example, when $X_t^N = 1$ and $X_t^M = t^m$ for some $m > 0$, or when the regressors X_t^M and X_t^N are asymptotically orthogonal. As we illustrate in the next subsection, $\bar{\theta}^*$ is easier to determine than $\bar{\theta}$, so it is simpler to provide examples of (7.23).

Extensions of the Corollary for the case of h-step ahead forecasting with $h > 1$ can be obtained by using $\eta(h, \theta)(B)$ in place of $\theta(B)$ in the definition of the GLS coefficient estimate of α^M, see Remark 6. However, these extensions do not carry over to h-step-ahead forecasts of nonstationary models like regARIMA models and their generalizations, because the h-step ahead forecast errors of such models involve forecast errors at all lags $1 \leq j \leq h$ of the A.S. "differenced" data, see (8.4) below.

For certain situations in which y_t is *i.i.d.* and $\xi_t = X_t$ is a realization of a stationary time series, so that $D_{X,T} = T^{-1/2}$, Thursby (1987) considers some theoretical examples of misspecified regressors with a focus on determining which of OLS or GLS has smaller asymptotic bias for estimating an individual coefficient of A^M, either outcome being possible in general. Corollary 7.3 shows that, for either outcome, the bias of GLS is better for one-step-head forecasting asymptotically.

Remark 6. The basic result (7.22) of the Corollary can be generalized considerably, for example, to accommodate noninvertible models, i.e. models not in Θ_{is} whose spectral densities (which are proportional to $\left|\tilde{\theta}\left(e^{i\lambda}\right)\right|^2 = \left|\theta\left(e^{i\lambda}\right)\right|^{-2}$) are zero for one or more values of λ. All that is required is a set of models or, more generally, filters Θ such that, for all $\theta \in \Theta$,

$$\int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG_y(\lambda) < \infty$$

and

$$tr\left[\int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG_X(\lambda)\right] < \infty$$

hold. Then, for

$$F(\theta, \theta^*) = \int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG_y(\lambda) + B^N(\theta^*)\left[\int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG_X(\lambda)\right] B^N(\theta^*)',$$

a straightforward modification of the proof of (7.22) yields that, for every $\theta^* \in \Theta$,

$$\inf_{\theta \in \Theta} F(\theta, \theta) \leq \inf_{\theta \in \Theta} F(\theta, \theta^*). \tag{7.24}$$

When minimizing values $\bar{\theta}$ and $\bar{\theta}^*$ exist as in the Corollary, then $F\left(\bar{\theta},\bar{\theta}\right) < F_0^M\left(\bar{\theta}^*,\bar{\theta}^*\right)$ yields strict inequality and (7.23) does also if A^N is such that (7.20) holds for $\theta = \bar{\theta}^*$. The inequality (7.24) is valid when $\Gamma_0^{MM}(\theta) = \int_{-\pi}^{\pi} \left|\theta\left(e^{i\lambda}\right)\right|^2 dG^{MM}(\lambda)$ is singular, because (7.18) still holds if $\Gamma_0^{MM}(\theta)^{-1}$ in (4.8) is interpreted as a generalized inverse. However, Theorems 6.2 and 7.1 have not been established for this case.

7.3. The special case of AR(1) models, $h = 1$ and $\dim X_t^N = 1$

We now present some illustrative formulas and examples related to the minimum asymptotic average squared forecast errors $\Gamma_0^M\left(1,1,\bar{\theta},\bar{\theta}\right)$ and $\Gamma_0^M\left(1,1,\bar{\theta}^*,\theta^*\right)$ of (7.22) for $\theta^* = (1,0,0,\dots)$, i.e. when $C^{NM}(\theta^*) = C^{NM}$. These are for the case $\dim X_t^N = 1$ when first order autoregressive modeling, with $\theta = (1,-\phi,0,0,\dots)$, is used for the regression error series y_t^M in (4.3).

We start with the simple situation in which X_t^M and X_t^N are asymptotically orthogonal, so $C^{NM}(\theta) = 0$ for all θ. In this case, for all ϕ, $\Gamma_0^M(1,1,\theta,\theta^*) = \Gamma_0^M(1,1,\theta,\theta)$ and

$$\begin{aligned}\Gamma_0^M(1,1,\theta,\theta) &= \int_{-\pi}^{\pi}\left|1-\phi e^{i\lambda}\right|^2 dG_y(\lambda) + \left(A^N\right)^2\int_{-\pi}^{\pi}\left|1-\phi e^{i\lambda}\right|^2 dG^{NN}(\lambda) \quad (7.25)\\ &= \left(1+\phi^2\right)\left\{\gamma_0^y + \left(A^N\right)^2\Gamma_0^{NN}\right\} - 2\phi\left\{\gamma_1^y + \left(A^N\right)^2\Gamma_1^{NN}\right\}. \quad (7.26)\end{aligned}$$

This is minimized by $\bar{\theta} = \left(1,-\bar{\phi},0,0,\dots\right)$, with

$$\bar{\phi} = \frac{\gamma_1^y + \left(A^N\right)^2\Gamma_1^{NN}}{\gamma_0^y + \left(A^N\right)^2\Gamma_0^{NN}}.$$

Define $\rho_1^y = \gamma_1^y/\gamma_0^y$ and $\rho_1^{NN} = \Gamma_1^{NN}/\Gamma_0^{NN}$. Rewriting $1-\phi e^{i\lambda}$ in the first integral in (7.25) as $\left(1-\rho_1^y e^{i\lambda}\right) + \left(\rho_1^y - \phi\right)e^{i\lambda}$ and as $\left(1-\rho_1^{NN}\phi e^{i\lambda}\right) + \left(\rho_1^{NN}-\phi\right)e^{i\lambda}$, one obtains

$$\begin{aligned}\Gamma_0^M\left(1,1,\bar{\theta},\bar{\theta}\right) &= \gamma_0^y\left(1-\left(\rho_1^y\right)^2\right)\\ &\quad + \gamma_0^y\left(\rho_1^y-\bar{\phi}\right)^2 + \left(A^N\right)^2\Gamma_0^{NN}\left\{\left(1-\left(\rho_1^{NN}\right)^2\right) + \left(\rho_1^{NN}-\bar{\phi}\right)^2\right\}. \quad (7.27)\end{aligned}$$

The lower expression (7.27) is the amount of increase in asymptotic average square forecast error due to misspecification of the regressor.

When X_t^M and X_t^N are not asymptotically orthogonal and $\dim X_t = 2$, i.e. $\dim X_t^M = 1$, then the formula for $\Gamma_0^M(1,1,\theta,\theta)$ is obtained by replacing Γ_k^{NN} in (7.26) by
$\Gamma_k^{NN} + C^{NM}(\theta)^2\Gamma_k^{MM} - C^{NM}(\theta)\left(\Gamma_k^{NM} + \Gamma_{-k}^{NM}\right)$, $k = 0,1$, with

$$C^{NM}(\theta) = \frac{\Gamma_0^{NM} - \phi\Gamma_1^{NM}}{\Gamma_0^{MM} - \phi\Gamma_1^{MM}}.$$

Thus, in this case $\Gamma_0^M(1,1,\theta,\theta)$ is a rational function of ϕ whose minimizing value $\bar{\phi}$ is a zero of a polynomial in ϕ of degree five and does not have a simple closed-form formula. Therefore, to obtain examples for which (7.23) holds for this case, in place of a generalization of (7.27), we will use a formula for the minimizer $\bar{\theta}^* = \left(1,-\bar{\phi}^*,0,0,\dots\right)$ of $\Gamma_0^M(1,1,\theta,\theta^*)$.

This $\bar{\phi}^*$ is the value of ϕ minimizing

$$\int_{-\pi}^{\pi} \left|1 - \phi e^{i\lambda}\right|^2 dG_{M,\theta^*}(\lambda).$$

Therefore $\bar{\phi}^*$ is the lag one asymptotic "autocorrelation" of

$$\breve{y}_t^M = y_t + A^N \left(X_t^N - T^{1/2} D_{M,T} C^{NM} X_t^M\right),$$

given by

$$\bar{\phi}^* = \frac{\gamma_1^y + \left(A^N\right)^2 \left\{\Gamma_1^{NN} + \left(C^{NM}\right)^2 \Gamma_1^{MM} - C^{NM}\left(\Gamma_1^{NM} + \Gamma_{-1}^{NM}\right)\right\}}{\gamma_0^y + \left(A^N\right)^2 \left\{\Gamma_0^{NN} + \left(C^{NM}\right)^2 \Gamma_0^{MM} - 2C^{NM}\Gamma_0^{NM}\right\}}. \tag{7.28}$$

Except possibly at a single value of A^N, this optimal $\bar{\phi}^*$ will be non-zero, i.e. be such that $\bar{\theta}^* \neq \theta^*$, when $\gamma_1^y \neq 0$ or when

$$\Delta^{NM} = \Gamma_1^{NN} + \left(C^{NM}\right)^2 \Gamma_1^{MM} - C^{NM}\left(\Gamma_1^{NM} + \Gamma_{-1}^{NM}\right) \neq 0. \tag{7.29}$$

The property $\bar{\phi}^* \neq 0$ yields (7.23) when the value $C^{NM}(\theta^*)$ is unique,

$$C^{NM}(\theta) \neq C^{NM}(\theta^*), \; \theta \neq \theta^*. \tag{7.30}$$

To cover the two most common kinds of regressors, we give two examples of (7.23), one in which X_t is periodic and the other in which X_t is a realization of a linear stationary process. The first example is motivated by the idea that if X_t^M inadequately represents a periodic effect, then X_t^N will include one or more compensating regressors with the same period. Consider the simple bivariate X_t with period 4 given by

$$X_t = \begin{bmatrix} X_t^M & X_t^N \end{bmatrix}' = b \cos\frac{\pi}{2}t + c(-1)^t,$$

with linearly independent coefficient vectors $b = \begin{bmatrix} b^M & b^N \end{bmatrix}'$ and $c = \begin{bmatrix} c^M & c^N \end{bmatrix}'$ such that $b^M, b^N \neq 0$. We have

$$\Gamma_k^X = \frac{1}{2} bb' \cos\frac{\pi}{2}k + cc'(-1)^k, \; k = 0, 1, \ldots,$$

so

$$C^{NM} = \frac{\frac{1}{2} b^N b^M + c^N c^M}{\frac{1}{2}\left(b^M\right)^2 + \left(c^M\right)^2}.$$

Since $\Gamma_0^X(\theta) = \Gamma_0^X - \phi\Gamma_1^X$,

$$C^{NM}(\theta) = \frac{\frac{1}{2}\left(1+\phi^2\right) b^N b^M + (1+\phi)^2 c^N c^M}{\frac{1}{2}\left(1+\phi^2\right)\left(b^M\right)^2 + (1+\phi)^2 \left(c^M\right)^2}.$$

Because $b^M, b^N \neq 0$ and c is not scalar multiple of b, we have $C^{NM}(\theta) \neq C^{NM}$ when $\phi \neq 0$, i.e. when $\theta \neq \theta^*$ verifying (7.30). Thus (7.23) follows from $\Delta^{NM} = -\left(c^N - C^{NM} c^M\right)^2 \neq 0$.

For the second example, consider the case in which X_t^N is scalar with $\Gamma_k^{NN} = 0$ for some $k \geq 1$ and $X_t^M = X_{t-k}^N$, with the result that $\Gamma_0^{NM} = 0$ and therefore $C^{NM}(\theta^*) = C^{NM} = 0$. We seek an example with $C^{NM}\left(\bar{\theta}^*\right) \neq 0$, or equivalently, $\Gamma_0^{NM}\left(\bar{\theta}^*\right) = \Gamma_k^{NN}\left(\bar{\theta}^*\right) \neq 0$. The formula (7.28) yields

$$\bar{\phi}^* = \frac{\gamma_1^y + \left(A^N\right)^2 \Gamma_1^{NN}}{\gamma_0^y + \left(A^N\right)^2 \Gamma_0^{NN}}.$$

This is nonzero when $\gamma_1^y + \left(A^N\right)^2 \Gamma_1^{NN} \neq 0$. Since

$$\begin{aligned}
\Gamma_k^{NN}(\theta) &= \int_{-\pi}^{\pi} e^{-ik\lambda} \left|1 - \phi e^{i\lambda}\right|^2 dG^{NN}(\lambda) = \left(1 + \phi^2\right) \Gamma_k^{NN} - \phi\left(\Gamma_{k+1}^{NN} + \Gamma_{k-1}^{NN}\right) \\
&= -\phi\left(\Gamma_{k+1}^{NN} + \Gamma_{k-1}^{NN}\right),
\end{aligned}$$

it follows that $\Gamma_k^{NN}\left(\bar{\theta}^*\right) \neq 0$ whenever $\Gamma_{k+1}^{NN} + \Gamma_{k-1}^{NN} \neq 0$. For example, consider the case in which $\gamma_1^y \neq 0$ and X_t^N is a realization of a second order autoregression satisfying $X_t^N - \phi_2 X_{t-2}^N = \varepsilon_t$ with $0 < |\phi_2| < 1$ for some i.i.d. white noise process ε_t, then $\bar{\phi}^* = \gamma_1^y / \left(\gamma_0^y + \left(A^N\right)^2 \Gamma_0^{NN}\right) \neq 0$. For $k = 2m + 1$ with $m \geq 0$, we have $\Gamma_k^{NN} = 0$ and $\Gamma_{k+1}^{NN} + \Gamma_{k-1}^{NN} = (1 + \phi_2)\phi_2^m \Gamma_0^{NN} \neq 0$, so $\Gamma_k^{NN}\left(\bar{\theta}^*\right) \neq 0$.

8. Joint Asymptotic Stationarity of Forecast Errors from Misspecified Nonstationary Models such as regARIMA Models

Many time series require linear transformations such as differencing operations before they have properties like those we have assumed for (1.1). We now present an extension of Theorem 7.1 for models for such series, e.g. invertible regARIMA models.

Suppose we have observations of the form

$$W_t = \alpha \zeta_t + w_t \ \ (t \geq -d + 1), \tag{8.1}$$

where w_t satisfies a d-th order difference equation with $d \geq 1$,

$$w_t + \delta_1 w_{t-1} + \cdots + \delta_d w_{t-d} = y_t \ \ (t \geq 1),$$

in which y_t together with

$$\xi_t = \zeta_t + \delta_1 \zeta_{t-1} + \cdots + \delta_d \zeta_{t-d} \ \ (t \geq 1)$$

satisfy the assumptions of Subsection **??**. Observe that

$$Y_t = W_t + \delta_1 W_{t-1} + \cdots \delta_d W_{t-d}, \ (t \geq 1) \tag{8.2}$$

has the formula $Y_t = \alpha \xi_t + y_t$ as in (1.1).

With $\delta_0 = 1$ and $\delta(z) = \sum_{j=0}^{d} \delta_j z^j$, define $\tilde{\delta}_0 = 1$ and $\tilde{\delta}_j = \sum_{i=0}^{j-1} \tilde{\delta}_i \delta_{j-i}, \ j = 1, 2, \ldots$. It is not difficult to verify from (8.2), see Bell (1984, p. 650), 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

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

Therefore, when $t \geq 1$, given forecasts of $Y_{t+h-j},\ 0 \leq j \leq h-1$, say $Y^M_{t+h-j|t}(\theta, \theta^*, T)$, $0 \leq j \leq h-1$ defined as in (7.1), we can define the forecast of W_{t+h} to be

$$W^M_{t+h|t}(\theta, \theta^*, T) = \sum_{j=0}^{h-1} \tilde{\delta}_j Y^M_{t+h-j|t}(\theta, \theta^*, T) + \sum_{j=0}^{d-1} c_{j,h} W_{t-j}. \tag{8.3}$$

For the forecast errors, we then have, for $t \geq 1$,

$$W_t - W^M_{t|t-h}(\theta, \theta^*, T) = \sum_{j=0}^{h-1} \tilde{\delta}_j \left(Y_{t-j} - Y^M_{t-j|t-j-h}(\theta, \theta^*, T) \right). \tag{8.4}$$

Consequently, these forecast errors inherit joint and uniform asymptotic stationarity from the corresponding properties of the $Y_{t-j} - Y^M_{t-j|t-j-h}(\theta, \theta^*, T),\ 0 \leq j \leq h-1$. Thus, with

$$\begin{aligned} \sigma^{M,\delta}_{hh}(\theta, \theta^*) &= \int_{-\pi}^{\pi} \left| \sum_{j=0}^{h-1} e^{ij\lambda} \tilde{\delta}_j \eta(h-j, \theta)\left(e^{i\lambda}\right) \right|^2 dG_{M,\theta^*}(\lambda) \\ &= \sum_{j,k=0}^{h-1} \tilde{\delta}_j \tilde{\delta}_k \Gamma^M_{k-j}(h-j, h-k, \theta, \theta^*), \end{aligned} \tag{8.5}$$

where $G_{M,\theta^*}(\lambda)$ is defined as in (7.10) and $\Gamma^M_{k-j}(h-j, h-k, \theta, \theta^*)$ as in (7.12), an immediate consequence of Theorem 7.1 is

Theorem 8.1. *Suppose $\Theta, \Theta^* \subseteq \Theta_{is}$ are such that the assumptions of Theorem 7.1 hold. Then the forecast error sequences $W_t - W^M_{t|t-h}(\theta, \theta^*, T)$ $t \geq 1$ with $h \geq 1$ are jointly uniformly asymptotically stationary on $\Theta \times \Theta^*$. In particular,*

$$\sup_{\theta \in \Theta, \theta^* \in \Theta^*} \left| \frac{1}{T} \sum_{t=1}^{T} \left(W_t - W^M_{t|t-h}(\theta, \theta^*, T) \right)^2 - \sigma^{M,\delta}_{hh}(\theta, \theta^*) \right| \to 0 \ \ a.s.\ [i.p.]. \tag{8.6}$$

The limit function $\sigma^{M,\delta}_{hh}(\theta, \theta^*)$ is bounded and continuous on $\Theta \times \Theta^*$.

9. Finite-Past Predictors and GLS

Our results above for the truncated predictors $y_{t+h|t}(\theta) = y_t[\pi(h, \theta)]$ of y_{t+h} will serve as stepping stones for deriving results for the more commonly used *finite-past* predictors. These are defined by

$$\hat{y}_{t+h|t}(\theta) = \begin{cases} \sum_{j=0}^{t-1} \pi_{t,j}(h, \theta) y_{t-j} & , 1 \leq t \leq T \\ 0 & , 1-h \leq t \leq 0 \end{cases}, \tag{9.1}$$

using as coefficient vector $[\pi_{t,j}(h, \theta)]_{1 \leq j \leq t-1}$ the solution of

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

where $\gamma_k(\theta)$ is as in (7.21), see, for example, Newton and Pagano (1983). Hence, if we define $\eta_{t,j}(h,\theta), 0 \leq j \leq t-1$, by

$$\eta_{t,j}(h,\theta) = \begin{cases} 1 & ,j=0 \\ 0 & ,1 \leq j \leq h-1 \\ -\pi_{t-h,j-h}(h,\theta) & ,h \leq j \leq t-1 \end{cases}$$

for $t \geq h+1$, and by

$$\eta_{t,j}(h,\theta) = \begin{cases} 1 & ,j=0 \\ 0 & ,1 \leq j \leq t-1 \end{cases},$$

for $1 \leq t \leq h$, then the observable prediction error values from data $y_t, 1 \leq t \leq T$ have the formulas

$$y_t - \hat{y}_{t|t-h}(\theta) = \sum_{j=0}^{t-1} \eta_{t,j}(h,\theta) y_{t-j},\ 1 \leq t \leq T.$$

These prediction errors and their coefficients are sometimes normalized by dividing the coefficients by the square root of the mean square prediction error quantity (9.2), particularly in the case $h = 1$ used for GLS estimation as in Amemiya (1973) and for maximum likelihood estimation based on the functions (9.12) below. The normalization is done to obtain prediction errors with constant variance (when the θ-model is correct).

$$\begin{aligned} u_{t|t-h}(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \sum_{j=0}^{t-1} \eta_{t,j}(h,\theta) e^{-ij\lambda} \right|^2 \left| \sum_{j=0}^{\infty} \theta_j e^{ij\lambda} \right|^{-2} d\lambda \\ &= \sum_{j=0}^{t-1} \eta_{t,j}(h,\theta)\gamma_j(\theta). \end{aligned} \quad (9.2)$$

Thus the normalized coefficients are

$$\breve{\eta}_{t,j}(h,\theta) = \eta_{t,j}(h,\theta)/u_{t|t-h}^{1/2}(\theta),\ j = 0,1,\ldots,t-1. \quad (9.3)$$

(Note that $u_{t|t-h}^{1/2}(\theta) = \gamma_0(\theta)^{1/2}$ for $1 \leq t \leq h$.) By a straightforward modification of the proof of (5.17) of FPW (2003), it can be shown that the sequence $u_{t|t-h}(\theta)$ is bounded above by $\gamma_0(\theta) = \sum_{i=0}^{\infty} \tilde{\theta}_i^2$ and decreases uniformly as $t \to \infty$ to $\sum_{j=0}^{h-1} \tilde{\theta}_j^2$ on the subsets $\Theta \subseteq \Theta_{is}$ considered in (b) of Proposition 9.1 below. Consequently, to convert limiting forecast error second moment formulas obtained without the normalization to formulas that apply when the normalized coefficients are used, the functions $\Gamma_k^M(h,l,\theta,\theta^*)$ with appearing in the limit must be divided by $\left\{ \sum_{j=0}^{h-1} \tilde{\theta}_j^2 \sum_{j=0}^{l-1} \tilde{\theta}_j^2 \right\}^{1/2}$. Note this is equal to one when $h = l = 1$. For simplicity, we shall use the normalization only in our definitions of the finite-past GLS coefficient vector and the Gaussian log-likelihood function both of which involve $h = 1$ exclusively.

With $Y_t(\theta^*) = \sum_{j=0}^{t-1} \breve{\eta}_{t,j}(1,\theta^*) Y_{t-j}$ and $\xi_t^M(\theta^*) = \sum_{j=0}^{t-1} \breve{\eta}_{t,j}(1,\theta^*) \xi_{t-j}^M$, the GLS estimator of α^M is defined by

$$\hat{\alpha}_T^M(\theta^*) = \sum_{t=1}^{T} Y_t(\theta^*) \xi_t^M(\theta^*)' \left(\sum_{t=1}^{T} \xi_t^M(\theta^*) \xi_t^M(\theta^*)' \right)^{-1}, \quad (9.4)$$

and the finite-past variant of the predictor $Y_{t+h|t}(\theta, \theta^*, T)$ of (7.1) is defined as

$$\hat{Y}^M_{t+h|t}(\theta, \theta^*, T) = \sum_{j=0}^{t-1} \pi_{t+h,j}(h, \theta) Y_{t-j} + \hat{\alpha}^M_T(\theta^*) \sum_{j=0}^{t+h-1} \eta_{t+h,j}(h, \theta) \xi^M_{t+h-j}, \tag{9.5}$$

see (7.2). With $U_t(T)$ as in (3.19), set

$$U^h_t(\theta, T) = \sum_{j=0}^{t-1} \eta_{t,j}(h, \theta) U_{t-j}(T), \ 1 \leq t \leq T,$$

and

$$V^h_t(\theta, T) = \begin{bmatrix} U^h_t(\theta, T) \\ T^{1/2} \sum_{j=0}^{t-1} \eta_{t,j}(h, \theta) x_{t-j} \end{bmatrix}, \ 1 \leq t \leq T.$$

Partition $\hat{\alpha}^M_T(\theta^*)$ as $\begin{bmatrix} \hat{A}^M_T(\theta^*) & \hat{a}^M_T(\theta^*) \end{bmatrix}$ corresponding to the partition $\begin{bmatrix} A^M & a^M \end{bmatrix}$ of α^M, and define

$$\hat{\beta}_T(\theta^*) = \left[1 \ \left(A^M - \hat{A}^M_T(\theta^*)\right) T^{-1/2} D^{-1}_{M,T} \ A^N \ T^{-1/2} \left(a^M - \hat{a}^M_T(\theta^*)\right) \ T^{-1/2} \alpha^N\right].$$

Then the observable forecast errors are given by

$$Y_t - \hat{Y}^M_{t|t-h}(\theta, \theta^*, T) = \hat{\beta}_T(\theta^*) V^h_t(\theta, T), \ 1 \leq t \leq T. \tag{9.6}$$

When no transitory regressors x_t are present in ξ_t, so that $V^h_t(\theta, T)$ in (9.6) is replaced by $U^h_t(\theta, T)$ and $\hat{\beta}_T(\theta^*)$ by $\left[1 \ \left(A^M - \hat{A}^M_T(\theta^*)\right) T^{-1/2} D^{-1}_{M,T} \ A^N\right]$, then the fact that the A.S. array $U_t(T)$ has both negligibility properties (2.8) and (2.7) has the consequence that the "finite-past" generalizations of Theorems 7.1 and 8.1 follow by arguments analogous to those given in Subsection 7.1 except that use is made of Proposition 5.2 and part (a) of Proposition 2.1 of FPW (2003) together with the assumptions that (6.4) holds and that, for some $\varepsilon > 0$, the sums $\sum_{j=0}^{\infty} \left(1 + j^{\frac{1}{2}+\varepsilon}\right) |\theta_j|$ and $\sum_{j=0}^{\infty} \left(1 + j^{\frac{1}{2}+\varepsilon}\right) \left|\theta^*_j\right|$ converge uniformly on Θ and Θ^* respectively. When transitory regressors are ‘present, the simplest way to achieve generalizations seems to be to require a bit more of the model sets in order that (b) and (c) of the following result, proved in Appendix B, can be applied. In (d), typical choices of $\nu(j)$ are $\nu(j) = 1 + j^{\mu}$ for some $\mu \geq 0$ or (in the case of invertible ARMA models) $\nu(j) = \tau^{-j}$ with $0 < \tau < 1$.

Proposition 9.1. *(a) Let T_0 be the smallest value of T for which $\sum_{t=1}^{T} \xi_t \xi_t^{'} > 0$. For any $\theta \in \Theta_{is}$, the one-step-ahead prediction errors $\xi_t(\theta) = \sum_{j=0}^{t-1} \breve{\eta}_{t,j}(1, \theta) \xi_t$ have the property that*

$$0 < \min_{-\pi \leq \lambda \leq \pi} \left| \sum_{j=0}^{\infty} \theta_j e^{ij\lambda} \right|^2 \sum_{t=1}^{T} \xi_t \xi_t \leq \sum_{t=1}^{T} \xi_t(\theta) \xi_t(\theta)' \leq \max_{-\pi \leq \lambda \leq \pi} \left| \sum_{j=0}^{\infty} \theta_j e^{ij\lambda} \right|^2 \sum_{t=1}^{T} \xi_t \xi_t \tag{9.7}$$

holds for every $T \geq T_0$.
(b) If $\Theta \subseteq \Theta_{is}$ is any subset of Θ_{is} on which (6.4) holds and $\sum_{j=0}^{\infty} (1 + j) |\theta_j|$ converges uniformly,

then $\sum_{j=1}^{\infty}\sum_{j=0}^{t-1}\left|\eta_{t,j}(h,\theta)-\eta_j(h,\theta)\right|$ *converges uniformly on* Θ *for each* $h\geq 1$*. Consequently, given regressors* x_t *satisfying* $\sum_{t=1}^{\infty}\|x_t\|<\infty$*, the sums* $\sum_{t=1}^{\infty}\left\|\sum_{j=0}^{t-1}\eta_{t,j}(h,\theta)\,x_{t-j}\right\|$ *converge uniformly on* Θ*.*
(c) With Θ *as in (b), for* $x_t(\theta)=\sum_{j=0}^{t-1}\breve{\eta}_{t,j}(1,\theta)\,x_{t-j}$ *and* $y_t^M(\theta)=\sum_{j=0}^{t-1}\breve{\eta}_{t,j}(1,\theta)\,y_{t-j}^M$*, with* y_{t-j}^M *defined by (4.3), we have*

$$\sup_{\theta\in\Theta}\left\|T^{-1/2}\sum_{t=1}^{T}y_t^M(\theta)\,x_t(\theta)'\left(\sum_{t=1}^{T}x_t(\theta)\,x_t(\theta)'\right)^{-1}\right\|\rightarrow 0\ a.s.\ [i.p.]. \tag{9.8}$$

(d) Consider any nondecreasing sequence of weights $1\leq\nu(1)\leq\nu(2)\leq\cdots$ *with the properties that* $\nu(t)\leq\nu(j)\,\nu(t-j)$ *holds whenever* $0\leq j\leq t$*. Let* x_t *be such that*

$$\sup_{t\geq 1}\nu(t)\,\|x_t\|<\infty.$$

Then, on any subset $\Theta\subseteq\Theta_{is}$ *on which* $\sum_{j=0}^{\infty}\nu(j)\,|\theta_j|$ *converges uniformly and (6.4) holds,* $\nu(t)\,x_t(\theta)$ *is uniformly bounded,*

$$\sup_{\theta\in\Theta,t\geq 1}\nu(t)\,\|x_t(\theta)\|<\infty. \tag{9.9}$$

The "finite-past" analogues of (8.3) and (8.4) are

$$\hat{W}_{t+h|t}^M(\theta,\theta^*,T)=\sum_{j=0}^{h-1}\tilde{\delta}_j\hat{Y}_{t+h-j|t}^M(\theta,\theta^*,T)+\sum_{j=0}^{d-1}c_{j,h}W_{t-j}. \tag{9.10}$$

and

$$W_t-\hat{W}_{t|t-h}^M(\theta,\theta^*,T)=\sum_{j=0}^{h-1}\tilde{\delta}_j\left(Y_{t-j}-\hat{Y}_{t-j|t-h}^M(\theta,\theta^*,T)\right)$$

$$=\sum_{j=0}^{h-1}\tilde{\delta}_j\left(\sum_{j=0}^{t-1}\eta_{t,j}(h,\theta)\,Y_{t-j}+\hat{\alpha}_T^M(\theta^*)\sum_{j=0}^{t-1}\eta_{t,j}(h,\theta)\,\xi_{t-j}^M\right). \tag{9.11}$$

We also consider the Gaussian log-likelihood function for the observations W_t, $1\leq t\leq T$, the autocovariance structure determined by $f_{\theta,\sigma}(\lambda)$ of (5.1), and the GLS mean function estimates determined by θ^*, for $\theta,\theta^*\in\Theta_{is}$. From its decomposition into a sum of logarithms of univariate conditional densities, this function can be calculated as $-T/2$ times

$$L_T(\theta,\theta^*,\sigma)=\frac{1}{T}\sum_{t=1}^{T}\log\left(2\pi\sigma^2u_{t|t-1}(\theta)\right)$$

$$+\frac{1}{\sigma^2T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\breve{\eta}_{t,j}(1,\theta)\,Y_{t-j}+\hat{\alpha}_T^M(\theta^*)\sum_{j=0}^{t-1}\breve{\eta}_{t,j}(1,\theta)\,\xi_{t-j}^M\right)^2. \tag{9.12}$$

Part (c) of Proposition 9.1 in conjunction with other results is shown, in Appendix B, to yield the following extension of Theorem 8.1.

Theorem 9.2. *Let Θ and Θ^* be two model sets in Θ_{is} for which the assumptions of Proposition 9.1(b) hold. Suppose that the components y_t and ξ_t of the model (1.1) satisfy the assumptions of Subsection* **??**, *the condition (4.5), and $\sum_{t=1}^{\infty} \|x_t\| < \infty$. Then*

$$\sup_{\theta^* \in \Theta^*} \left\| \left(\hat{\alpha}_T^M (\theta^*) - \alpha^M \right) T^{-1/2} D_{\xi,M,T}^{-1} - \left[\begin{array}{cc} A^N C^{NM} (\theta^*) & 0 \end{array} \right] \right\| \to 0 \; a.s. \; [i.p.]. \tag{9.13}$$

Also, for $h = 1, 2, \ldots$, the forecast error sequences $W_t - \hat{W}_{t|t-h}^M (\theta, \theta^*, T)$ are jointly uniformly asymptotically stationary on $\Theta \times \Theta^*$ with the same asymptotic second moment functions as the sequences $W_t - W_{t|t-h}^M (\theta, \theta^*, T)$ of (8.4). In particular,

$$\sup_{\theta \in \Theta, \theta^* \in \Theta^*} \left| \frac{1}{T} \sum_{t=1}^{T} \left(W_t - \hat{W}_{t|t-h}^M (\theta, \theta^*, T) \right)^2 - \sigma_{hh}^{M,\delta} (\theta, \theta^*) \right| \to 0 \quad a.s. \; [i.p.], \tag{9.14}$$

with $\sigma_{hh}^{M,\delta} (\theta, \theta^*)$ defined by (8.5).
Furthermore, we have, for every $\varepsilon > 0$,

$$\lim_{T \to \infty} \sup_{\substack{\theta \in \Theta, \theta^* \in \Theta^* \\ \varepsilon \leq \sigma < \infty}} \left| L_T(\theta, \theta^*, \sigma) - \left\{ \log \left(2\pi\sigma^2 \right) + \frac{\sigma_{11}^{\delta,M}(\theta, \theta^*)}{\sigma^2} \right\} \right| = 0 \quad a.s. \; [i.p.], \tag{9.15}$$

and

$$\lim_{T \to \infty} \left| \inf_{\substack{\theta \in \Theta, \theta^* \in \Theta^* \\ 0 < \sigma < \infty}} L_T(\theta, \theta^*, \sigma) - \left\{ \log \left(2\pi \inf_{\theta \in \Theta, \theta^* \in \Theta^*} \sigma_{11}^{\delta,M}(\theta, \theta^*) \right) + 1 \right\} \right| = 0 \quad a.s. \; [i.p.]. \tag{9.16}$$

Remark 7. Part (d) of Proposition 9.1 can be used with $\nu(t) = 1 + t^{1/2+\varepsilon}$ with $\varepsilon > 0$ to establish the uniform convergence of $\sum_{t=1}^{\infty} x_t(\theta) x_t(\theta)'$ and $\left(\sum_{t=1}^{\infty} x_t(\theta) x_t(\theta)' \right)^{-1}$. With $\nu(t) = 1 + t^{3/2+\varepsilon}$, it yields the uniform convergence of $\sum_{t=1}^{\infty} t^{1/2} \|x_t(\theta)\|$. From this, the uniform convergence of $\sum_{t=1}^{\infty} y_t^M(\theta) x_t(\theta)' = \sum_{t=1}^{\infty} \left\{ t^{-1/2} y_t^M(\theta) \right\} \left\{ t^{1/2} x_t(\theta)' \right\}$ follows because $t^{-1/2} y_t^M(\theta)$ converges uniformly to 0 a.s. [i.p.], by Theorem 2.1 and Proposition 5.2 of FPW (2003). From these facts, one obtains the continuity of the function $\sum_{t=1}^{\infty} y_t^M(\theta) x_t(\theta)' \left(\sum_{t=1}^{\infty} x_t(\theta) x_t(\theta)' \right)^{-1}$ on Θ and the uniform convergence of the bias $\hat{a}_T^M(\theta) - a^M$ to this function.

10. Convergence of ARIMA Parameter Estimates When the Regressors Are Estimated By GLS

Theorem 9.2 can be applied to obtain convergence results for parameter estimates determined by h-step-ahead squared forecast error minimization or by Gaussian likelihood maximization over compact subsets of $\Theta_{is} \times \Theta_{is}$ with the regression coefficient vector α^M being estimated by the GLS estimates of Sections 5 or 9, or by OLS. Such results extend Theorems 4.1 and 5.1 of FPW (2003) to the cases of regARIMA models and regression models with a more general covariance structure. From the various possibilities, we present specific results for the two cases that seem to be of most practical interest, the case in which regression and autocovariance structure parameter estimates are obtained simultaneously and the case in which the regression coefficient estimates are available in advance of the estimate of the covariance structure, as when OLS estimates are used, or previously obtained GLS estimates of α^M are used for estimating θ optimally for multi-step-ahead forecasting.

In preparation, note that for any pair of compact subsets Θ and Θ^* of $\bar{\Theta}_{is}$ with ,Θ^* satisfying (6.7), it follows from their defining formula (8.3) that the forecast functions $W^M_{t+h|t}(\theta, \theta^*, T)$ are continuous on $\Theta \times \Theta^*$, so the same is true of the functions $\sum_{u=1}^{t}\left(W_u - W^M_{u|u-h}(\theta, \theta^*, T)\right)^2$. Under the assumptions of Theorem 10.1, the function $\sigma^{\delta,M}_{hh}(\theta, \theta^*)$ defined by (8.5) is continuous on this set, by Theorem 7.1. Therefore minimizers of these functions exist and the assertions of the following theorem are well known consequences of uniform convergence results obtained in Theorems 8.1 and 9.2. In the situation considered in (b) in which $\Gamma^{NM}(\theta^*) = 0$ for all θ^*, then $B^{NM}(\theta^*)$ in (7.15) does not depend on θ^*, so neither does $\sigma^{M,\delta}_{hh}(\theta, \theta^*)$. Recall that a random sequence $\theta^T = \theta^T(Y_1, \ldots, Y_T)$ is said to converge to a set Θ_0 almost surely if, on each realization of the time series Y_t except realizations belonging to an event with probability zero, every subsequence of θ^T has a convergent subsequence whose limit belongs to Θ_0. Convergence in probability to Θ_0 means that every subsequence of θ^T has a subsequence that converges almost surely to Θ_0. For the i.p. case, only measurable minimizers are considered, the existence of which is guaranteed by standard results, e.g. Lemma 3.4 of Pötscher and Prucha (1997).

Theorem 10.1. *Under assumptions of Section 3and (4.5), let Θ be a compact subset of Θ_{is}, and let $\theta^{*,T}$, $T \geq 1$ be a convergent random sequence, $\theta^{*,T} \to \theta^{*,M}$ a.s. [i.p.], contained in a compact subset Θ^* of $\bar{\Theta}_{is}$ with the properties of (6.2) and (6.7).*
(a) For a given $h \geq 1$ and each $T \geq 1$, let θ^T denote a minimizer of
$\sum_{t=1}^{T}\left(W_t - W^M_{t|t-h}(\theta, \theta, T)\right)^2$ over Θ. Then

$$T^{-1}\sum_{t=1}^{T}\left(W_t - W^M_{t|t-h}\left(\theta^T, \theta^T, T\right)\right)^2 \to \min_{\theta\in\Theta} \sigma^{\delta,M}_{hh}(\theta, \theta) \ \ a.s.\ [i.p.],$$

and the sequence θ^T, $T \geq 1$ converges a.s. [i.p.] to the set of minimizers of the function $\sigma^{\delta,M}_{hh}(\theta, \theta)$ over Θ, where $\sigma^{M,\delta}_{hh}(\theta, \theta)$ is defined by (8.5). In particular, if $\sigma^{\delta,M}_{hh}(\theta, \theta)$ has a unique minimizer $\theta^{h,M}$ over Θ, then $\theta^T \to \theta^M$ a.s. [i.p.].
(b) For a given $h \geq 1$, let $\theta^{h,T}$ denote a minimizer of $\sum_{t=1}^{T}\left(W_t - W^M_{t|t-h}\left(\theta, \theta^{*,T}, T\right)\right)^2$ over Θ. Then the sequence $\theta^{h,T}$ converges to the set of minimizers of $\sigma^{M,\delta}_{hh}\left(\theta, \theta^{*,M}\right)$ over Θ a.s. [i.p.]. If $\sigma^{M,\delta}_{hh}\left(\theta, \theta^{*,M}\right)$ has a unique minimizer $\theta^{h,M}$ in Θ, then $\theta^{h,T} \to \theta^{h,M}$ a.s. [i.p.]. The same conclusion obtains with no convergence requirement on $\theta^{*,T}$ when the regressors sequences X^M_t and X^N_t are asymptotically orthogonal.
(c) In the case of the finite-past forecasts (9.10), if Θ and Θ are compact subsets of Θ_{is} on which $\sum_{j=0}^{\infty}(1+j)|\theta_j|$ converges uniformly, then the conclusions of (a) and (b) also apply to the minimizers $\hat{\theta}^T$ of
$\sum_{t=1}^{T}\left(W_t - \hat{W}^M_{t|t-1}(\theta, \theta, T)\right)^2$ and $\hat{\theta}^{h,T}$ of $\sum_{t=1}^{T}\left(W_t - \hat{W}^M_{t|t-h}\left(\theta, \theta^{*,T}, T\right)\right)^2$ respectively, as

well as to the minimizers with respect to θ of $L_T(\theta,\theta,\sigma)$ and $L_T\left(\theta,\theta^{*,T},\sigma\right)$.

The interesting cases are usually those in which the limit $\theta^{*,M}$ of the sequence $\theta^{*,T}$ coincides with a minimizer $\theta^{h,M}$ of $\sigma_{hh}^{M,\delta}(\theta,\theta)$ over Θ for some $h \geq 1$. For example, when $h = 1$, one version of iterated GLS estimation is defined by $\theta^{*,T} = \hat{\theta}^{1,T-1}$ in which case $\theta^{*,T}$ will have the same limit as $\hat{\theta}^{1,T}$ if the latter sequence has a limit $\theta^{1,M}$.

Because $\sigma_{11}^{M,\delta}(\theta,\theta^*) = \Gamma_0^M(1,1,\theta,\theta^*)$, the conclusions of Corollary 7.3 concerning the optimality of GLS for one-step-ahead forecast apply in the estimation situations considered in Theorem 10.1 when $h = 1$.

Remark 8. An examination of the proofs of the theorems and corollaries of Sections 6.1–8 show their assertions for the case $h = 1$, and those of (a)–(b) of Theorem 10.1, hold for all $\|\cdot\|_1$-compact subsets Θ, Θ^* of absolutely summable filters, when (6.7) holds for Θ^* and the first coordinate of each $\theta^* \in \Theta^*$ is nonzero. That is, it is not necessary to restrict the zeroes of $\theta(z)$ and $\theta^*(z)$ to lie in $\{|z| \geq 1\}$. This is of interest because certain classes of multistep forecasting filters, for example, the direct autoregressive predictors discussed Subsection 4a of Findley (1984) and in the literature reviewed by Bhansali (1999), do not obey such restrictions.

11. Concluding Remarks

Under weak assumption on the regressor and regression errors of data of the form (1.1), we have shown that standard model coefficient estimates converge with increasing series length even when the model is incorrect with regard to its regressor specification or its regression error autocovariance specification. The limiting behavior is described by spectral calculus second moment formulas that generalize those that apply when the modeling assumptions are correct. Further, in the case of one-step-ahead forecasting with an underspecified mean function regressor, the formulas yield that the use of generalized least squares estimates of the regression models typically leads to smaller asymptotic average squared forecast errors than those obtained from ordinary least squares coefficient estimates. The results for the case of truncated infinite-past forecast functions did not require the restriction we imposed for the finite-past forecast functions that the zeroes of the generating function of the limiting prediction error filter have magnitude larger than one. Our development included a study of effects of regressors that decay too rapidly for consistent coefficient estimation. Formula were obtained for the asymptotic bias of their GLS coefficient estimates when their decay rate is at least $t^{-\frac{3}{2}-\varepsilon}$ for some $\varepsilon > 0$, as is the case for the intervention variables of Box and Tiao (1975).

The results and methods of the present paper can be used in conjunction with other results to obtain formulas for the limiting average of squared "out-of-sample" (real time) forecast errors of regARIMA models under slightly more restrictive assumptions on the regressor sequence X_t that are satisfied by all of the specific regressor types we have mentioned. The limit formulas are the same as those of the present article when the nontransitory regression component X_t is asymptotically stationary, i.e., when $D_{X,T} = T^{-1/2}I_X$, see Findley (2001, 2003).

When $D_{X,T} = T^{-1/2}I_X$, there are relatively straightforward multivariate time series extensions of those results of the present paper that do not involve the normalized forecast errors (9.3), for the reasons described in Section 7.2 (ii) of FPW (2003).

Acknowledgement. The author is grateful to John Aston for comments on a draft of this report that led to an improved exposition.

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

[1] Amemiya, T. (1973). "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica* 41, 723-732.

[2] Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York: Wiley.

[3] Bhansali, R. J. (1999). "Parameter Estimation and Model Selection for Multistep Prediction of Time Series: A Review," In S. Ghosh (Ed.), *Asymptotics, Nonparametrics and Time Series.* Marcel Dekker: New York.

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

[5] Bloomfield, P. (1973). "An exponential model for the spectrum of a scalar time series," *Biometrika* 60, 217-226.

[6] 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.

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

[8] Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods,* 2nd ed., Springer: New York.

[9] Dunford, N. and J. T. Schwartz (1957). *Linear Operators*, *Part I.* Wiley-Interscience: New York.

[10] Findley, D. F. (1984). "On Some Ambiguities Associated with the Fitting of ARMA models to Time Series," *Journal of Time Series Analysis* 5, 213–225.

[11] Findley, D. F. (1991). "Convergence of Finite Multistep Predictors from Incorrect Models and Its Role in Model Selection," *Note di Matematica* XI, 145-155.

[12] Findley, D. F. (2001). "Asymptotic Stationarity Properties of Out-of-Sample Forecast Errors of Misspecified RegARIMA Models," 2001 Proceedings of the Business and Economic Statistics Section, American Statistical Association, Alexandria.

[13] Findley, D. F. (2003). Asymptotic Stationarity Properties of Out-of-Sample Forecast Errors of Misspecified RegARIMA Models and the Optimality of GLS for One-Step-Ahead Forecasting, manuscript submitted for publication

[14] 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).

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

[16] Findley, D. F., Pötscher, B. M., and Wei, C. Z. (2003). "Modeling of Time Series Arrays by Multistep Prediction or Likelihood Methods," *Journal of Econometrics* (in press).

[17] 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

[18] Grenander, U. (1954). "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance," *Annals of Mathematical Statistics* 25, 252-272.

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

[20] Hirschman, I., Jr. (1965). "Finite Sections of Wiener-Hopf Equations and Szegö Polynomials," *J. Mathematical Analysis and Applications* 11, 290-320.

[21] Koreisha, S. G. and Fang, Y. (2001), "Generalized Least Squares with Misspecified Correlation Structure," *J. R. Statist. Soc. B* 63, 515-532.

[22] Lai, T. L. and Wei, C. Z. (1984). "Moment Inequalities with Applications to Regression and Time Series Models," in *Inequalities in Statistics and Probability* (ed. Y. L. Tong). Hayward: IMS.

[23] Newton, H. J. and Pagano, M. (1983). "The Finite Memory Prediction of Covariance Stationary Time Series," *SIAM J. Sci. Stat. Comput.* 4, 330-339.

[24] Parzen, E. (1962). "Spectral analysis of asymptotically stationary time series," *Bull. Inst. Internat. Statist.* 39, 87–103.

[25] Pierce, D. A. (1971). "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors," *Biometrika* 58, 299–312.

[26] Pötscher, B. M. (1991). "Noninvertibility and Pseudo-Maximum Likelihood Estimation of Misspecified ARMA Models," *Econometric Theory* 7, 435–449. Corrections: *Econometric Theory* 10, 811.

[27] Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Berlin: Springer.

[28] Rudin, W. (1973). *Functional Analysis.* New York: McGraw Hill.

[29] Thursby, J. G. (1987). "OLS or GLS in the Presence of Misspecification Error?" *Journal of Econometrics* 35, 359-374.

[30] Titchmarsh, E. C. (1939), *Theory of Functions,* 2nd Ed., Oxford University Press, Oxford

12. Appendix A: Uniform Convergence Results for Arrays

Proposition 12.1. *Let $V_t(T)$, $1 \le t \le T$, $T = 1, 2, \ldots$ be an n-dimensional column vector array satisfying (2.6) and (2.7). Let H and Z be sets of filters whose absolute coefficient sums converges uniformly. Then the filter output arrays $V_t[\eta](T)$, $V_t[\zeta](T)$, $1 \le t \le T$ defined as in (5.7) have the following properties:*

(a)

$$\lim_{T\to\infty} \sup_{\eta\in H} \left\| T^{-1/2} V_{T-j,T}[\eta] \right\| = 0, \ a.s. \ [i.p.] \ \ j \ge 0, \tag{12.1}$$

and, if (2.4) holds, then so does

$$\lim_{T\to\infty} \sup_{\eta\in H} \left\| T^{-1/2} V_{1+j,T}[\eta] \right\| = 0, \ a.s. \ [i.p.] \ \ j \ge 0.$$

(b) For any $k \ge 0$, as $T \to \infty$ one has

$$\sup_{\eta\in H, \zeta\in Z} \left\| \frac{1}{T} \sum_{t=1}^{T-k} V_{t+k}[\eta](T) V_t[\zeta](T)' - \Gamma_k^V(\eta, \zeta) \right\| \to 0 \ \ a.s. \ [i.p.], \tag{12.2}$$

with

$$\Gamma_k^V(\eta, \zeta) = \int_{-\pi}^{\pi} e^{-ik\lambda} \eta\left(e^{i\lambda}\right) \zeta\left(e^{-i\lambda}\right) dG_V(\lambda). \tag{12.3}$$

(c) The functions $\Gamma_k^V(\eta, \zeta)$ are bounded on $H \times Z$,

$$\left\| \Gamma_k^V(\eta, \zeta) \right\| \le \left\| \Gamma_0^V \right\| \sup_{\eta\in H} \left| \eta\left(e^{i\lambda}\right) \right| \sup_{\zeta\in Z} \left| \zeta\left(e^{i\lambda}\right) \right| < \infty,$$

and are jointly continuous in η, ζ in the sense that if $\eta^T \in H, \zeta^T \in Z$ are such that $\eta^T \to \eta, \zeta^T \to \zeta$ (coordinatewise convergence) with $\eta \in H, \zeta \in Z$, then $\Gamma_k^V\left(\eta^T, \zeta^T\right) \to \Gamma_k^V(\eta, \zeta)$. Also, if $Z = H$, then

$$\inf_{\eta\in H, -\pi\le\lambda\le\pi} \left| \eta\left(e^{i\lambda}\right) \right|^2 \Gamma_0^V \le \Gamma_0^V(\eta, \eta) \le \sup_{\eta\in H, -\pi\le\lambda\le\pi} \left| \eta\left(e^{i\lambda}\right) \right|^2 \Gamma_0^V, \tag{12.4}$$

so $\inf_{\eta\in H} \lambda_{\min}\left(\Gamma_0^V(\eta, \eta)\right) > 0$ if $\inf_{\eta\in H, -\pi\le\lambda\le\pi} \left| \eta\left(e^{i\lambda}\right) \right| > 0$.
(d) If there is a partition $V_t(T) = \begin{bmatrix} V_t^*(T) & V_t^{**}(T) \end{bmatrix}'$ with asymptotically orthogonal components, i.e.

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=k+1}^{T-k} V_{t\pm k}^*(T) V_t^{**}(T) = 0, \ a.s. \ [i.p.] \ (k = 0, 1, \ldots), \tag{12.5}$$

then

$$\lim_{T\to\infty} \sup_{\eta\in H, \zeta\in Z} \left\| \frac{1}{T} \sum_{t=k+1}^{T-k} V_{t\pm k}^*[\eta](T) V_t^{**}[\zeta](T)' \right\| = 0, \ a.s. \ [i.p.] \ (k = 0, 1, \ldots). \tag{12.6}$$

(e) Let B be an index set for a family of arrays $V_t(\beta, T)$, $1 \leq t \leq T$, $T = 1, 2, \ldots$, $\beta \in B$ such that

$$\sup_{\beta \in B} \left\| \frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' - \Gamma_0^V(\beta) \right\| \to 0 \ a.s., \tag{12.7}$$

holds, where the $\Gamma_0(\beta)$ are positive definite matrices whose minimum eigenvalues are bounded away from zero, i.e.

$$\inf_{\beta \in B} \lambda_{\min}(\Gamma_0(\beta)) \geq m_B$$

holds for some $m_B > 0$. Then

$$\sup_{\beta \in B} \left\| \left(\frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' \right)^{-1} - \Gamma_0(\beta)^{-1} \right\| \to 0 \ a.s. \tag{12.8}$$

Proof. Parts (a)–(c) are straightforward vector extensions of special cases of Theorem 2.1 and Proposition 2.1 of FPW (2001), and (d) follows from (12.3) because, under (12.5), $G_V(\lambda)$ can be written in the block diagonal form $G_V(\lambda) = diag[G_*(\lambda), G_{**}(\lambda)]$, where $G_*(\lambda)$ and $G_{**}(\lambda)$ are the spectral density matrices of the arrays $V_t^*(T)$ and $V_t^{**}(T)$, respectively.

For (e), it follows from (12.7) that, given $\varepsilon > 0$, for each realization except those of an event with probability zero, there is a T_ε such that for $T \geq T_\varepsilon$ the inequalities

$$\sup_{\beta \in B} \left\| \frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' - \Gamma_0^V(\beta) \right\| < \frac{\varepsilon}{2} m_H^2$$

and

$$\inf_{\beta \in B} \lambda_{\min} \left(\frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' \right) \geq \frac{m_B}{2}$$

hold. Hence for these T and all $\beta \in B$,

$$\begin{aligned}
& \sup_{\beta \in B} \left\| \left(\frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' \right)^{-1} - \Gamma_0(\beta)^{-1} \right\| \\
\leq \ & \sup_{\beta \in B} \left\{ \left\| \left(\frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' \right)^{-1} \right\| \left\| \frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' - \Gamma_0(\beta) \right\| \left\| \Gamma_0(\beta)^{-1} \right\| \right\} \\
\leq \ & \frac{1}{m_H} \sup_{\beta \in B} \left\{ \lambda_{\min}^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' \right) \right\} \sup_{\beta \in B} \left\{ \left\| \frac{1}{T} \sum_{t=1}^{T} V_t(\beta, T) V_t(\beta, T)' - \Gamma_0(\beta) \right\| \right\} \\
< \ & \varepsilon,
\end{aligned}$$

which establishes (12.8).

We also need the following lemma.

Lemma 12.2. *Suppose that, on a set Θ^*, the sequence $\beta_T(\theta^*)$, $T \geq 1$ of row vector functions converges uniformly to a bounded function$\beta(\theta^*)$, i.e. (7.7) holds, and similarly for $\tau_T(\theta^*)$, $T \geq 1$ and $\tau(\theta^*)$. Let $U_t(\eta, T), \eta \in H$ and $W_t(\zeta, T), \zeta \in Z$, $1 \leq t \leq T$, $T = 1, 2, \ldots$ be families of column vector arrays of the same dimension as $\beta(\theta^*)$ and $\tau(\theta^*)$, respectively, such that*

$$\sup_{\eta \in H, \zeta \in Z} \left\| \frac{1}{T} \sum_{t=1}^{T-k} U_{t+k}(\eta, T) W_t(\zeta, T)' - \Gamma_k(\eta, \zeta) \right\| \to 0 \ a.s. \ [i.p.], \ k = 0, 1, \ldots$$

and

$$\sup_{\eta \in H, \zeta \in Z} \|\Gamma_0(\eta, \zeta)\| < \infty$$

hold. Then

$$\sup_{\substack{\theta^* \in \Theta^* \\ \eta \in H, \zeta \in Z}} \left\| \frac{1}{T} \sum_{t=1}^{T-k} \beta_T(\theta^*) U_{t+k}(\eta, T) W_t(\zeta, T)' \tau_T(\theta^*)' - \beta(\theta^*) \Gamma_k(\eta, \zeta) \tau(\theta^*)' \right\|$$

$$\to 0 \ a.s. \ [i.p.], \ k = 0, 1, \ldots.$$

Proof. First note that, with $M_\beta = \sup_{\theta^* \in \Theta^*} \|\beta(\theta^*)\|$ and $M_\tau = \sup_{\theta^* \in \Theta^*} \|\tau(\theta^*)\|$, since

$$\begin{aligned} &\sup_{\theta^* \in \Theta^*} \left\| \frac{1}{T} \sum_{t=1}^{T-k} \beta(\theta^*) U_{t+k}(\eta, T) W_t(\zeta, T)' \tau'(\theta^*) - \beta(\theta^*) \Gamma_k(\eta, \zeta) \tau'(\theta^*) \right\| \\ \leq \quad & M_\beta M_\tau \sup_{\eta \in H, \zeta \in Z} \left\| \frac{1}{T} \sum_{t=1}^{T-k} U_{t+k}(\eta, T) W_t(\zeta, T)' - \Gamma_k(\eta, \zeta) \right\| \to 0 \ a.s. \ [i.p.], \end{aligned}$$

it suffices to verify

$$\sup_{\substack{\theta^* \in \Theta^* \\ \eta \in H, \zeta \in Z}} \left\| \frac{1}{T} \sum_{t=1}^{T-k} \left\{ \beta_T(\theta^*) U_{t+k}(\eta, T) W_t(\zeta, T)' \tau_T'(\theta^*) - \beta(\theta^*) U_{t+k}(\eta, T) W_t'(\zeta, T) \tau'(\theta^*) \right\} \right\|$$

$$\to 0 \quad a.s. \ [i.p.]. \tag{12.9}$$

By the usual difference of products decomposition, this is reduced to the proof of the uniform convergence to zero of three expressions, for example,

$$\sup_{\substack{\theta^* \in \Theta^* \\ \eta \in H, \zeta \in Z}} \left\| \frac{1}{T} \sum_{t=1}^{T-k} (\beta_T(\theta^*) - \beta(\theta^*)) U_{t+k}(\eta, T) W_t(\zeta, T)' \tau(\theta^*)' \right\| \to 0 \ a.s. \ [i.p.]. \tag{12.10}$$

our proof for which is representative. The expression on the left in (12.10) is bounded above by

$$M_\tau \sup_{\theta \in \theta^*} \|\beta_T(\theta^*) - \beta(\theta^*)\| \sup_{\eta \in H, \zeta \in Z} \left\| \frac{1}{T} \sum_{t=1}^{T-k} U_{t+k}(\eta, T) W_t(\zeta, T)' \right\|.$$

For the factor on the right in this bound, we have

$$\begin{aligned}
& \lim_{T\to\infty} \sup \sup_{\eta\in H,\zeta\in Z} \left\| \frac{1}{T} \sum_{t=1}^{T-k} U_{t+k}(\eta,T) W_t(\zeta,T)' \right\| \\
\leq \;& \lim_{T\to\infty} \sup_{\eta\in H,\zeta\in Z} \left\| \frac{1}{T} \sum_{t=1}^{T-k} U_{t+k}(\eta,T) W_t(\zeta,T)' - \Gamma_k(\eta,\zeta) \right\| + \sup_{\eta\in H,\zeta\in Z} \|\Gamma_k(\eta,\zeta)\| \\
\leq \;& 0 + \sup_{\eta\in H,\zeta\in Z} \|\Gamma_0(\eta,\zeta)\| < \infty.
\end{aligned}$$

Hence (12.10) follows from (7.7).

Finally, we need an instance of the basic uniform bounding inequalities for second moment sums of normalized one-step-ahead array forecast errors:

Lemma 12.3. *Let Θ be a subset of Θ_{is} on which $\sum_{j=0}^{\infty} |\theta_j|$ is bounded. Given a vector array $V_t(T)$, $1 \leq t \leq T, T = 1, 2, \ldots$, define, for every $\theta \in \Theta$,*

$$V_t(\theta,T) = \sum_{j=0}^{t-1} \breve{\eta}_{t,j}(1,\theta) V_{t-j}(T) \;\; 1 \leq t \leq T, T = 1, 2, \ldots,$$

where the coefficients are defined as in (9.3) with $h = 1$. Then for all $T \geq 1$,

$$\begin{aligned}
\min_{-\pi\leq\lambda\leq\pi} \left| \sum_{j=0}^{\infty} \theta_j e^{ij\lambda} \right|^2 \sum_{t=1}^{T} V_t(T) V_t(T)' &\leq \sum_{t=1}^{T} V_t(\theta,T) V_t(\theta,T)' \\
&\leq \max_{-\pi\leq\lambda\leq\pi} \left| \sum_{j=0}^{\infty} \theta_j e^{ij\lambda} \right|^2 \sum_{t=1}^{T} V_t(T) V_t(T)'. \qquad (12.11)
\end{aligned}$$

Proof. We only need verify (12.11) for the case of scalar arrays $v_t(T)$, $1 \leq t \leq T, T = 1, 2, \ldots$: the vector case is reduced to this one by multiplying each expression in (12.11) on the left by an unrestricted nonzero row vector of length $\dim V_t(T)$ and on the right by the transpose of this vector. For every $T \geq 1$, define the column vector $\mathbf{V}_T = [v_1(T) \; \ldots \; v_T(T)]'$, the covariance matrix $\Gamma_T(\theta) = [\gamma_{j-k}(\theta)]_{0\leq j,k\leq T-1}$, and the lower triangular matrix $L_T(\theta)$ of order T whose t-th row is given by $[\breve{\eta}_{t,t-1}(1,\theta) \; \ldots \; \breve{\eta}_{t,0}(1,\theta) \; \mathbf{0}_{T-t}]$, where $\mathbf{0}_{T-t}$ denotes a vector of zeros of length $T - t$. Then for each $T \geq 1$,

$$\sum_{t=1}^{T} v_t(\theta,T)^2 = \mathbf{V}_T' L_T(\theta)' L_T(\theta) \mathbf{V}_T = \mathbf{V}_T' \Gamma_T(\theta)^{-1} \mathbf{V}_T.$$

Because the θ-model's spectral density is proportional to $\left|\sum_{j=0}^{\infty} \theta_j e^{ij\lambda}\right|^{-2}$, it follows from Lemma 10.2.6 of Anderson (1971) that

$$\min_{-\pi\leq\lambda\leq\pi} \left|\theta\left(e^{ij\lambda}\right)\right|^2 \sum_{t=1}^{T} v_t(T)^2 \leq \sum_{t=1}^{T} v_t(T,\theta)^2 \leq \max_{-\pi\leq\lambda\leq\pi} \left|\theta\left(e^{ij\lambda}\right)\right|^2 \sum_{t=1}^{T} v_t(T)^2,$$

for each $T \geq 1$, which is (12.11) for scalar arrays. This completes the proof.

13. Appendix B: Proofs

13.1. Proof of Proposition 3.1

From (3.10) and (3.11), it follows that for any $T \geq k+1$,

$$\sum_{t=T}^{\infty} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\| \leq \left\{ \sup_{t \geq T, \zeta \in Z} \left\| V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\| \right\} \sup_{\eta \in H} \sum_{t=T}^{\infty} \| x_t(\eta) \| < \infty.$$

Hence, by (3.12), for a given $\varepsilon > 0$, there is a $T_0 \geq k+2$ such that for $T \geq T_0 + k$, $\sup_{\eta \in H, \zeta \in Z} \sum_{t=T_0}^{T-k} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\| < \varepsilon/2$. Because the $D_{V,T}$ sequence is decreasing, in this sum, $D_{V,t\pm k} \geq D_{V,T}$. Hence, for $T \geq T_0 + k$,

$$\sum_{t=T_0}^{T-k} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,T} \right\| \leq \sum_{t=T_0}^{T-k} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\| \left\| D_{V,t\pm k}^{-1} D_{V,T} \right\| < \frac{\varepsilon}{2}. \tag{13.1}$$

Also for $T \geq T_0 + k$, calculating as above, and then using the fact that $\left\| D_{V,t\pm k}^{-1} D_{V,T} \right\| \leq \left\| D_{V,t\pm k}^{-1} \right\| \| D_{V,T} \|$, we obtain

$$\begin{aligned}
\sum_{t=k+1}^{T_0-1} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,T} \right\| &\leq \left\{ \max_{k+1 \leq t \leq T_0-1} \left\| D_{V,t\pm k}^{-1} D_{V,T} \right\| \right\} \sum_{t=k+1}^{T_0-1} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\| \\
&\leq \| D_{V,T} \| \left\{ \max_{k+1 \leq t \leq T_0-1} \left\| D_{V,t\pm k}^{-1} \right\| \right\} \sum_{t=k+1}^{T_0-1} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\| \\
&= \| D_{V,T} \| \left\| D_{V,T_0-1}^{-1} \right\| \sum_{t=k+1}^{T_0-1} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,t\pm k} \right\|.
\end{aligned}$$

Since $\| D_{V,T} \| \to 0$, it follows that there is a $T_1 \geq T_0$ such that for $T \geq T_1 + k$, we have $\sum_{t=k+1}^{T_0-1} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,T} \right\| < \varepsilon/2$, and therefore, from (13.1), also

$$\sum_{t=k+1}^{T-k} \left\| x_t(\eta) V'_{t\pm k}(\zeta) D_{V,T} \right\| < \varepsilon.$$

This establishes (3.13).

13.2. Proof of Lemma 5.1

It follows from (6.3) that $\theta^T(z)$ converges to $\theta(z)$ uniformly on $\{|z| \leq 1\}$.Since the $\theta^T(z)$ have no zeros in $\{|z| < 1\}$, the Theorem of Hurwitz (see Titchmarsh, 1939, p. 119), shows that the same must be true of $\theta(z)$, as asserted. Next, given any $\theta \in \bar{\Theta}_{is}$ and any positive, strictly increasing sequence ρ_T converging to 1, define $\theta^T = \left(\rho_T^j \theta_j \right)_{j \geq 0}$. Then since $\theta^T(z) = \theta(\rho_T z)$, we have $\theta^T \in \Theta_{is}$ as well as

$$\left\|\theta^T - \theta\right\|_1 = \sum_{j=1}^{\infty}\left(1-\rho_T^j\right)|\theta_j| \le \left(1-\rho_T^N\right)\sum_{j=1}^{N}|\theta_j| + \sum_{j=N+1}^{\infty}|\theta_j|$$

for every $N \ge 1$ from which $\left\|\theta^T - \theta\right\|_1 \to 0$ follows.

13.3. Proof of Lemma 6.1

As we mentioned when (6.4) was introduced, under (6.2), the condition (6.4) guarantees that the compact set $\bar{\Theta}$ consisting of all limit points of Θ is a subset of Θ_{is}. Since the transformation $\theta \mapsto \tilde{\theta}$ is continuous on Θ_{is} for $\|\cdot\|_1$-convergence, the image set $\left\{\tilde{\theta} : \theta \in \bar{\Theta}\right\}$ is $\|\cdot\|_1$-compact. Consequently, $\sum_{j=1}^{\infty}\left|\tilde{\theta}_j\right|$ converges uniformly on this set and therefore on the subset $\tilde{\Theta}$. Conversely, uniform convergence of $\sum_{j=1}^{\infty}\left|\tilde{\theta}_j\right|$ on $\tilde{\Theta}$, implies that

$$\left\{\min_{-\pi\le\lambda\le\pi,\theta\in\Theta}\left|\theta\left(e^{i\lambda}\right)\right|\right\}^{-1} = \sup_{-\pi\le\lambda\le\pi,\theta\in\Theta}\left|\tilde{\theta}\left(e^{i\lambda}\right)\right| \le \sup_{\theta\in\Theta}\sum_{j=1}^{\infty}\left|\tilde{\theta}_j\right| < \infty,$$

from which (6.4) follows.

13.4. Proof of Theorem 6.2

We have

$$\left(\alpha_T^M(\theta) - \alpha^M\right)T^{-1/2}D_{\xi,M,T}^{-1}$$

$$\begin{aligned}
&= T^{-1/2}\sum_{t=1}^{T} y_t^M[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\left(D_{\xi,M,T}\sum_{t=1}^{T}\xi_t^M[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\right)^{-1} \\
&= T^{-1/2}\sum_{t=1}^{T} y_t[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\left(D_{\xi,M,T}\sum_{t=1}^{T}\xi_t^M[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\right)^{-1} && (13.2)\\
&\quad + T^{-1/2}\alpha^N\sum_{t=1}^{T}\xi_t^N[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\left(D_{\xi,M,T}\sum_{t=1}^{T}\xi_t^M[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\right)^{-1}, && (13.3)
\end{aligned}$$

see (4.3). By (b) and (d) of Proposition 12.1, $D_{\xi,M,T}\sum_{t=1}^{T}\xi_t^M[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}$ converges uniformly to $diag\left(\Gamma_0^{MM}(\theta),\ \sum_{t=1}^{\infty}x_t^M[\theta]\,x_t^M[\theta]'\right)$ a.s. [i.p.], which, is bounded below by

$$diag\left(\inf_{\theta\in\Theta}\lambda_{\min}\left(\Gamma_0^{MM}(\theta)\right)I_{X,M},\ \inf_{\theta\in\Theta}\lambda_{\min}\left(\sum_{t=1}^{\infty}x_t^M[\theta]\,x_t^M[\theta]'\right)I_{x,M}\right), \tag{13.4}$$

where $I_{X,M}$ and $I_{x,M}$ denote the identity matrix of orders $\dim X_t^M$ and $\dim x_t^M$ respectively. By (6.7), $\inf_{\theta\in\Theta}\lambda_{\min}\left(\Gamma_0^{MM}(\theta)\right)>0$. We verify that also

$$\inf_{\theta\in\Theta}\lambda_{\min}\left(\sum_{t=1}^{\infty}x_t^M[\theta]\,x_t^M[\theta]'\right)>0. \tag{13.5}$$

Indeed, because $\sum_{t=1}^{\infty}x_t^M[\theta]\,x_t^M[\theta]'$ is continuous on $\bar{\Theta}$, failure of (13.5) would mean that for some $\theta\in\bar{\Theta}$ and some vector c of dimension $\dim x_t^M$, we have $\sum_{t=1}^{\infty}\left(c'x_t^M[\theta]\right)^2=0$, i.e. $c'x_t^M[\theta]=0$ for $t=1,2,\ldots$ .Since $\theta_0=1$, this yields $c'x_t^M=0$ for $t=1,2,\ldots$ in contradiction to (3.4). Hence (13.4) is positive definite.

Now (e) of Proposition 12.1 yields that $\left(D_{\xi,M,T}\sum_{t=1}^{T}\xi_t^M[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\right)^{-1}$ converges uniformly to $diag\left(\Gamma_0^{MM}(\theta)^{-1},\ \left(\sum_{t=1}^{\infty}x_t^M[\theta]\,x_t^M[\theta]'\right)^{-1}\right)$. By Proposition 12.1 again, $T^{-1/2}\sum_{t=1}^{T}y_t[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}$, and therefore also the expression (13.2), converges uniformly to zero a.s. [i.p.]. With $I_{X,N}$ and $I_{x,N}$ denoting identity matrices with orders the dimensions of X_t^N and x_t^N respectively, observe that

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\xi_t^N[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}\\
=&\left[\begin{array}{cc} I_{X,N} & 0\\ 0 & T^{-1/2}I_{x,N}\end{array}\right]\sum_{t=1}^{T}D_{\xi,N,T}\xi_t^N[\theta]\,\xi_t^M[\theta]'\,D_{\xi,M,T}.
\end{aligned}$$

Applying Proposition 12.1 as before and Lemma 12.2, we obtain that the expression (13.3) converges uniformly a.s. [i.p.] to α^N times

$$\left[\begin{array}{cc} I_{X,N} & 0\\ 0 & 0\end{array}\right]\left[\begin{array}{cc}\Gamma_0^{NM}(\theta)\,\Gamma_0^{MM}(\theta)^{-1} & 0\\ 0 & \sum_{t=1}^{\infty}x_t^N[\theta]\,x_t^M[\theta]'\left(\sum_{t=1}^{\infty}x_t^M[\theta]\,x_t^M[\theta]'\right)^{-1}\end{array}\right]$$
$$=diag\left(C^{NM}(\theta),\ 0\right).$$

This yields (6.8).

Finally, $C^{NM}(\theta)$ is bounded and continuous on Θ because this is true of $\Gamma_0^{NM}(\theta)$ and $\Gamma_0^{MM}(\theta)^{-1}$, by (c) of Proposition 12.1 and (6.7).

13.5. Proof of Corollary 7.3

By (c) of Proposition 12.1, $\Gamma_0^M(1,1,\theta,\theta)$ and $\Gamma_0^M(1,1,\theta,\theta^*)$ are continuous functions of θ on Θ. Hence the respective minimizers $\bar{\theta}$ and $\bar{\theta}^*$ exist. The inequality (7.22) follows from

$$\Gamma_0^M\left(1,1,\bar{\theta},\bar{\theta}\right)\le\Gamma_0^M\left(1,1,\bar{\theta}^*,\bar{\theta}^*\right)\le\Gamma_0^M\left(1,1,\bar{\theta}^*,\theta^*\right),$$

where the first inequality results from the minimizing property of $\bar{\theta}$ and the second from (7.19) with $C=C^{NM}(\theta^*)$. The strictness of either inequality implies the strictness of (7.22). If the first inequality is not strict but (7.20) holds for $\theta=\bar{\theta}^*$, then the second inequality is strict.

13.6. Proof of Proposition 9.1

(a) follows by setting $V_t(T) = \xi_t$ in Lemma 12.3 and using (3.4) to obtain that $\sum_{t=1}^{T} \xi_t \xi_t' > 0$ for T sufficiently large.

For (b), because each $\eta_{t,j}(h,\theta) - \eta_j(h,\theta)$ is continuous on Θ_{is} by Proposition 3.1(a) of FPW (2003) and therefore bounded on the compact set $\bar{\Theta}$ of limit points of Θ, and hence also on Θ, to prove the first uniform convergence assertion, it suffices to verify

$$\lim_{t_0 \to \infty} \sup_{\theta \in \Theta} \sum_{t=t_0}^{\infty} \sum_{j=0}^{t-1} \left| \eta_{t,j}(h,\theta) - \eta_j(h,\theta) \right| = 0. \tag{13.6}$$

From the proof of the Baxter inequality (3.6) of Findley (1991), for any weights $\nu(j)$, $j \geq 0$ as in (d) and for any $t_0 = t_0(\theta) \geq 1$ large enough that

$$\sum_{j=0}^{\infty} \nu(j) |\tilde{\theta}_j| \cdot \sum_{j=t_0-h+1}^{\infty} \nu(j) |\theta_j| \leq \frac{1}{2}$$

holds we have, for all $t \geq t_0$,

$$\sum_{j=0}^{t-1} \nu(j) \left| \eta_{t,j}(h,\theta) - \eta_j(h,\theta) \right|$$

$$\leq 9 \left(\sum_{j=0}^{\infty} \nu(j) |\tilde{\theta}_j| \right)^2 \left(\sum_{j=0}^{h-1} \nu(j) |\tilde{\theta}_j| \right) \left(\sum_{j=0}^{\infty} \nu(j) |\theta_j| \right)^2 \sum_{j=t-h+1}^{\infty} \nu(j) |\theta_j|. \tag{13.7}$$

From the case $\nu(j) \equiv 1$, we obtain

$$\begin{aligned}
&\sum_{t=t_0}^{\infty} \sum_{j=0}^{t-1} \left| \eta_{t,j}(h,\theta) - \eta_j(h,\theta) \right| \\
&\leq 9 \left(\sum_{j=0}^{\infty} |\tilde{\theta}_j| \right)^2 \left(\sum_{j=0}^{h-1} |\tilde{\theta}_j| \right) \left(\sum_{j=0}^{\infty} |\theta_j| \right)^2 \sum_{t=t_0}^{\infty} \sum_{j=t-h+1}^{\infty} |\theta_j| \\
&= 9 \left(\sum_{j=0}^{\infty} |\tilde{\theta}_j| \right)^2 \left(\sum_{j=0}^{h-1} |\tilde{\theta}_j| \right) \left(\sum_{j=0}^{\infty} |\theta_j| \right)^2 \sum_{j=t_0-h+1}^{\infty} (j+h-t_0) |\theta_j|.
\end{aligned}$$

By (6.2), $D_t = \sup_{\theta \in \Theta} \sum_{j=t}^{\infty} |\theta_j| < \infty$, for all $t \geq 0$, and $D_t \searrow 0$ as $t \to \infty$. By Lemma 6.1, $\tilde{D}_0 = \sup_{\theta \in \Theta} \sum_{j=0}^{\infty} |\tilde{\theta}_j| < \infty$. Therefore, for t_0 large enough that $\tilde{D}_0 D_{t_0} \leq 1/2$, we have

$$\sup_{\theta \in \Theta} \sum_{t=t_0}^{\infty} \sum_{j=0}^{t-1} \left| \eta_{t,j}(h,\theta) - \eta_j(h,\theta) \right| \leq 9 \tilde{D}_0^3 D_0^2 \sup_{\theta \in \Theta} \sum_{j=t_0-h+1}^{\infty} (j+h-t_0) |\theta_j|,$$

from which (13.6) follows by the uniform convergence of $\sum_{j=0}^{\infty}(1+j)\,|\theta_j|$.

To prove the uniform convergence of $\sum_{t=1}^{\infty}\left\|\sum_{j=0}^{t-1}\eta_{t,j}\left(h,\theta\right)x_{t-j}\right\|$, consider the inequalities

$$\sum_{t=t_0}^{\infty}\left\|\sum_{j=0}^{t-1}\eta_{t,j}\left(h,\theta\right)x_{t-j}\right\| \leq \sum_{t=t_0}^{\infty}\left\|\sum_{j=0}^{t-1}\left|\eta_{t,j}\left(h,\theta\right)-\eta_j\left(h,\theta\right)\right|x_{t-j}\right\| + \sum_{t=t_0}^{\infty}\left\|\sum_{j=0}^{t-1}\eta_j\left(h,\theta\right)x_{t-j}\right\|$$

for $t_0 \geq 1$. The first expression on the right is bounded above by

$$\sup_{t\geq 1}\|x_t\|\sum_{t=t_0}^{\infty}\sum_{j=0}^{t-1}\left|\eta_{t,j}\left(h,\theta\right)-\eta_j\left(h,\theta\right)\right|$$

and so converges uniformly on Θ by (13.6). To complete the proof of (b), it remains to that same is true of the second expression.

By (b) of Proposition 3.1 of FPW (2003), $\sum_{j=0}^{\infty}\left|\eta_j\left(h,\theta\right)\right|$ is continuous and uniformly convergent on the larger, compact set $\bar{\Theta}$. Define $x_0 = 0$. Because, for a fixed absolutely summable sequence $\mathbf{x} = \left(x_t\right)_{j\geq 0}$, the sequence mapping $\eta \mapsto \eta * \mathbf{x} = \left(\sum_{j=0}^{t-1}\eta_j x_{t-j}\right)_{t\geq 0}$ is $\|\cdot\|_1$-continuous on the space l^1 of absolutely summable sequences $\left(\eta_j\right)_{j\geq 0}$, it follows that $\left\{\left(\sum_{j=0}^{t-1}\eta_j\left(h,\theta\right)x_{t-j}\right)_{t\geq 0} : \theta\in\Theta\right\}$ is compact. Hence $\sum_{t=1}^{\infty}\left\|\sum_{j=0}^{t-1}\eta_j\left(h,\theta\right)x_{t-j}\right\|$ is absolutely convergent on $\bar{\Theta}$.

For (c) it follows from (6.7) and (9.7) that $\left\|\left(\sum_{t=1}^{T}x_t\left(\theta\right)x_t\left(\theta\right)'\right)^{-1}\right\|$ is uniformly bounded. Thus it suffices to verify

$$\sup_{\theta\in\Theta}\left\|T^{-1/2}\sum_{t=1}^{T}y_t^M\left(\theta\right)x_t\left(\theta\right)'\right\| \to 0 \ a.s.\ [i.p.].$$

By Proposition 3.1, this follows from the result $\sup_{\theta\in\Theta}\sum_{t=1}^{\infty}\|x_t\left(\theta\right)\| < \infty$ of (b) together with $\sup_{\theta\in\Theta}t^{-1/2}\left|y_t^M\left(\theta\right)\right| \to 0$ *a.s.* [*i.p.*], which is established by (a2) of Theorem 2.1 of FPW (2001) and (13.6), by virtue of the asymptotic stationarity of y_t^M.

For (d), we return to (13.7) and use $\nu\left(t\right) \leq \nu\left(j\right)\nu\left(t-j\right),\ 0\leq j\leq t$ to calculate

$$\nu\left(t\right)\sum_{j=0}^{t-1}\left\|\left(\eta_{t,j}\left(h,\theta\right)-\eta_j\left(h,\theta\right)\right)x_{t-j}\right\| \leq \sum_{j=0}^{t-1}\nu\left(j\right)\left|\left(\eta_{t,j}\left(h,\theta\right)-\eta_j\left(h,\theta\right)\right)\right|\nu\left(t-j\right)\|x_{t-j}\|$$

$$\leq 9\left\{\sup_{t\geq 1}\nu\left(t\right)\|x_t\|\right\}\left(\sum_{j=0}^{\infty}\nu\left(j\right)|\tilde{\theta}_j|\right)^2\left(\sum_{j=0}^{h-1}\nu\left(j\right)|\tilde{\theta}_j|\right)\left(\sum_{j=0}^{\infty}\nu\left(j\right)|\theta_j|\right)^2\sum_{j=t-h+1}^{\infty}\nu\left(j\right)|\theta_j|.$$

From $\nu(t) \leq \nu(j)\,\nu(t-j)\,,\ 0 \leq j \leq t$ again, we obtain

$$\begin{aligned}
\nu(t) \sum_{j=0}^{t-1} \left\|\eta_j(h,\theta)\, x_{t-j}\right\| &\leq \left\{\sup_{t\geq 1} \nu(t)\,\|x_t\|\right\} \sum_{j=0}^{t-1} \nu(j)\left|\eta_j(h,\theta)\right| \\
&\leq \left\{\sup_{t\geq 1} \nu(t)\,\|x_t\|\right\} \left(\sum_{j=0}^{h-1} \nu(j)\,|\tilde{\theta}_j|\right) \left(\sum_{j=0}^{\infty} \nu(j)\,|\theta_j|\right).
\end{aligned}$$

Hence, (d) follows from

$$\nu(t) \left\|\sum_{j=0}^{t-1} \eta_{t,j}(h,\theta)\, x_{t-j}\right\| \leq \nu(t) \sum_{j=0}^{t-1} \left\|\left(\eta_{t,j}(h,\theta) - \eta_j(h,\theta)\right) x_{t-j}\right\| + \nu(t) \sum_{j=0}^{t-1} \left\|\eta_j(h,\theta)\, x_{t-j}\right\|$$

$$\leq \left\{\sup_{t\geq 1} \nu(t)\,\|x_t\|\right\} \left(\sum_{j=0}^{\infty} \nu(j)\,|\tilde{\theta}_j|\right) \left\{9\left(\sum_{j=0}^{\infty} \nu(j)\,|\tilde{\theta}_j|\right)^2 \left(\sum_{j=0}^{\infty} \nu(j)\,|\theta_j|\right)^3 + \left(\sum_{j=0}^{\infty} \nu(j)\,|\theta_j|\right)\right\}.$$

13.7. Proof of Theorem 9.2

The similarity of (9.11) and (8.4) makes clear that it suffices to consider the case $d = 1$, i.e. $W_t = Y_t$. In analogy with the proof of Theorem 7.1, it therefore suffices to prove that

$$\sup_{\theta^* \in \Theta^*} \left\|\hat{\beta}_T(\theta^*) - \beta(\theta^*)\right\| \to 0 \ a.s.\ [i.p.] \tag{13.8}$$

and that $T^{-1}\sum_{t=1}^{T-k} V_{t+k}^h(\theta,T)\, V_t^l(\theta,T)'$ converges uniformly a.s. on Θ to

$$\left[\begin{array}{ccc} \gamma_k^y(h,l,\theta) & 0 & 0 \\ 0 & \Gamma_k^X(h,l,\theta) & 0 \\ 0 & 0 & \sum_{t=1}^{\infty}\left(\sum_{j=0}^{t-1} \eta_{t+k,j}(h,\theta)\, x_{t+k-j}\right)\left(\sum_{j=0}^{t-1} \eta_{t,j}(l,\theta)\, x_{t-j}\right)' \end{array}\right], \tag{13.9}$$

with

$$\gamma_k^y(h,l,\theta) = \int_{-\pi}^{\pi} e^{-ik\lambda}\eta(h,\theta)\left(e^{i\lambda}\right)\eta(l,\theta)\left(e^{-i\lambda}\right) dG_y(\lambda)\,,$$

etc. We start with the latter. The uniform convergence of $T^{-1}\sum_{t=1}^{T-k} U_{t+k}^h(\theta,T)\, U_t^l(\theta,T)'$ to $diag\left(\gamma_k^y(h,l,\theta)\,,\Gamma_k^X(h,l,\theta)\right)$ a.s. [i.p.] follows from (3.18) by way of Proposition 5.2 and part (a) of Proposition 2.1 of FPW (2003), which also yield the $U_t^h(\theta,T)$ version of (3.11) and (3.12),

$$\sup_{t\geq 1,\theta\in\Theta} \left\|t^{-1/2} U_t^h(\theta,t)\right\| < \infty \ a.s.\ [i.p.]$$

and

$$\lim_{t\to\infty} \sup_{\theta\in\Theta} \left\|t^{-1/2} U_t^h(\theta,t)\right\| = 0 \ a.s.\ [i.p.],$$

because $U_t(T)$ has both negligibility properties (2.8) and (2.7). The uniform convergence of $\sum_{t=1}^{\infty} \left\| \sum_{j=0}^{t-1} \eta_{t,j}(l,\theta)\, x_{t-j} \right\|$ established in (c) of Proposition 9.1 therefore enables us to apply Proposition 3.1 to obtain the off-diagonal 0's in (13.9) and to directly obtain the last diagonal entry.

For (13.8), it suffices to establish (9.13), which follows by direct adaptation of the arguments used to obtain (13.9) as a uniform limit to generalize the proof of Theorem 6.2 to the finite-past case.

Finally, (9.15) and (9.16) follow from an argument completely analogous to the proof of (b) of Theorem 5.1 of FPW (2003), using (9.14) instead of the corresponding result for the case of no regressors.