Econometric Theory, **23**, 2007, 1083–1107. Printed in the United States of America.
DOI: 10.1017/S0266466607070430

OPTIMALITY OF GLS FOR ONE-STEP-AHEAD FORECASTING WITH REGARIMA AND RELATED MODELS WHEN THE REGRESSION IS MISSPECIFIED

David F. Findley
U.S. Census Bureau

We consider the modeling of a time series described by a linear regression component whose regressor sequence satisfies the generalized asymptotic sample second moment stationarity conditions of Grenander (1954, *Annals of Mathematical Statistics* 25, 252–272). The associated disturbance process is only assumed to have sample second moments that converge with increasing series length, perhaps after a differencing operation. The model's regression component, which can be stochastic, is taken to be underspecified, perhaps as a result of simplifications, approximations, or parsimony. Also, the autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) model used for the disturbances need not be correct. Both ordinary least squares (OLS) and generalized least squares (GLS) estimates of the mean function are considered. An optimality property of GLS relative to OLS is obtained for one-step-ahead forecasting. Asymptotic bias characteristics of the regression estimates are shown to distinguish the forecasting performance. The results provide theoretical support for a procedure used by Statistics Netherlands to impute the values of late reporters in some economic surveys.

1. INTRODUCTION

For many economic indicator series, modeling requires specification of both a regression function and an autocovariance structure for the disturbance process. Suppose that, possibly after a variance stabilizing transformation (e.g., differencing), one has data W_t, $1 \leq t \leq T$ of the form

$$W_t = AX_t + y_t, \tag{1.1}$$

The author thanks two referees and the co-editor for comments and suggestions that led to substantial improvements in the exposition and also thanks John Aston and Tucker McElroy for helpful comments on an earlier draft. Any views expressed are the author's and not necessarily those of the U.S. Census Bureau. Address correspondence to David F. Findley, U.S. Census Bureau, Washington, DC 20233, USA; e-mail: david.f.findley@census.gov.

 0266-4666/07 $15.00

where the X_t are column vectors and the y_t are real variates that are asymptotically orthogonal to the X_t in a sense to be defined, whose lagged sample second moments converge as $T \to \infty$. With monthly or quarterly seasonal economic data, AX_t might describe a linear or higher degree trend, stable seasonal effects, moving holiday effects (Bell and Hillmer, 1983), trading day effects (Findley, Monsell, Bell, Otto, and Chen, 1998), or other periodic effects. The term X_t might also include values of related stochastic variables, perhaps at leads or lags. We address the situation in which the modeler considers a model

$$W_t = A^M X_t^M + y_t^M \tag{1.2}$$

whose regressor X_t^M is not able to reproduce AX_t for all t because of known or unknown omissions, approximations, simplifications, etc. We assume that the modeler, perhaps starting from the ordinary least squares (OLS) estimate for A^M given by (1.5) later in this section, has decided upon an autoregressive moving average (ARMA) model family, not necessarily correct, for the disturbance (or residual) process $y_t^M = W_t - A^M X_t^M$. Such a model for (1.2) is called a regARMA model.

Generalized least squares (GLS) estimation of A^M occurs simultaneously with ARMA estimation. The simplest definition of (feasible) GLS estimates of A^M, given by (1.3), makes use of the ARMA model's innovation filter that is defined as follows. With L denoting the lag operator, let $\phi(L)$ be the autoregressive polynomial (AR) and $\alpha(L)$ the moving average (MA) polynomial of a (perhaps incorrect) candidate ARMA model for y_t^M and let $\theta = (1, \theta_1, \theta_2, \ldots)$ denote the coefficient sequence of the power series expansion $\phi(L)/\alpha(L) = \sum_{j=0}^{\infty} \theta_j L^j$. When y_t in (1.1) and the regressors missing from X_t^M are weakly (i.e., first and second moment) stationary with mean zero, then y_t^M will be weakly stationary with mean zero. In this case, assuming that values of y_t^M are available at all past times, $y_{t|t-1}^M(\theta) = -\sum_{j=1}^{\infty} \theta_j y_{t-j}^M$ is the model's linear forecast of y_t^M from y_s^M, $-\infty < s \le t-1$; see Section 5.3.3 of Box and Jenkins (1976) or Hannan (1970, p. 147). The forecast errors $a_t(\theta) = y_t^M - y_{t|t-1}^M(\theta) = \sum_{j=0}^{\infty} \theta_j y_{t-j}^M$ are called the model's innovations series, and the coefficient sequence θ is its innovation filter. If the ARMA model is correct, then for each t, $a_t(\theta)$ is uncorrelated with y_s^M, $-\infty < s \le t-1$, and it follows that $y_{t|t-1}^M(\theta)$ has minimum mean square error among all such linear forecasts of y_t^M and that the innovations $a_t(\theta)$ are uncorrelated (white noise). However, *we do not assume that a correct ARMA model exists* or that y_t^M is weakly stationary. For example, when a missing regressor is deterministic, e.g., periodic, y_t^M will not be weakly stationary even when y_t is but will instead be *asymptotically stationary*, meaning that its lagged sample second moments will converge as T increases. Their limits form the autocovariance sequence of a weakly stationary process. In effect, it is this autocovariance sequence for which an ARMA model is sought. All ARMA model–related quantities of interest in this paper depend only on θ and on the W_t and X_t^M. Thus we can express model dependence in terms of θ, as we

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

do throughout the paper. Further motivation for this "parameterization" is given in Section 3. We refer to each θ as a model.

For given W_t, X_t^M, $1 \leq t \leq T$ and θ, define $W_t[\theta] = \sum_{j=0}^{t-1} \theta_j W_{t-j}$ and $X_t^M[\theta] = \sum_{j=0}^{t-1} \theta_j X_{t-j}^M$ for $1 \leq t \leq T$ and let $'$ denote transpose. Following Pierce (1971), we define the θ-model's GLS estimator of A^M to be

$$A_T^M(\theta) = \sum_{t=1}^{T} W_t[\theta] X_t^M[\theta]' \left(\sum_{t=1}^{T} X_t^M[\theta] X_t^M[\theta]' \right)^{-1}. \tag{1.3}$$

(We discuss another GLS estimator in Section 8.) With these $A_T^M(\theta)$, an estimate of θ (and of the ARMA coefficients determining θ when they are identified) can be obtained by conditional or unconditional maximum likelihood estimation (MLE). (As usual, Gaussian likelihood functions are used without requiring the data to be Gaussian.) For the conditional MLE estimates on which we focus for simplicity (see Box and Jenkins, 1976, Sect. 7.1.2), for each $1 \leq t \leq T$, one defines the θ-model's forecast of W_t from W_s, $1 \leq s \leq t-1$ to be $A_T^M(\theta) X_t^M + \sum_{j=0}^{t-2} (-\theta_{j+1})(W_{t-1-j} - A_T^M(\theta) X_{t-1-j}^M)$, with the convention $\sum_{j=0}^{-1} = 0$. This is the special case $W_{t|t-1}^M(\theta, \theta, T)$ of the more general forecast function $W_{t|t-1}^M(\theta, \theta^*, T)$ defined in (1.6), which follows. Conditional MLE estimates θ^T leading to GLS estimates $A_T^M(\theta^T)$ are the minimizers

$$\theta^T = \arg\min_{\theta \in \bar{\Theta}} \frac{1}{T} \sum_{t=1}^{T} (W_t - W_{t|t-1}^M(\theta, \theta, T))^2, \tag{1.4}$$

where $\bar{\Theta}$ is a compact set of θ specified by ARMA(p,q) models whose AR and MA polynomials have all zeroes in $\{|z| \geq 1 + \varepsilon\}$, for some $\varepsilon > 0$.

Responding to the extensive literature comparing GLS with OLS, we also consider model estimates and forecasts based on the OLS estimate of A^M,

$$A_T^M = \sum_{t=1}^{T} W_t X_t^{M\prime} \left[\sum_{t=1}^{T} X_t^M X_t^{M\prime} \right]^{-1}. \tag{1.5}$$

This is the special case $A_T^M(\theta^*)$ of (1.3) with $\theta^* = (1, 0, 0, \ldots)$, the white noise model for y_t^M. The forecast function of W_t associated with A_T^M is obtained by using this choice of θ^* in

$$W_{t|t-1}^M(\theta, \theta^*, T) = A_T^M(\theta^*) X_t^M + \sum_{j=0}^{t-2} (-\theta_{j+1})(W_{t-1-j} - A_T^M(\theta^*) X_{t-1-j}^M). \tag{1.6}$$

With this formula, for any fixed θ^*, conditional MLE yields a specification $\theta^{*T} = \arg\min_{\theta \in \bar{\Theta}} T^{-1} \sum_{t=1}^{T} (W_t - W_{t|t-1}^M(\theta, \theta^*, T))^2$.

In this paper, we obtain formulas for the limiting values of average squared one-step-ahead prediction errors obtained from these two types of MLEs,

$$\lim_{T \to \infty} \min_{\theta \in \bar{\Theta}} T^{-1} \sum_{t=1}^{T} (W_t - W_{t|t-1}^M(\theta, \theta, T))^2 \tag{1.7}$$

and, for fixed θ^*,

$$\lim_{T\to\infty} \min_{\theta\in\bar{\Theta}} T^{-1} \sum_{t=1}^{T} (W_t - W_{t|t-1}^{M}(\theta,\theta^*,T))^2. \tag{1.8}$$

With Theorems 5.1 and 7.1, which are given later in the paper, we show, under general assumptions on X_t and X_t^M given subsequently, that (1.7) is always less than or equal to (1.8), typically less. This is the optimality property of GLS referred to in the title of this paper. (By contrast, in the correct regressor case, when all our assumptions hold except (2.9) requiring asymptotic non-negligibility of the omitted regressors, the two limits are equal.) Further, using OLS with the white noise model $\theta^* = (1,0,0,\ldots)$ for y_t^M, as is often done, usually leads to even worse forecasts, in the sense that $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}(W_t - W_{t|t-1}^{M}(\theta^*,\theta^*,T))^2$ has a larger value than (1.8); see Section 7.1.1.

1.1. Overview of the Paper

The regressor sequence X_t, $t \geq 1$ is required to satisfy the conditions of Grenander (1954), which define a property we call scalable asymptotic stationarity; see Section 2 and Appendix B. Grenander introduced this generalization of stationarity to investigate the efficiency of OLS estimates for a large class of nonstochastic regressors in models with a broad range of weakly stationary disturbances. We indicate in Section 7.2 why efficiency in Grenander's sense is rarely applicable in the context of misspecified nonstochastic regressors. For the models we consider, the regressor X_t^M in (1.2), which can be stochastic, is taken to be a proper subvector of X_t. The remaining entries of X_t can be those of any vector X_t^N, compatible with our assumptions, whose variables compensate for the inadequacies of X_t^M in such a way that, for some A^M and A^N, the regression function in (1.1) can be decomposed as

$$AX_t = A^M X_t^M + A^N X_t^N. \tag{1.9}$$

Then, in (1.2), $y_t^M = A^N X_t^N + y_t$.

Our requirements for X_t^M, X_t^N, and y_t are comprehensively stated in Section 2 and verified for some important classes of models in Sections 2.1 and 6.1.1. More information about ARMA model parameterization with innovations coefficient sequences $\theta = (1,\theta_1,\theta_2,\ldots)$ is provided in Section 3, which includes some elementary examples. For diagonal scaling matrices $D_{M,T}$ such that a.s.-$\lim_{T\to\infty} D_{M,T}\sum_{t=1}^{T} X_t^M X_t^{M\prime} D_{M,T}$ is nonsingular, Theorem 4.1 gives a formula for $\lim_{T\to\infty}\ (A_T^M(\theta) - A^M)T^{-1/2}D_{M,T}^{-1}$ and establishes that convergence is uniform on the compact sets $\bar{\Theta}$ defined in Appendix A. For a given θ, this limit is called the *asymptotic bias characteristic* of $A_T^M(\theta)$ for A^M. Section 5 obtains formulas for the limits of the sample second moments of the forecast errors $W_t - W_{t|t-1}^{M}(\theta,\theta,T)$ and $W_t - W_{t|t-1}^{M}(\theta,\theta^*,T)$. The analogous results for regARIMA-type nonstationary models, for situations in which the disturbance process requires a differencing transformation prior to ARMA modeling, are discussed in Section 6. We describe, in Theorem 7.1 in Section 7, how the optimality property of GLS mentioned previously arises: the better performance of GLS relative to OLS occurs when the OLS estimate has an asymptotic bias characteristic different from that of the GLS estimate. These results provide support for an imputation procedure used by Statistics Netherlands (Aelen, 2004), which uses one-step-ahead forecasts from regARIMA models with stochastic distributed lag regressors to impute the net contribution of late-reporting firms to economic time series from certain monthly surveys; see Section 6.1. Section 7.1 provides elementary expressions for some asymptotic quantities associated with GLS and OLS estimation when y_t^M is modeled as a first-order autoregression. These are used to illustrate the generality of GLS's optimality. Section 8 discusses related results and extensions.

Proofs of the theorems are given in Appendix E. They use the auxiliary results of Appendix D obtained mainly from Findley, Pötscher, and Wei (2001).

2. THE DATA AND REGRESSOR ASSUMPTIONS

In (1.1), we require y_t, $t \geq 1$ to be *asymptotically stationary* (A.S.) in the sense of Pötscher (1987), meaning that for each $k = 0, \pm 1, \ldots$, the lag k sample second moments have asymptotic limits almost surely (i.e., with probability one), denoted a.s. That is, the limits

$$\gamma_k^y = \lim_{T\to\infty} \frac{1}{T} \sum_{t=|k|+1}^{T-|k|} y_{t+k} y_t \quad \text{a.s.} \tag{2.1}$$

exist. (By convention, $\sum_{t=a}^{b} = 0$, if $a < b$.) From a well-known result of Herglotz, the sequence of *asymptotic lag k second moments* γ_k^y has a spectral distribution function $G_y(\lambda)$ such that $\gamma_k^y = \int_{-\pi}^{\pi} e^{-ik\lambda}\, dG_y(\lambda)$ for $k = 0, \pm 1, \ldots$.

We require X_t, $t \geq 1$ in (1.1) to be *scalably asymptotically stationary* (S.A.S.), meaning that the limits

$$\Gamma_k^X = \lim_{T\to\infty} D_{X,T} \sum_{t=|k|+1}^{T-|k|} X_{t+k} X_t' D_{X,T} \quad \text{a.s.}, \qquad k = 0, \pm 1, \ldots \tag{2.2}$$

exist, where the $D_{X,T}$ are diagonal scaling matrices, $D_{X,T} = diag(d_{1,T}, \ldots, d_{\dim X,T})$, which are positive definite, decrease to zero ($D_{X,T} \searrow 0$), and satisfy $\lim_{T\to\infty} D_{X,T+k}^{-1} D_{X,T} = I_X$ for each $k \geq 0$. Here I_X is the identity matrix of order $\dim X$. (Ordinary convergence is meant in (2.2) if no coordinate of X_t is stochastic.) The resulting sequence Γ_k^X has a spectral distribution matrix function $G_X(\lambda)$: $\Gamma_k^X = \int_{-\pi}^{\pi} e^{-ik\lambda}\, dG_X(\lambda)$ for $k = 0, \pm 1, \ldots$; see Appendix B for further background, including examples.

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

Partition X_t as

$$X_t = \begin{bmatrix} X_t^M \\ X_t^N \end{bmatrix}, \tag{2.3}$$

where, as in the Introduction, the superscript N designates regressors *not* in the model (1.2). Let the corresponding partition of A in (1.1) be $A = [A^M \quad A^N]$ and let those of $D_{X,T}$, Γ_k^X, and $G_X(\lambda)$ be, respectively,

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

$$\Gamma_k^X = \begin{bmatrix} \Gamma_k^{MM} & \Gamma_k^{MN} \\ \Gamma_k^{NM} & \Gamma_k^{NN} \end{bmatrix}, \qquad G_X(\lambda) = \begin{bmatrix} G^{MM}(\lambda) & G^{MN}(\lambda) \\ G^{NM}(\lambda) & G^{NN}(\lambda) \end{bmatrix}. \tag{2.4}$$

From $D_{X,T} \searrow 0$, we have

$$D_{M,T} \searrow 0. \tag{2.5}$$

We require Γ_0^{MM} to be positive definite,

$$\Gamma_0^{MM} > 0, \tag{2.6}$$

and restrict X_t^N to being A.S.,

$$\Gamma_k^N = \lim_{T\to\infty} \frac{1}{T} \sum_{t=|k|+1}^{T-|k|} X_{t+k}^N X_t^{N\prime} \quad \text{a.s.,} \qquad k = 0, \pm 1, \dots . \tag{2.7}$$

Of course, (2.7) is equivalent to $D_{N,T} = T^{-1/2} I_N$, with I_N the identity matrix of order $\dim X^N$. We exclude omitted regressors of larger order, e.g., t^p with $p > 0$, because they yield unbounded y_t^M dominated by $A^N X_t^N$, which would clearly reveal the inadequacy of X_t^M with large enough T.

Further, the two series y_t and X_t must be *asymptotically orthogonal*, meaning that

$$\lim_{T\to\infty} T^{-1/2} \sum_{t=|k|+1}^{T-|k|} y_{t+k} X_t' D_{X,T} = 0 \quad \text{a.s.,} \qquad k = 0, \pm 1, \dots . \tag{2.8}$$

Finally, to keep the focus on the incorrect regressor situation, we assume that

$$A^N \Gamma_0^N A^{N\prime} > 0. \tag{2.9}$$

In summary, our assumptions concerning (1.1) are (2.1), (2.2), and (2.5)–(2.9).

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

2.1. Consequences of (2.1), (2.8), and (2.9) for y_t and y_t^M

First we note that, when X_t contains an entry equal to 1 for all t, then the corresponding scaling factor in $D_{X,T}$ can be taken to be $T^{-1/2}$, and so (2.8) yields $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} y_t = 0$ a.s. In this sense, y_t in (1.1) can be thought of as an asymptotically mean zero process. A similar result holds for the disturbances $y_t^M = A^N X_t^N + y_t$ of the misspecified model (1.2); see Section 4.

Now we establish the asymptotic stationarity of the y_t^M. From the requirement (2.7) that X_t^N be A.S. and from (2.1) and (2.8), for each k, $\gamma_k^M = \lim_{T\to\infty} T^{-1}\sum_{t=|k|+1}^{T-|k|} y_{t+k}^M y_t^M$ is given by

$$\gamma_k^M = A^N \Gamma_k^{NN} A^{N\prime} + \gamma_k^y = \int_{-\pi}^{\pi} e^{-ik\lambda}\, dG_{y^M}(\lambda), \tag{2.10}$$

where $G_{y^M}(\lambda) = A^N G^{NN}(\lambda) A^{N\prime} + G_y(\lambda)$. From (2.9), we have $\gamma_0^M > 0$. (The term γ_0^y can be zero.)

Finally, we note that, except in special situations such as that of Section 7.2, the disturbances and regressors in (1.2) will be asymptotically correlated, meaning $\lim_{T\to\infty} T^{-1/2}\sum_{t=|k|+1}^{T-|k|} y_{t+k}^M X_t^{M\prime} D_{M,T} = A^N \Gamma_k^{NM} \neq 0$ for some k, which will usually cause $A_T^M(\theta)$ defined in (1.3) to be biased asymptotically for some θ; see Theorem 4.1.

2.2. Sufficient Conditions for (2.1) and (2.8)

The properties (2.1) and (2.8) hold under reasonably general assumptions on y_t and X_t. The verification of (2.8) for a common type of stochastic regression model is discussed in Section 6.1.1. Here we consider the case in which y_t is weakly stationary with mean zero and X_t is nonstochastic with $\Gamma_0^X > 0$. Then, for almost sure convergence in (2.1) and (2.8), it suffices to have $y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$, with $\sum_{j=1}^{\infty} j b_j^2 < \infty$ for some independent white noise process ε_t such that $\sup_t E|\varepsilon_t|^r < \infty$ with $r > 2$ if y_t has a bounded spectral density or, if the spectral density of y_t is unbounded but square integrable, with $r > 4$; see Section 3.1 of Findley et al. (2001).

3. THE θ-PARAMETERIZATION OF ARMA MODELS

Three features of our ARMA model situation may be new to readers not familiar with the vein of research literature of which the papers by Pötscher (1987, 1991) are representative: (a) the disturbances y_t^M, $1 \le t \le T$ are not required to have means or covariances but only the asymptotic stationarity property; (b) no ARMA model is assumed to be correct in the sense of being able to exactly model the asymptotic lagged second moment sequence (2.10); (c) the ARMA coefficients of a model envisioned as $\phi(L) y_t^M = \alpha(L) a_t$ are replaced by the innovations filter $\theta = (1, \theta_1, \theta_2, \ldots)$ defined by the property that $\theta(z) = \sum_{j=0}^{\infty} \theta_j z^j$ satisfies $\theta(z) = \phi(z)/\alpha(z)$ for $|z| < 1$. In this section, we provide some orienting discussion and examples.

We assume that $\alpha(z) \neq 0$ for all $|z| \leq 1$, i.e., that the model is invertible. When y_t^M is weakly stationary with mean zero and defined for all t, then there always exists a weakly stationary series $a_t = a_t(\theta)$ such that the preceding ARMA model formula holds, namely, $a_t = \sum_{j=0}^{\infty} \theta_j y_{t-j}^M$. When y_t^M is only A.S. and defined only for $t \geq 1$, we define $a_t[\theta] = \sum_{j=0}^{t-1} \theta_j y_{t-j}^M$, $t \geq 1$. This series is A.S. with asymptotic lag k second moment given by $\gamma_k^a(\theta) = \int_{-\pi}^{\pi} e^{-ik\lambda} |\theta(e^{i\lambda})|^2 \, dG_{y^M}(\lambda)$, with $G_{y^M}(\lambda)$ as in (2.10); see (ii) of Proposition D.1 in Appendix D. We would call the θ-model correct if the white noise property, $\gamma_k^a(\theta) = 0$ for $k \neq 0$, obtains or, equivalently, if $\gamma_k^M = \sigma^2 \int_{-\pi}^{\pi} e^{-ik\lambda} |\theta(e^{i\lambda})|^{-2} \, d\lambda$ for all k for some $\sigma^2 > 0$. However, our theorems do not require any model for y_t^M, $t \geq 1$ to be correct in this sense.

For subsequent discussions, it will be useful to have in mind the θ's of some simple ARMA models. As was indicated in Section 1, a white noise model has $\theta = (1,0,0,\ldots)$. For the invertible ARMA(1,1) model, $(1 - \phi L) y_t^M = (1 - \alpha L) a_t$, with $|\alpha|, |\phi| < 1$, one has $\theta_j = \alpha^{j-1}(\alpha - \phi)$, $j \geq 1$. For AR(1) and MA(1) models, we have $\theta = (1, -\phi, 0, 0, \ldots)$ and $\theta = (1, \alpha, \alpha^2, \ldots)$, respectively.

Model parameterization by θ is useful because the θ's that are determined by likelihood-maximizing ARMA coefficients have uniquely defined large-sample limits in situations where the ARMA coefficients themselves do not, because of common zeroes in limiting AR and MA polynomials. For example, when an ARMA(1,1) model is fitted to white noise, the sequence of maximum likelihood pairs (ϕ^T, α^T) has multiple limit (or cluster) points, all on the line $\{(\alpha, \alpha) : |\alpha| \leq 1\}$; see Hannan (1982). However, when $\phi = \alpha$ for an ARMA(1,1) model, then $\theta = (1,0,0,\ldots)$, and so this is the only limit point of the filter sequence θ^T defined by the maximum likelihood estimates ϕ^T, α^T. That is, $\theta^T \to \theta$ a.s. coordinatewise, i.e., $\theta_j^T \to \theta_j$ a.s., $j \geq 0$.

As in the preceding examples, the coordinates of θ are always continuous functions of the ARMA coefficients. The converse holds only if the ARMA model is identifiable, i.e., the AR and MA polynomials have no common zero; also see the Appendix of Pötscher (1991) for additional background on the θ-parameterization. (Pötscher's parameter is the coefficient sequence of $\tilde{\theta}(z) = \alpha(z)/\phi(z)$. The relationship between θ and $\tilde{\theta}$ is continuous and invertible; see Section 3 of Findley, Pötscher, and Wei, 2004.)

To obtain the uniform convergence and continuity properties needed to establish the results indicated in the Introduction, ARMA(p,q), model coefficient estimation is restricted to compact sets of AR and MA coefficient vectors whose polynomials have all zeroes in $\{|z| \geq 1 + \varepsilon\}$ for some $\varepsilon > 0$. Such sets specify compact sets $\bar{\Theta}$ of the type discussed in Appendix A.

4. UNIFORM CONVERGENCE OF GLS ESTIMATES

We now present a fundamental convergence property of the $A_T^M(\theta)$ defined in (1.3). A generalized inverse is to be used in (1.3) when the inverse matrix fails to exist. This can (with probability one when X_t^M is stochastic) only happen for a finite number of T values, because of (2.6) and (iv) of Proposition D.1 in Appendix D. For any matrix M, define $\|M\| = \lambda_{\max}^{1/2}(MM')$, with $\lambda_{\max}(\cdot)$ denoting the maximum eigenvalue. If M is a vector with real coordinates $m_1, \ldots, m_n$, then $\|M\| = (\sum_1^n m_i^2)^{1/2}$.

Partition $\Gamma_0^X(\theta) = \int_{-\pi}^{\pi} |\theta(e^{i\lambda})|^2 \, dG_X(\lambda)$ analogously to (2.4), 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 $\Gamma_0^{MM}(\theta) = \int_{-\pi}^{\pi} |\theta(e^{i\lambda})|^2 \, dG^{MM}(\lambda)$, etc. For θ from an invertible model, define

$$C^{NM}(\theta) = \Gamma_0^{NM}(\theta)\Gamma_0^{MM}(\theta)^{-1}. \tag{4.1}$$

In Appendix E, we prove the theorem that follows.

THEOREM 4.1. *Let $\bar{\Theta}$ be a compact set of models as described in Appendix A. Under the assumptions (2.1), (2.2), and (2.5)–(2.8), we have, uniformly on $\bar{\Theta}$,*

$$\lim_{T\to\infty} (A_T^M(\theta) - A^M)T^{-1/2}D_{M,T}^{-1} = A^N C^{NM}(\theta) \quad a.s. \tag{4.2}$$

The function $C^{NM}(\theta)$ is continuous on $\bar{\Theta}$ and thus bounded there, $\max_{\theta\in\bar{\Theta}} \|C^{NM}(\theta)\| < \infty$.

For a given θ, $\lim_{T\to\infty} (A_T^M(\theta) - A^M)T^{-1/2}D_{M,T}^{-1} = A^N C^{NM}(\theta)$ is called the *asymptotic bias characteristic* of $A_T^M(\theta)$ for A^M. It is nonzero for some θ if $\Gamma_k^{NM} \neq 0$ for some k, i.e., if the series $A^N X_t^N$ and X_t^M are asymptotically correlated. When $D_{M,T} = T^{-1/2}$, then $A^N C^{NM}(\theta)$ is the asymptotic bias of $A_T^M(\theta)$ for A^M. Omitted variable bias is a fundamental modeling issue; see, e.g., Stock and Watson (2002, pp. 143–149). Section 7 will show that, when $A^N C^{NM}(\theta)$ varies with θ, there is usually an optimal value of $A^N C^{NM}(\theta)$ for one-step-ahead forecasting that is determined by the θ^T sequence of (1.4).

If X_t^M has one or more coordinates that are A.S., then for any $\check{A}^M$ that differs from A^M only in these coordinates we have, uniformly on $\bar{\Theta}$,

$$\lim_{T\to\infty} (A_T^M(\theta) - \check{A}^M)T^{-1/2}D_{M,T}^{-1} = A^N C^{NM}(\theta) + (A^M - \check{A}^M) \quad \text{a.s.} \tag{4.3}$$

This reveals the important fact that the asymptotic bias characteristic associated with an alternative omitted-regressor decomposition, $AX_t = \check{A}^M X_t^M + \check{X}_t^N$ with $\check{X}_t^N = A^N X_t^N + (A^M - \check{A}^M)X_t^M$, differs from the right-hand side of (4.2) by a term that is independent of θ.

Except in special situations, e.g., when the omitted regressors are precisely known, there is always ambiguity concerning X_t^N and A^M. However, it is useful to note that if a coordinate $X_{i,t}^M$ of X_t^M is constant with value one, then $\bar{X}^N = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T X_t^N$ can be assumed to be zero: by defining $\check{A}^M$ to differ from A^M in that $\check{A}_i^M = A_i^M + A^N\bar{X}^N$ replaces A_i^M, and by defining $\check{X}_t^N = X_t^N - \bar{X}^N$, one obtains $AX_t = \check{A}\check{X}_t + A^N\check{X}_t^N$ with $\lim_{T\to\infty} T^{-1}\sum_{t=1}^T \check{X}_t^N = 0$. Then, for $\check{y}_t^M = A^N\check{X}_t^N + y_t$ we have $\lim_{T\to\infty} T^{-1}\sum_{t=1}^T \check{y}_t^M = 0$.

5. UNIFORM ASYMPTOTIC STATIONARITY OF FORECAST ERRORS

We consider sample second moments of the errors of the one-step-ahead forecasts $W_{t|t-1}^M(\theta,\theta^*,T)$ from (1.6). For $1 \le t \le T$, the forecast errors $W_t - W_{t|t-1}^M(\theta,\theta^*,T)$ are observable and equal to $W_t[\theta] - A_T^M(\theta^*)X_t^M[\theta]$, which yields

$$W_t - W_{t|t-1}^M(\theta,\theta^*,T) = y_t[\theta] + \{AX_t - A_T^M(\theta^*)X_t^M\}[\theta], \qquad 1 \le t \le T. \tag{5.1}$$

Thus, setting $U_t(T) = [y_t \quad T^{1/2}D_{M,T}X_t^M \quad X_t^N]'$, $1 \le t \le T$ and $\beta_T(\theta^*) = [1 \quad (A^M - A_T^M(\theta^*))T^{-1/2}D_{M,T}^{-1} \quad A^N]$, we have

$$W_t - W_{t|t-1}^M(\theta,\theta^*,T) = \beta_T(\theta^*)U_t[\theta](T), \qquad 1 \le t \le T. \tag{5.2}$$

Let $\bar{\Theta}^*$ be a compact set in the sense of Appendix A. For $\beta(\theta^*) = [1 \quad -A^N C^{NM}(\theta^*) \quad A^N]$, Theorem 4.1 yields

$$\sup_{\theta^*\in\bar{\Theta}^*} \|\beta(\theta^*)\| < \infty, \qquad \sup_{\theta^*\in\bar{\Theta}^*} \|\beta_T(\theta^*) - \beta(\theta^*)\| \to 0 \quad \text{a.s.} \tag{5.3}$$

This fact and the properties of the $U_t(T)$ array described in Appendix C lead to the following theorem, which is proved in Appendix E. Define

$$B^{NM}(\theta^*) = A^N[-C^{NM}(\theta^*) \quad I_N] \tag{5.4}$$

and

$$G_{M,\theta^*}(\lambda) = G_y(\lambda) + B^{NM}(\theta^*)G_X(\lambda)B^{NM}(\theta^*)'. \tag{5.5}$$

For any $\bar{\Theta}, \bar{\Theta}^*$, let $\bar{\Theta}\times\bar{\Theta}^*$ denote the Cartesian product set $\{(\theta,\theta^*) : \theta \in \bar{\Theta}, \theta^* \in \bar{\Theta}^*\}$ and define convergence $(\theta^T,\theta^{*T}) \to (\theta,\theta^*)$ in $\bar{\Theta}\times\bar{\Theta}^*$ to mean $\theta_j^T \to \theta_j$ and $\theta_j^{*T} \to \theta_j^*$ for all $j \ge 0$.

THEOREM 5.1. *Let $\bar{\Theta}$ and $\bar{\Theta}^*$ be compact sets of models as described in Appendix A. Under the assumptions (2.1), (2.2), and (2.5)–(2.8), the forecast-error arrays $W_t - W_{t|t-1}^M(\theta,\theta^*,T)$, $1 \le t \le T$ are continuous on $\bar{\Theta}\times\bar{\Theta}^*$ and also jointly uniformly A.S. there. Specifically, for each $k = 0, \pm 1, \ldots,$ as $T \to \infty$, with*

$$\Gamma_k^M(\theta,\theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda}|\theta(e^{i\lambda})|^2\, dG_{M,\theta^*}(\lambda), \tag{5.6}$$

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

for $G_{M,\theta^*}(\lambda)$ *as in (5.5), the limits*

$$\frac{1}{T}\sum_{t=|k|+1}^{T-|k|}(W_{t+k}-W^M_{t+k|t+k-1}(\theta,\theta^*,T))(W_t-W^M_{t|t-1}(\theta,\theta^*,T))\to\Gamma^M_k(\theta,\theta^*) \tag{5.7}$$

hold uniformly a.s. on $\bar{\Theta}\times\bar{\Theta}^*$. *Further, the functions* $\Gamma^M_k(\theta,\theta^*)$ *are continuous and uniformly bounded on* $\bar{\Theta}\times\bar{\Theta}^*$. *Also, from (5.7) and (5.1), for given* θ *and* θ^*, *the values of* $\Gamma^M_k(\theta,\theta^*)$ *depend only on the values of the series* AX_t, X^M_t *and* $y_t = W_t - AX_t$, *not on the specification of the compensating regressor* X^N_t *in decompositions* $AX_t = A^M X^M_t + A^N X^N_t$ *(see Sect. 4).*

Theorem 5.1 shows that the quantities $\Gamma^M_0(\theta,\theta^*)$ are of special interest because they describe limiting average squared one-step-ahead forecast errors. With

$$\gamma^y_0(\theta)=\int_{-\pi}^{\pi}|\theta(e^{i\lambda})|^2\,dG_y(\lambda), \tag{5.8}$$

(5.5) yields the decomposition

$$\Gamma^M_0(\theta,\theta^*)=\gamma^y_0(\theta)+B^{NM}(\theta^*)\Gamma^X_0(\theta)B^{NM}(\theta^*)'. \tag{5.9}$$

By specializing the argument used to establish Theorem 5.1, $\gamma^y_0(\theta)$ is seen to be the limiting average squared error of the θ-model's one-step-ahead forecast of W_t when $X^M_t = X_t$. Similarly, using (4.2), the final quantity in (5.9) is seen to be the limit of the average of the squares of one-step-ahead forecast errors of the regression-function error array $AX_t - A^M_T(\theta^*)X^M_t$, $1\le t\le T$,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(\{AX_t-A^M_T(\theta^*)X^M_t\}[\theta])^2$$
$$=B^{NM}(\theta^*)\left[\int_{-\pi}^{\pi}|\theta(e^{i\lambda})|^2\,dG_X(\lambda)\right]B^{NM}(\theta^*)'\quad\text{a.s.} \tag{5.10}$$

It follows from the results for $k=0$ in Theorem 5.1 by standard arguments (see Pötscher and Prucha, 1997, Ch. 3 and Lem. 4.2) that the conditional maximum likelihood estimators θ^T of (1.4) converge a.s. to the compact set $\bar{\Theta}_0$ of minimizers of $\Gamma^M_0(\theta,\theta)$ over $\bar{\Theta}$,

$$\theta^T\to\bar{\Theta}_0\quad\text{a.s.} \tag{5.11}$$

That is, on a set of realizations of the random variables in (1.1) with probability one, the limit point of each (coordinatewise) convergent subsequence of $\theta^T, T\ge 1$ belongs to $\bar{\Theta}_0$. (So if there is a unique minimizer $\bar{\theta}$, then $\theta^T\to\bar{\theta}$ a.s.) Equivalently, in terms of the l^1-norm (see Appendix A), $\lim_{T\to\infty}\min_{\theta\in\bar{\Theta}_0}\|\theta^T-\theta\|_1=0$ a.s.

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

Similarly, the conditional maximum likelihood estimators θ^{*T} associated with $A_T^M(\theta^*)$ for fixed $\theta^* \in \bar{\Theta}$ converge a.s. to the set of minimizers of $\Gamma_0^M(\theta,\theta^*)$, which usually does not include θ^*; see Section 7.1.1.

6. EXTENSION TO ARIMA DISTURBANCE MODELS

Now suppose the observed data are $\widetilde{W}_t$, $1-d \le t \le T$ from a time series of the form $\widetilde{W}_t = A\widetilde{X}_t + \tilde{y}_t$ to which a model of the form $\widetilde{W}_t = A^M\widetilde{X}_t^M + \tilde{y}_t^M$ is being fit. Suppose also that it has been correctly determined that the disturbances $\tilde{y}_t^M$ require "differencing" with an operator $\delta(L) = \sum_{j=0}^{d} \delta_j L^j$, whose zeroes are on the unit circle, to obtain residuals for which an ARMA model can be considered. The resulting model is called a regARIMA model for $\widetilde{W}_t$. Such models are extensively used for seasonal time series in the context of seasonal adjustment (see Findley et al., 1998; Peña, Tiao, and Tsay, 2001), often with $\delta(L) = (1-L)(1-L^s), s = 4,12$. We assume that (2.1), (2.2), and (2.5)–(2.8) hold for $W_t = \delta(L)\widetilde{W}_t$, $y_t = \delta(L)\tilde{y}_t$, $X_t = \delta(L)\widetilde{X}_t$, and $X_t^M = \delta(L)\widetilde{X}_t^M$ and that X_t^M is a subvector of X_t. For any $1 \le t \le T$, because $\widetilde{W}_t = W_t - \sum_{j=1}^{d} \delta_j \widetilde{W}_{t-j}$, for given θ and θ^* a natural one-step-ahead forecast for $\widetilde{W}_t$ is $\widetilde{W}_{t|t-1}^M(\theta,\theta^*,T) = W_{t|t-1}^M(\theta,\theta^*,T) - \sum_{j=1}^{d} \delta_j \widetilde{W}_{t-j}$, with $W_{t|t-1}^M(\theta,\theta^*,T)$ defined by (1.6). This leads to $\widetilde{W}_t - \widetilde{W}_{t|t-1}^M(\theta,\theta^*,T) = W_t - W_{t|t-1}^M(\theta,\theta^*,T)$ for $1 \le t \le T$ and therefore to forecast-error limiting results as in Theorem 5.1 with the same functions $\Gamma_k^M(\theta,\theta^*)$.

6.1. Forecasting a Stochastic Regressor to Impute Values for Late Survey Responders

We briefly consider an application involving regARIMA models with stochastic regressors. Section 3.3 of Aelen (2004) provides an interesting one-step-ahead forecasting application involving a variety of seasonal time series $\widetilde{W}_t$ whose values come from enterprises that report economic data to Statistics Netherlands a month late, and $\widetilde{X}_t^M$ includes the sum of the values for month t from all enterprises of the same type that report on time, i.e., in the desired month, and sometimes also lagged values of these sums. Thus $\widetilde{X}_t^M$ is stochastic. In conjunction with the following discussion of distributed lag models, Theorem 5.1 and Theorem 7.1 in Section 7 provide theoretical support for Aelen's use of the regARIMA model GLS estimation and one-step-ahead forecasting procedures of X-12-ARIMA (Findley et al., 1998) to obtain Statistics Netherlands' imputed value for $\widetilde{W}_t$ in the month in which $\widetilde{X}_t^M$ becomes available.

6.1.1. A Class of Distributed Lag Models Satisfying the Assumptions of Theorem 5.1. After differencing, Aelen's model becomes a distributed lag model with regressors and correlated disturbances that are both treated as

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

stationary. We consider a broad class of such models. Suppose that W_t and Z_t are jointly covariance stationary variates with zero means and that the spectral density matrix of Z_t is Hermitian positive definite at all frequencies. Then, when the autocovariance sequence Γ_k^V of $V_t = [W_t \quad Z_t']'$ satisfies $\sum_{k=-\infty}^{\infty} \|\Gamma_k^V\| < \infty$, there exist coefficients A_k satisfying $\sum_{k=-\infty}^{\infty} \|A_k\| < \infty$ such that $W_t = \sum_{k=-\infty}^{\infty} A_k Z_{t-k} + y_t$ holds, with $Ey_t Z_{t-k}' = 0$, $k = 0, \pm 1, \ldots$; see Theorem 8.3.1 of Brillinger (1975). For any $m, n \geq 0$, setting $X_t^M = [Z_{t+n}' \ldots Z_{t-m}']'$, $X_t^N = \sum_{k \neq -n, \ldots, m} A_k Z_{t-k}$, $A^M = [A_{-n} \ldots A_m]$, and $A^N = 1$ leads to (1.1) and (1.2) having the form of a distributed lag model with stationary disturbances (see, e.g., Stock and Watson, 2002) and to the assumptions of Theorem 5.1 holding under Gaussianity or weaker assumptions on V_t; see Theorem IV.3.6 of Hannan (1970).

7. OPTIMALITY OF GLS

Because of the uniform convergence and continuity results established in Theorem 5.1, for any compact $\bar{\Theta}$ as described in Appendix A, we have

$$\min_{\theta \in \bar{\Theta}} \left\{ \frac{1}{T} \sum_{t=1}^{T} (W_t - W_{t|t-1}^M(\theta, \theta, T))^2 \right\} \to \min_{\theta \in \bar{\Theta}} \Gamma_0^M(\theta, \theta) \quad \text{a.s.,} \tag{7.1}$$

and, for any fixed $\theta^* \in \bar{\Theta}$,

$$\min_{\theta \in \bar{\Theta}} \left\{ \frac{1}{T} \sum_{t=1}^{T} (W_t - W_{t|t-1}^M(\theta, \theta^*, T))^2 \right\} \to \min_{\theta \in \bar{\Theta}} \Gamma_0^M(\theta, \theta^*) \quad \text{a.s.} \tag{7.2}$$

In Appendix E, we establish the theorem that follows.

THEOREM 7.1. *Let $\bar{\Theta}$ be a compact set as described in Appendix A and suppose that (2.1), (2.2), and (2.5)–(2.8) hold. Then for any fixed $\theta^* \in \bar{\Theta}$,*

$$\min_{\theta \in \bar{\Theta}} \Gamma_0^M(\theta, \theta) \leq \min_{\theta \in \bar{\Theta}} \Gamma_0^M(\theta, \theta^*), \tag{7.3}$$

with equality holding if and only if a minimizer $\bar{\theta}^*$ of $\Gamma_0^M(\theta, \theta^*)$ over $\bar{\Theta}$ is always a minimizer of $\Gamma_0^M(\theta, \theta)$,

$$\Gamma_0^M(\bar{\theta}^*, \bar{\theta}^*) = \min_{\theta \in \bar{\Theta}} \Gamma_0^M(\theta, \theta), \tag{7.4}$$

and, simultaneously, the asymptotic bias characteristic of $A_T^M(\bar{\theta}^*)$ as an estimator of A^M coincides with that of $A_T^M(\theta^*)$,

$$A^N C^{NM}(\bar{\theta}^*) = A^N C^{NM}(\theta^*). \tag{7.5}$$

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

As a consequence, strict inequality obtains in (7.3) if and only if

$$A^N C^{NM}(\theta^*) \neq A^N C^{NM}(\bar{\theta}) \tag{7.6}$$

holds for every minimizer $\bar{\theta}$ of $\Gamma_0^M(\theta,\theta)$ over $\bar{\Theta}$. For the maximum likelihood estimators θ^T of (1.4), this condition implies

$$\lim \inf_{T\to\infty} \|(A_T^M(\theta^T) - A_T^M(\theta^*))T^{-1/2}D_{M,T}^{-1}\| > 0 \quad a.s. \tag{7.7}$$

Conversely, if $\Gamma_0^M(\theta,\theta)$ has a unique minimizer $\bar{\theta}$, then (7.7) implies (7.6).

Unless θ^* is a minimizer of $\Gamma_0^M(\theta,\theta)$, we expect that both $\min_{\theta\in\bar{\Theta}} \Gamma_0^M(\theta,\theta) < \Gamma_0^M(\bar{\theta}^*,\bar{\theta}^*)$ and $A^N C^{NM}(\bar{\theta}^*) \neq A^N C^{NM}(\theta^*)$ will hold except in quite special situations, the only one known to us being when $A^N C^{NM}(\theta^*)$, and therefore also $\Gamma_0^M(\theta,\theta^*)$, does not depend on θ^*. In Section 7.1, this is shown to occur with AR(1) models for y_t^M only in a singular situation. Otherwise $\bar{\theta}^*$ is unique. Whenever $\bar{\theta}^*$ is unique, failure of (7.5), which implies $\min_{\theta\in\bar{\Theta}} \Gamma_0^M(\theta,\theta) < \Gamma_0^M(\bar{\theta}^*,\theta^*)$ and $\bar{\theta}^* \neq \theta^*$, also yields $\Gamma_0^M(\bar{\theta}^*,\theta^*) < \Gamma_0^M(\theta^*,\theta^*)$.

Model sets $\bar{\Theta}$ usually include the white noise model $\theta^* = (1,0,0,\dots)$ as a degenerate case. Hence the conclusions of Theorem 7.1 are generally applicable to OLS as an alternative to GLS. They indicate the following optimality property of GLS: In conjunction with maximum likelihood estimation of θ, *asymptotically, OLS estimation is never better than GLS estimation for one-step-ahead forecasting. When the regressor is underspecified and $A^N C^{NM}(\theta)$ is nonconstant, OLS will typically have greater average mean square error than GLS, for large enough T, because of its asymptotic bias characteristics being different from that of GLS.*

Thursby (1987) provides comparisons of OLS and GLS biases when y_t is known to be independent and identically distributed (i.i.d.) (white noise), $\dim X_t^M = 2$, $\dim X_t^N = 1$, the coordinates of X_t are correlated first-order AR processes, and the loss function is the posterior mean squared bias associated with a prior for the parameters that determine the covariance structure between X_t^N and X_t^M. With the aid of numerical integrations for the GLS quantities, he establishes that, depending on the choice of the autocovariance structure of X_t^M, the mean squared asymptotic bias of GLS is sometimes less and sometimes greater than that of OLS. Theorem 7.1 shows that, for either outcome, GLS has an asymptotic advantage over OLS for one-step-ahead forecasting.

7.1. Examples Involving AR(1) Models and dim X_t^M = dim X_t^N = 1

The condition (7.5) is the easiest to investigate, because, for AR models, $\bar{\theta}^*$ is the solution of a linear system of equations. For simplicity, we consider only the case in which $\dim X_t^M = \dim X_t^N = 1$ and a first-order AR model, i.e.,

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

$\theta = \theta(\phi) = (1,-\phi,0,0,\ldots)$, is used for the disturbance series y_t^M in (1.2). From (5.8) and (5.9), this leads to

$$\Gamma_0^M(\theta,\theta^*) = \int_{-\pi}^{\pi} |1-\phi e^{i\lambda}|^2\, dG_y(\lambda) + B^{NM}(\theta^*) \times \int_{-\pi}^{\pi} |1-\phi e^{i\lambda}|^2\, dG_X(\lambda) B^{NM}(\theta^*)', \quad \textbf{(7.8)}$$

where $\int_{-\pi}^{\pi}|1-\phi e^{i\lambda}|^2 dG_y(\lambda) = (1+\phi^2)\gamma_0^y - 2\phi\gamma_1^y$ and $\int_{-\pi}^{\pi}|1-\phi e^{i\lambda}|^2 dG_X(\lambda) = (1+\phi^2)\Gamma_0^X - \phi(\Gamma_1^X + \Gamma_{-1}^X)$. Also, with $\theta^* = (1,-\phi^*,0,\ldots)$, the $C^{NM}(\theta^*)$ component of $B^{NM}(\theta^*)$ is

$$C^{NM}(\theta^*) = \frac{(1+\phi^{*2})\Gamma_0^{NM} - \phi^*(\Gamma_1^{NM}+\Gamma_{-1}^{NM})}{(1+\phi^{*2})\Gamma_0^{MM} - 2\phi^*\Gamma_1^{MM}}.$$

When

$$2\Gamma_0^{NM}\Gamma_1^{MM} - (\Gamma_1^{NM}+\Gamma_{-1}^{NM})\Gamma_0^{MM} \neq 0, \quad \textbf{(7.9)}$$

the derivative of $C^{NM}(\theta^*)$ is nonzero on $-1 < \phi^* < 1$ and $C^{NM}(\theta^*)$ is strictly monotonic; see Section 6.3 in the paper by Findley (2005), whose derivation also shows that the unique $\bar{\theta}^* = (1,-\bar{\phi}^*,0,\ldots)$ minimizing (7.8) is the lag one autocorrelation of $G_{M,\theta^*}(\lambda)$ in (5.5),

$$\bar{\phi}^* = \frac{\gamma_1^y + (A^N)^2\{\Gamma_1^{NN} + (C^{NM}(\theta^*))^2\Gamma_1^{MM} - C^{NM}(\theta^*)(\Gamma_1^{NM}+\Gamma_{-1}^{NM})\}}{\gamma_0^y + (A^N)^2\{\Gamma_0^{NN} - (C^{NM}(\theta^*))^2\Gamma_0^{MM}\}}. \quad \textbf{(7.10)}$$

There is no such simple formula for $\bar{\phi}$ minimizing $\Gamma_0^M(\theta,\theta)$ because the critical point equation for $\bar{\phi}$ provides $\bar{\phi}$ as a zero of a polynomial of degree five in general. However, from strict monotonicity of $C^{NM}(\theta^*(\phi^*))$, if $\bar{\phi}^* \neq \phi^*$ then (7.5) fails, and therefore strict inequality holds in (7.3) by Theorem 7.1. For the OLS choice, $\phi^* = 0$, when $C^{NM}(\theta^*) = C^{NM}$, (7.10) shows that $\bar{\phi}^* \neq 0$ (except possibly at a single value of $(A^N)^2$), when either γ_1^y or $\Delta^{NM} = \Gamma_1^{NN} + (C^{NM})^2\Gamma_1^{MM} - C^{NM}(\Gamma_1^{NM}+\Gamma_{-1}^{NM})$ is nonzero, which will usually be the case. A periodic X_t satisfying (7.9) and $\Delta^{NM} \neq 0$ is given in Section 7.1.2.

When (7.9) fails, $C^{NM}(\theta^*) = C^{NM} = \Gamma_0^{NM}/\Gamma_0^{MM}$ for all θ^*, and so equality holds in (7.3).

7.1.1. The Inferiority of White Noise Modeling with OLS when $\bar{\phi}^* \neq 0$. If $\bar{\Theta}$ is a compact model set containing the AR(1) models $\theta = \theta(\phi)$, then $\Gamma_0^M(\bar{\theta}^*,\theta^*) \leq \Gamma_0^M(\theta^*(\bar{\phi}^*),\theta^*)$. So, under (7.9) and $\bar{\phi}^* \neq \phi^*$, we have, from (7.3), that $\min_{\theta\in\bar{\Theta}} \Gamma_0^M(\theta,\theta) \leq \Gamma_0^M(\bar{\theta}^*,\theta^*) < \Gamma_0^M(\theta^*,\theta^*)$. Thus, for $\theta^* = (1,0,0,\ldots)$, it follows from (7.1) and (7.2) that when $\bar{\phi}^* \neq 0$, using OLS

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

estimation of A^M with the white noise model for y_t^M leads to asymptotically worse one-step-ahead forecasts than GLS with (1.4), for any such model set $\bar{\Theta}$.

7.1.2. Periodic X_t and an Example of Δ^{NM}. The trading day and holiday regressors discussed in Findley et al. (1998), Bell and Hillmer (1983), and Findley and Soukup (2000) are effectively periodic functions; i.e., $X_{t+P}^M = X_t^M$ holds for all t, for rather large periods P (e.g., $12 \times 28 = 336$ months for trading day regressors, $12 \times 19 = 228$ months for some lunar holiday regressors, more for other holidays, e.g., Easter). The simplest holiday regressors are one-dimensional and specify that the effect of the holiday is the same for each day in some interval near the holiday, a dubious but simplifying assumption. For such regressors, the compensating X_t^N can be assumed to be one-dimensional and have the same period.

Every regressor of period P has a Fourier representation $\sum_j \alpha_j \cos(2\pi jt/P) + \beta_j \sin(2\pi jt/P)$ with at most P nonzero coefficients, which are uniquely determined linear functions of P consecutive values of the regressor; see Section 4.2.3 of Anderson (1971). To give a more complete analysis of (7.3) for the function (7.8), we consider a simplified period $P = 4$ regressor X_t^M having the representation $X_t^M = a_1^M \cos(\pi/2)t + a_2^M(-1)^t$, with $a_1^M, a_2^M \neq 0$, for which $X_t^N = a_1^N \cos(\pi/2)t + b_1^N \sin(\pi/2)t$, with $a_1^N, b_1^N \neq 0$. Thus $X_t = [X_t^M \quad X_t^N]' = \alpha_1 \cos(\pi/2)t + \alpha_2(-1)^t + \beta_1 \sin(\pi/2)t$, where $\alpha_1 = [a_1^M \quad a_1^N]$, $\alpha_2 = [a_2^M \quad 0]$, and $\beta_1 = [0 \quad b_1^N]$. Consequently, $\Gamma_k^X = \frac{1}{2}\alpha_1'\alpha_1 \cos(\pi/2)k + \alpha_2'\alpha_2(-1)^k + \frac{1}{2}\beta_1'\beta_1 \sin(\pi/2)k$, $k = 0, \pm 1, \dots$, and $G_X(\lambda)$ is piecewise constant with upward jumps at $\lambda = \pm\pi/2, \pi$; see Anderson (1971, p. 581).

For this X_t, the left-hand side of (7.9) has the value $-a_1^M a_1^N (a_2^M)^2$, and so (7.9) holds. Further, $C^{NM} = a_1^M a_1^N \{(a_1^M)^2 + 2(a_2^M)^2\}^{-1}$ and $\Delta^{NM} = -(a_2^M C^{NM})^2$. Strict inequality holds in (7.3) for OLS estimation except when $\gamma_1^y > 0$ and $(A^N)^2 = \gamma_1^y (a_2^M C^{NM})^{-2}$, in which case $\bar{\phi}^* = 0 = \phi^*$.

7.2. Regarding Asymptotic Efficiency in the Sense of Grenander (1954)

Here we restrict attention to nonrandom regressors X_t in (1.1) whose components are polynomials, periodic functions, or realizations of stationary processes with continuous spectral densities and with convergent sample second moments. The disturbance process y_t is assumed to be a mean zero stationary process with the last-mentioned properties. Grenander (1954) considers the correct regressor case and calls the OLS estimates $A_T^M = \sum_{t=1}^T W_t X_t' (\sum_{t=1}^T X_t X_t')^{-1}$ asymptotically efficient if $\lim_{T\to\infty} D_T^{-1} E\{(A_T - A)'(A_T - A)\} D_T^{-1}$ is minimal (in the ordering of symmetric matrices) among all linear, unbiased estimates A_T of A. For this situation, his result, given on p. 244 of Grenander and Rosenblatt (1984), is that OLS is efficient if and only if the spectral distribution function $G_X(\lambda)$ has at most $\dim X_t$ jumps and the sum of the ranks of the jumps

$G_X(\lambda+) - G_X(\lambda+)$, $0 \le \lambda \le \pi$ is equal to $\dim X_t$. These conditions are not satisfied, and OLS is not efficient, for most of the regressors discussed in Section 7.1.2, including the calendar effect regressors and the period four regressor with $b^N \neq 0$; see Chapter 7.7 and case (1) on p. 253 of Grenander and Rosenblatt (1984): usually, the number of terms in the Fourier representation of X_t, and thus also the number of jumps in $G_X(\lambda)$, exceeds $\dim X_t$.

To be able to apply Grenander's result to our underspecified regression situation, assume that X_t^M and y_t^M have the properties hypothesized previously for X_t and y_t. Thus X_t^N has a continuous spectral density and so cannot have periodic components. If we consider X_t^M having only polynomial and periodic components, then X_t^N and X_t^M are asymptotically orthogonal; see Section 6.1 of Findley (2005). This implies $A^N C^{NM}(\theta^*) = 0$ for all θ^*, resulting in equality in (7.3) always, because $\Gamma_0^M(\theta, \theta^*)$ does not depend on θ^*.

On the other hand, with regressors in X_t^M that are realizations of stationary processes, if $A^N C^{NM}(\theta^*)$ is nonzero, then the analogue for $A_T^M(\theta^*)$ of Grenander's efficiency measure fails by being infinite, because some entries of $(A_T^M(\theta^*) - A^M)D_{M,T}^{-1}$ will have order $T^{1/2}$; see (4.2).

Thus this concept of efficiency is not useful in our context.

8. EXTENSIONS AND RELATED RESULTS

From their connection to one-step-ahead forecast error filters, it is not very surprising that GLS estimates of regARMA and regARIMA models have an optimality property for one-step-ahead forecasting. Yet a systematic investigation of the topic has been lacking. A pleasingly simple result, such as Theorem 7.1's connection of optimality with asymptotic bias characteristics, seems possible only for the incorrect regressor case. Indeed, if asymptotic efficiency results are indicative, the correct regressor case will be quite complex. In this case, when the ARMA model for y_t is incorrect, GLS can be more or less efficient than OLS; see Koreisha and Fang (2001). Even when the ARMA model is also correct, the analysis and examples of Grenander and Rosenblatt (1984) and of Section 7.2 show, for nonstochastic regressors, that OLS is asymptotically efficient only for a limited range of relatively simple regressors.

For any fixed θ^*, in the incorrect nonstochastic regressor case, a referee conjectures that, under additional assumptions and with the aid of a result like Theorem 4.1 of West (1996), it can be shown that the limit as $T \to \infty$ of the variance of $T^{-1/2} \sum_{t=1}^{T} (W_t - W_{t|t-1}^M(\theta^*, \theta^*, T))$ does not depend on θ^*.

So far, we have only provided asymptotic results for the most simply defined GLS estimates, which are obtained by truncating the infinite-past forecast error filters and using conditional maximum likelihood estimation of the ARMA model. Section 2.4 of Findley (2005) and (d) of Lemma 10 of Findley (2005) reveal that the same limits are obtained if the errors of the *finite-past* one-step-ahead forecasts discussed in Newton and Pagano (1983) are used to define GLS estimates in conjunction with unconditional maximum likelihood estimation of the ARMA model. (Analogous GLS estimates from AR models were considered in Amemiya, 1973.) See Section 9 of the technical report Findley (2003) for additional details, including details about how to weaken the assumptions on X_t^M to include the frequently used intervention variables of Box and Tiao (1975). These decay exponentially to zero and so have weight one in $D_{M,T}$, causing (2.5) to fail. Also, with the restriction to measurable minimizers θ^T discussed in Findley et al. (2001, 2004), in the case of nonstochastic X_t, all almost sure convergence results hold with convergence in probability when convergence in (2.1) holds only in this weaker sense.

Findley (2003) also shows how to use the results of Appendix D to generalize Theorem 5.1 to the case of multi-step-ahead forecast errors and to establish the convergence of θ-parameter estimates that minimize average squared multi-step-ahead forecast errors (allowing for y_t^M the more comprehensive model classes of Findley et al., 2004).

Findley (2005) uses the results of Theorems 4.1 and 7.1 to obtain formulas and GLS optimality results for the limiting average of squared *out-of-sample* (real time) forecast errors of regARIMA models under assumptions on the regressors X_t that are slightly more restrictive than those of Section 2 but are satisfied by all of the specific regressor types we have mentioned. The limit formulas are the same as those of the present paper when X_t^M is A.S. Empirical results are available from the author showing that GLS usually leads to better one-step-ahead out-of-sample forecasting performance than OLS for a suite of monthly series that are modeled with trading day and Easter holiday regressors by the U.S. Census Bureau for the purpose of seasonal adjustment.

REFERENCES

Aelen, F. (2004) Improving Timeliness of Industrial Short-Term Statistics Using Time Series Analysis. Statistics Netherlands, Division of Technology and Facilities, Methods and Informatics Discussion paper 04005. Statistics Netherlands. http://www.cbs.nl/NR/rdonlyres/EC77EDFA-1579-4AF6-9763-8EFEEB9F4CC5/0/discussionpaper04005.pdf.

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

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. Wiley.

Bell, W.R. & S.C. Hillmer (1983) Modelling time series with calendar variation. *Journal of the American Statistical Association* 78, 526–534.

Box, G.E.P. & G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, rev. ed. Holden-Day.

Box, G.E.P. & G.C. Tiao (1975) Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* 70, 70–79.

Brillinger, D.R. (1975) *Time Series*. Holt, Rinehart and Winston.

Findley, D.F. (2003) Convergence of Estimates of Misspecified regARIMA Models and Generalizations and the Optimality of Estimated GLS Regression Coefficients for One-Step-Ahead Forecasting. Statistical Research Division Research Report Statistics 2003-03. U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/rrs2003-06.pdf.

Findley, D.F. (2005) Asymptotic stationarity properties of out-of-sample forecast errors of misspecified regARIMA models and the optimality of GLS for one-step-ahead forecasting. *Statistica Sinica* 15, 447–476.

Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto, & B.C. Chen (1998) New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business & Economic Statistics* 16, 127–177 (with discussion).

Findley, D.F., B.M. Pötscher, & C.Z. Wei (2001) Uniform convergence of sample second moments of time series arrays. *Annals of Statistics* 29, 815–838.

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

Findley, D.F. & R.J. Soukup (2000) Modeling and model selection for moving holidays. In *2000 Proceedings of the Business and Economic Statistics Section of American Statistical Association*, pp. 102–107. American Statistical Association. Also http://www.census.gov/ts/papers/asa00_eas.pdf.

Grenander, U. (1954) On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics* 25, 252–272.

Grenander, U. & M. Rosenblatt (1984) *Time Series*, 2nd ed. Chelsea.

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

Hannan, E.J. (1982) Testing for autocorrelation and Akaike's criterion. In J. Gani & E.J. Hannan (eds.), *Essays in Statistical Science, Papers in Honour of P.A.P. Moran*, pp. 403–412. Applied Probability Trust.

Koreisha, S.G. & Y. Fang (2001) Generalized least squares with misspecified correlation structure. *Journal of the Royal Statistical Society, Series B* 63, 515–532.

Newton, H.J. & M. Pagano (1983) The finite memory prediction of covariance stationary time series. *SIAM Journal of Scientific and Statistical Computation* 4, 330–339.

Peña, D., G.E. Tiao, & R.S. Tsay (2001) *A Course in Time Series Analysis*. Wiley.

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

Pötscher, B.M. (1987) Convergence results for maximum likelihood type estimators in multivariate ARMA models. *Journal of Multivariate Analysis* 21, 29–52.

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

Pötscher, B.M. & I.R. Prucha (1997) *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer-Verlag.

Stock, J.H. & M.W. Watson (2002) *Introduction to Econometrics*. Addison-Wesley.

Thursby, J.G. (1987) OLS or GLS in the presence of misspecification error? *Journal of Econometrics* 35, 359–374.

West, K.D. (1996) Asymptotic inference about predictive ability. *Econometrics* 64, 1067–1084.

APPENDIX A. Compact θ-Sets for Estimation

For each $\varepsilon > 0$ and integer pair $p, q \geq 0$, we define $\Theta_{p,q,\varepsilon}$ to be the set all of $\theta = (1, \theta_1, \theta_2, \ldots)$ from invertible ARMA(r,s) models with $r \leq p$, $s \leq q$ such that the zeroes of the minimal degree AR and MA polynomials $\phi(z)$ and $\alpha(z)$ such that $\theta(z) = \phi(z)/\alpha(z)$ all belong to $\{|z| \geq 1 + \varepsilon\}$. Every sequence $\theta^T = (1, \theta_1^T, \theta_2^T, \ldots)$, $T = 1, 2, \ldots$ in $\Theta_{p,q,\varepsilon}$ has a subsequence $\theta^{S(T)}$ that converges coordinatewise to some $\theta \in \Theta_{p,q,\varepsilon}$, i.e., $\theta_j^{S(T)} \to \theta_j$, $j \geq 1$. Thus $\Theta_{p,q,\varepsilon}$ is compact for coordinatewise convergence. Further, for $0 \leq \varepsilon_0 < \varepsilon$, the sums $\sum_{j=0}^{\infty} (1 + \varepsilon_0)^j |\theta_j|$ converge uniformly on $\Theta_{p,q,\varepsilon}$; i.e.,

$\sup_{\theta\in\Theta_{p,q,\varepsilon}} \sum_{j=0}^{\infty}(1+\varepsilon_0)^j|\theta_j| < \infty$ and $\lim_{J\to\infty}\sup_{\theta\in\Theta_{p,q,\varepsilon}} \sum_{j=J}^{\infty}(1+\varepsilon_0)^j|\theta_j| = 0$. See Lemmas 2 and 10 of Findley (2005) for these and other properties mentioned. Our uniform convergence results that are presented subsequently follow from these facts as do some other important properties. First, the functions $\theta(e^{i\lambda}) = \sum_{j=0}^{\infty}\theta_j e^{i\lambda}$ are continuous on $-\pi \le \lambda \le \pi$ and uniformly bounded and bounded away from zero on $\Theta_{p,q,\varepsilon}$:

$$\min_{-\pi\le\lambda\le\pi,\,\theta\in\Theta_{p,q,\varepsilon}} |\theta(e^{i\lambda})| > 0, \qquad \max_{-\pi\le\lambda\le\pi,\,\theta\in\Theta_{p,q,\varepsilon}} |\theta(e^{i\lambda})| < \infty.$$

Second, if a sequence θ^T, $T = 1,2,\ldots$ in $\Theta_{p,q,\varepsilon}$ converges coordinatewise to some θ, then it also converges in the stronger sense that $\lim_{T\to\infty}\sum_{j=0}^{\infty}(1+\varepsilon_0)^j|\theta_j^T - \theta_j| = 0$ whenever $0 \le \varepsilon_0 < \varepsilon$. In particular, the topology of coordinatewise convergence on $\Theta_{p,q,\varepsilon}$ coincides with that of the l^1-norm $\|\theta\|_1 = \sum_{j=0}^{\infty}|\theta_j|$.

Our theorems apply to any compact $\bar{\Theta}$ for which $\bar{\Theta} \subseteq \Theta_{p,q,\varepsilon}$ holds, for some $\varepsilon > 0$ and $p,q \ge 0$. A typical $\bar{\Theta}$ would arise from constraints on the zeroes of the AR and MA polynomials of the kind of ARMA model of interest.

APPENDIX B. Scalable Asymptotic Stationarity

Under the data assumptions made in Section 2, X_t and y_t in (1.1) together form a multivariate sequence that is S.A.S., a property we now consider in some detail. Let U_t, $t \ge 1$ be a real-valued column vector sequence that is S.A.S. and let I_U denote the identity matrix of order dim U, the dimension of U_t. Thus there is a decreasing sequence $D_1 \ge D_2 \ge \ldots$ of positive definite diagonal matrices, for which $D_T \searrow 0$ and

$$\lim_{T\to\infty} D_{T+k}^{-1}D_T = I_U, \qquad k = 1,2,\ldots \tag{B.1}$$

hold, such that, for each $k = 0, \pm 1, \ldots$, the limits

$$\Gamma_k^U = \lim_{T\to\infty} D_T \sum_{t=|k|+1}^{T-|k|} U_{t+k}U_t' D_T \quad \text{a.s.} \tag{B.2}$$

exist (finitely). The properties (B.1) and (B.2) yield $\lim_{T\to\infty} D_T U_{T-j} = 0$ a.s., $j \ge 0$. For example, when $j = 0$, as $T \to \infty$,

$$D_T U_T U_T' D_T = D_T \sum_{t=1}^{T} U_t U_t' D_T - (D_T D_{T-1}^{-1}) D_{T-1} \sum_{t=1}^{T-1} U_t U_t' D_{T-1}(D_{T-1}^{-1} D_T)$$

converges a.s. to $\Gamma_0^U - \Gamma_0^U = 0$, whence $D_T U_T \to 0$ a.s. Further, $D_T \searrow 0$ leads to $\lim_{T\to\infty} D_T U_{1+j} = 0$ a.s. for all $j \ge 0$.

Without a formal name, this generalization of stationarity was introduced for regressors in Grenander (1954) to encompass a variety of nonstochastic regressors, including polynomials. (Our notation is the inverse of his, using D_T where he uses D_T^{-1}. He only requires the diagonal elements of Γ_0^U to be positive, which is the nature of (2.9) for $\check{X}_t^N = A^N X_t^N$. Our requirement (2.10) for X_t^M is stronger.) Grenander shows that the real matrix sequence Γ_k^U, $k = 0, \pm 1, \ldots$ has a representation $\Gamma_k^U = \int_{-\pi}^{\pi} e^{-ik\lambda}\, dG_U(\lambda)$ in which $G_U(\lambda)$ is a Hermitian-matrix-valued function such that the eigenvalues of increments

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

$G_U(\lambda_2) - G_U(\lambda_1)$, $\lambda_2 \geq \lambda_1$, are nonnegative, or, equivalently, the increments are Hermitian nonnegative; see also Grenander and Rosenblatt (1984), Chapter II of Hannan (1970), and Chapter 10 of Anderson (1971). For example, if $U_t = t^p$, $p \geq 0$, then, with $D_T = T^{-(p+1/2)}$, one obtains $\Gamma_k^U = (2p+1)^{-1}$ for each k, and so $G_U(\lambda)$ can be taken to be 0 for $\lambda < 0$ and $(2p+1)^{-1}$ for $\lambda \geq 0$. Grenander (1954) and Grenander and Rosenblatt (1984, Ch. 7) verify the joint scalable asymptotic stationarity property for regressors whose entries $X_{i,t}$ are polynomials, linear combinations (perhaps infinite) of sinusoids, i.e., of $\cos \omega_j t$ and/or $\sin \omega_j t$, for various $0 \leq \omega_j \leq \pi$ (scaling sequence $T^{-1/2}$), and, finally, products of polynomials t^p with linear combinations of sinusoids (scaling sequence $T^{-p-1/2}$). By contrast, exponentially increasing regressors, e.g., $U_t = e^{bt}$ with $b > 0$, are not S.A.S. because (B.1) fails for $D_T = (\sum_{t=1}^{T} e^{2bt})^{-1/2}$; see Hannan (1970, p. 77).

APPENDIX C. Vector Array Reformulation of Assumptions

The following reformulation of our assumptions (2.1), (2.2), and (2.5)–(2.9) concerning y_t and X_t will enable us to make use of the results of Findley et al. (2001, 2004). The vector array

$$U_t(T) = \begin{bmatrix} y_t \\ T^{1/2} D_{X,T} X_t \end{bmatrix} = \begin{bmatrix} y_t \\ T^{1/2} D_{M,T} X_t^M \\ X_t^N \end{bmatrix}, \qquad 1 \leq t \leq T, \tag{C.1}$$

is A.S. More specifically, for each $k = 0, \pm 1, \ldots,$

$$\Gamma_k^U = \lim_{T\to\infty} T^{-1} \sum_{t=|k|+1}^{T-|k|} U_{t+k}(T) U_t(T)' \quad \text{a.s.} \tag{C.2}$$

$$= \begin{bmatrix} \gamma_k^y & 0 & 0 \\ 0 & \Gamma_k^{MM} & \Gamma_k^{MN} \\ 0 & \Gamma_k^{NM} & \Gamma_k^{NN} \end{bmatrix}, \tag{C.3}$$

with $\Gamma_0^{MM} > 0$ and $A^N \Gamma_0^{NN} A^{N\prime} > 0$. Further, from Appendix B,

$$\lim_{T\to\infty} T^{-1/2} U_{1+j}(T) = 0 = \lim_{T\to\infty} T^{-1/2} U_{T-j}(T) \quad \text{a.s.,} \qquad j \geq 0. \tag{C.4}$$

Because of (C.3), the spectral distribution matrix of the Γ_k^U sequence has the block diagonal form $G_U(\lambda) = blockdiag(G_y(\lambda), G_X(\lambda))$.

APPENDIX D. Uniform Convergence Results for Filtered A.S. Arrays

The proposition and lemma that follow are formulated for proving some of the more general results indicated in Section 8.

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

PROPOSITION D.1. *Let $U_t(T)$, $1 \le t \le T$ be an A.S. column vector array satisfying (C.4) and let $G_U(\lambda)$ denote the spectral distribution matrix of the asymptotic lagged second moments matrices Γ_k^U defined by (C.2). Let H and Z be sets of filters $\eta = (\eta_0, \eta_1, \ldots)$ and $\zeta = (\zeta_0, \zeta_1, \ldots)$ such that $\sum_{j=0}^{\infty} |\eta_j|$ resp. $\sum_{j=0}^{\infty} |\eta_j|$ converges uniformly on H resp. Z. Then the filter output arrays $U_t[\eta](T) = \sum_{j=0}^{t-1} \eta_j U_{t-j}$ and $U_t[\zeta](T) = \sum_{j=0}^{t-1} \zeta_j U_{t-j}$, $1 \le t \le T$, $\eta \in H$, $\zeta \in Z$ have the following properties:*

(i) $\lim_{T\to\infty} \sup_{\eta\in H} \|T^{-1/2} U_{1+j,T}[\eta]\| = \lim_{T\to\infty} \sup_{\eta\in H} \|T^{-1/2} U_{T-j,T}[\eta]\| = 0$ *a.s. for all $j \ge 0$, and analogously for $U_t[\zeta](T)$.*

(ii) As $T \to \infty$, $\sup_{\eta\in H, \zeta\in Z} \|T^{-1} \sum_{t=|k|+1}^{T-|k|} U_{t+k}[\eta](T) U_t[\zeta](T)' - \Gamma_k^U(\eta,\zeta)\| \to 0$ *a.s., where $\Gamma_k^U(\eta,\zeta) = \int_{-\pi}^{\pi} e^{-ik\lambda} \eta(e^{i\lambda}) \zeta(e^{-i\lambda})\, dG_U(\lambda)$, for $k = 0, \pm 1, \ldots$.*

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

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

and are jointly continuous in η,ζ in the sense that, if $\eta^T \in H$, $\zeta^T \in Z$ are such that $\eta^T \to \eta$ and $\zeta^T \to \zeta$ (coordinatewise convergence) with $\eta \in H$, $\zeta \in Z$, then $\Gamma_k^U(\eta^T,\zeta^T) \to \Gamma_k^U(\eta,\zeta)$. Also, if $Z = H$, then $\inf_{\eta\in H, -\pi\le\lambda\le\pi} |\eta(e^{i\lambda})|^2 \Gamma_0^U \le \Gamma_0^U(\eta,\eta) \le \sup_{\eta\in H, -\pi\le\lambda\le\pi} |\eta(e^{i\lambda})|^2 \Gamma_0^U$.

(iv) Let H be an index set for a family of arrays $U_t(\eta,T)$, $1 \le t \le T$, $\eta \in H$ such that, as $T \to \infty$,

$$\sup_{\eta\in H} \left\| \frac{1}{T} \sum_{t=1}^{T} U_t(\eta,T) U_t(\eta,T)' - \Gamma_0(\eta) \right\| \to 0 \quad a.s., \qquad \textbf{(D.1)}$$

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

$$\inf_{\eta\in H} \lambda_{\min}(\Gamma_0(\eta)) \ge m_H \qquad \textbf{(D.2)}$$

holds for some $m_H > 0$. Then

$$\sup_{\eta\in H} \left\| \left(\frac{1}{T} \sum_{t=1}^{T} U_t(\eta,T) U_t(\eta,T)' \right)^{-1} - \Gamma_0(\eta)^{-1} \right\| \to 0 \quad a.s. \qquad \textbf{(D.3)}$$

Proof. Parts (i)–(iii) are straightforward vector extensions of special cases of Theorem 2.1 and Proposition 2.1 of Findley et al. (2001). For (iv), it follows from (D.1) and (D.2) that, given $\varepsilon > 0$, for each realization except those of an event with probability zero, there is a T_ε such that for $T \ge T_\varepsilon$ the inequalities $\sup_{\eta\in H} \|T^{-1} \sum_{t=1}^{T} U_t(\eta,T) U_t(\eta,T)' - \Gamma_0(\eta)\| < (\varepsilon/2) m_H^2$ and $\inf_{\eta\in H} \lambda_{\min}(T^{-1} \sum_{t=1}^{T} U_t(\eta,T) U_t(\eta,T)') \ge \frac{1}{2} m_H$ hold. Hence for these T and all $\eta \in H$,

$$\sup_{\eta\in H}\left\|\left(\frac{1}{T}\sum_{t=1}^{T}U_t(\eta,T)U_t(\eta,T)'\right)^{-1}-\Gamma_0(\eta)^{-1}\right\|$$
$$\leq \sup_{\eta\in H}\left\{\left\|\left(\frac{1}{T}\sum_{t=1}^{T}U_t(\eta,T)U_t(\eta,T)'\right)^{-1}\right\|\right.$$
$$\left.\times\left\|\frac{1}{T}\sum_{t=1}^{T}U_t(\eta,T)U_t(\eta,T)'-\Gamma_0(\eta)\right\|\|\Gamma_0(\eta)^{-1}\|\right\}$$
$$\leq \frac{1}{m_H}\sup_{\eta\in H}\left\{\lambda_{\min}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}U_t(\eta,T)U_t(\eta,T)'\right)\right\}$$
$$\times\sup_{\eta\in H}\left\{\left\|\frac{1}{T}\sum_{t=1}^{T}U_t(\eta,T)U_t(\eta,T)'-\Gamma_0(\eta)\right\|\right\}<\varepsilon,$$

which establishes (D.3).

We also need the following lemma, whose proof can be obtained by standard arguments, as in the proof of (5.18) of Findley et al. (2004).

LEMMA D.2. *Suppose that, on a set $\bar{\Theta}^*$, the sequence $\beta_T(\theta^*)$, $T=1,2,\ldots$ of row vector functions converges uniformly a.s. to a bounded function $\beta(\theta^*)$, i.e., (5.3) holds, and similarly for $\tau_T(\theta^*)$, $T\geq 1$ and its limit $\tau(\theta^*)$. Let $U_t(\eta,T)$, $\eta\in H$ and $W_t(\zeta,T)$, $\zeta\in Z$, $1\leq t\leq T$ be families of column vector arrays of the same dimension as $\beta(\theta^*)$ and $\tau(\theta^*)$, respectively, such that, for $k=0,\pm1,\ldots,$*

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

holds for functions $\Gamma_k(\eta,\zeta)$ with $\sup_{\eta\in H,\zeta\in Z}\|\Gamma_0(\eta,\zeta)\|<\infty$. Then, as $T\to\infty$,

$$\sup_{\substack{\theta^*\in\bar{\Theta}^*\\ \eta\in H,\zeta\in Z}}\left\|\frac{1}{T}\sum_{t=|k|+1}^{T-|k|}\beta_T(\theta^*)U_{t+k}(\eta,T)W_t(\zeta,T)'\tau_T(\theta^*)'-\beta(\theta^*)\Gamma_k(\eta,\zeta)\tau(\theta^*)'\right\|$$
$$\to 0\quad a.s.$$ ■

APPENDIX E. Proofs

Proof of Theorem 4.1. We have

$$(A_T^M(\theta)-A^M)T^{-1/2}D_{M,T}^{-1}$$
$$=T^{-1/2}\sum_{t=1}^{T}y_t^M[\theta]X_t^M[\theta]'D_{M,T}\left(D_{M,T}\sum_{t=1}^{T}X_t^M[\theta]X_t^M[\theta]'D_{M,T}\right)^{-1}$$
$$=T^{-1/2}\sum_{t=1}^{T}y_t[\theta]X_t^M[\theta]'D_{M,T}\left(D_{M,T}\sum_{t=1}^{T}X_t^M[\theta]X_t^M[\theta]'D_{M,T}\right)^{-1}$$
$$+A^N\left(T^{-1/2}\sum_{t=1}^{T}X_t^N[\theta]X_t^M[\theta]'D_{M,T}\right)\left(D_{M,T}\sum_{t=1}^{T}X_t^M[\theta]X_t^M[\theta]'D_{M,T}\right)^{-1}.$$

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press

By (ii) and (iii) of Proposition D.1, $T^{-1/2}\sum_{t=1}^{T} y_t[\theta]X_t^M[\theta]'D_{M,T}$ converges uniformly a.s. to 0 and $T^{-1/2}\sum_{t=1}^{T} X_t^N[\theta]X_t^M[\theta]'D_{M,T}$ and $D_{M,T}\sum_{t=1}^{T} X_t^M[\theta]X_t^M[\theta]'D_{M,T}$ converge uniformly a.s. to the continuous limits $\Gamma_0^{NM}(\theta)$ and $\Gamma_0^{MM}(\theta)$, respectively, with $\Gamma_0^{MM}(\theta)$ bounded below by the positive definite matrix $m_{\bar{\Theta}}^2\Gamma_0^{MM}$, where $m_{\bar{\Theta}} = \min_{\pi\leq\lambda\leq\pi,\theta\in\bar{\Theta}}|\theta(e^{i\lambda})| > 0$; see Appendix A. It follows from (iv) of Proposition D.1 that $(D_{M,T}\sum_{t=1}^{T} X_t^M[\theta]X_t^M[\theta]'D_{M,T})^{-1}$ converges uniformly to $\Gamma_0^{MM}(\theta)^{-1}$, which is therefore continuous (and bounded above by $m_{\bar{\Theta}}^{-2}(\Gamma_0^{MM})^{-1}$). Hence $(A_T^M(\theta) - A^M)$ $T^{-1/2}D_{M,T}^{-1}$ converges uniformly a.s. to $A^N C^{NM}(\theta)$, which is continuous on $\bar{\Theta}$ and also bounded. ■

Proof of Theorem 5.1. The assertions follow from (5.2) and Lemma D.2 with $\tau_T(\theta^*) = \beta_T(\theta^*)$, $H = Z = \bar{\Theta}$, and $U_t(\theta,T) = W_t(\theta,T) = U_t[\theta](T) = \sum_{j=0}^{t-1}\theta_j U_{t-j}(T)$, for $U_{t-j}(T)$ defined by (C.1), because the uniform convergence of $T^{-1}\sum_{t=|k|+1}^{T-|k|} U_{t+k}[\theta](T)U_t[\theta](T)'$ to $\Gamma_k^U(\theta) = \int_{-\pi}^{\pi} e^{-ik\lambda}|\theta(e^{i\lambda})|^2\, dG_U(\lambda)$ and the boundedness of $\|\Gamma_0^U(\theta)\|$ on $\bar{\Theta}$, which are required to apply Lemma D.2, follow from (ii) and (iii), respectively, of Proposition D.1. The uniform convergence of $\sum_{j=0}^{\infty}|\theta_j|$ required by the proposition is the special case $\varepsilon_0 = 0$ in Appendix A. The fact that $G_U(\lambda) = \mathit{blockdiag}(G_y(\lambda), G_X(\lambda))$, because of (C.3), yields the form of $G_{M,\theta^*}(\lambda)$ in (5.5). ■

Proof of Theorem 7.1. We start by establishing that, for any invertible θ and θ^*, we have $\Gamma_0^M(\theta,\theta) \leq \Gamma_0^M(\theta,\theta^*)$ with equality holding if and only if $A^N C^{NM}(\theta^*) = A^N C^{NM}(\theta)$. Indeed, the component of $\Gamma_0^M(\theta,\theta^*)$ that depends on θ^* can be reexpressed in terms of the analogues of $C^{NM}(\theta^*)$ and $\Gamma_0^X(\theta)$ obtained by replacing X_t^N with $\check{X}_t^N = A^N X_t^N$. Denoting these analogues by $\check{C}^{NM}(\theta^*)$ and $\check{\Gamma}_0^X(\theta)$, we have

$$A^N[-C^{NM}(\theta^*) \quad I_N]\Gamma_0^X(\theta)[-C^{NM}(\theta^*) \quad I_N]'A^{N\prime}$$

$$= [-\check{C}^{NM}(\theta^*) \quad 1]\check{\Gamma}_0^X(\theta)[-\check{C}^{NM}(\theta^*) \quad 1]'.$$

By a standard calculation, for any C with the dimensions of $\check{C}^{NM}(\theta)$,

$$[-\check{C}^{NM}(\theta) \quad 1]\check{\Gamma}_0^X(\theta)[-\check{C}^{NM}(\theta) \quad 1]' \leq [-C \quad 1]\check{\Gamma}_0^X(\theta)[-C \quad 1]', \tag{E.1}$$

with equality holding in (E.1) if and only if $C = \check{C}^{NM}(\theta)$ $(= A^N C^{NM}(\theta))$.

Next, note that because $\Gamma_0^M(\theta,\theta)$ and $\Gamma_0^M(\theta,\theta^*)$ are continuous functions of θ on $\bar{\Theta}$, they have minimizers $\bar{\theta}$, resp. $\bar{\theta}^*$ over $\bar{\Theta}$. From the result just established, we obtain

$$\Gamma_0^M(\bar{\theta},\bar{\theta}) \leq \Gamma_0^M(\bar{\theta}^*,\bar{\theta}^*) \leq \Gamma_0^M(\bar{\theta}^*,\theta^*). \tag{E.2}$$

Thus $\Gamma_0^M(\bar{\theta},\bar{\theta}) = \Gamma_0^M(\bar{\theta}^*,\theta^*)$ holds, i.e., equality in (7.3), if and only if (7.4) and $\Gamma_0^M(\bar{\theta}^*,\bar{\theta}^*) = \Gamma_0^M(\bar{\theta}^*,\theta^*)$ do, and the latter is equivalent to (7.5), as was just shown.

In particular, equality in (7.3) implies the failure of (7.6) for $\bar{\theta} = \bar{\theta}^*$ satisfying (7.4). Conversely, failure of (7.6) for some minimizer $\bar{\theta}$, i.e., $A^N C^{NM}(\theta^*) = A^N C^{NM}(\bar{\theta})$, implies $\Gamma_0^M(\bar{\theta}^*,\theta^*) \leq \Gamma_0^M(\bar{\theta},\theta^*) = \Gamma_0^M(\bar{\theta},\bar{\theta})$, which, from (E.2), yields $\Gamma_0^M(\bar{\theta}^*,\theta^*) = \Gamma_0^M(\bar{\theta},\bar{\theta}) = \Gamma_0^M(\bar{\theta}^*,\bar{\theta}^*)$, i.e., equality in (7.3). Therefore (7.6) for all $\bar{\theta}$ minimizing $\Gamma_0^M(\theta,\theta)$ is necessary and sufficient for strict inequality in (7.3).

From Theorem 4.1 and (5.11), it follows that the left-hand side of (7.7) is equal a.s. to the left-hand side of

$$\liminf_{T\to\infty} \|A^N(C^{NM}(\theta^T) - C^{NM}(\theta^*))\| \geq \min_{\bar{\theta}\in\bar{\Theta}_0} \|A^N(C^{NM}(\bar{\theta}) - C^{NM}(\theta^*))\| \quad \text{a.s.} \tag{E.3}$$

The assertions concerning (7.7) follow from (E.3) and the fact that, when $\bar{\Theta}_0 = \{\bar{\theta}\}$, equality holds in (E.3) because $\theta^T \to \bar{\theta}$ a.s., from (5.11). ■

https://doi.org/10.1017/S0266466607070430 Published online by Cambridge University Press