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

David F. Findley, U.S. Census Bureau

Abstract

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). Similarly, the associated disturbance term is assumed to have sample second moments that converge with increasing series length, perhaps after differencings. The model’s regression component is taken to be underspecified, due perhaps to simplifications, approximations, parsimony, etc. Also, the model’s ARMA or ARIMA-type structure for the disturbance term need not be correct. Both Ordinary Least Squares and Generalized Least Squares estimates of the mean function are considered. An optimality property of GLS relative to OLS (and other alternatives) is obtained for one-step-ahead forecasting. Asymptotic bias characteristics of the regression estimates are shown to distinguish the forecasting performance. These results provide support for the application by Statistics Netherlands (Aelen, 2004) of regARIMA models with stochastic regressors to forecast/impute the net contribution of late-reporting firms to monthly economic time series from surveys.

1. INTRODUCTION

Univariate modeling of many economic indicator series requires specification of both a regression function with extrinsic variables as well as an autocovariance structure for the disturbance process. Suppose that, after making any needed variance stabilizing transformations (such as taking logarithms and then differencings), one has observations $W_t, 1 \leq t \leq T$ of a time series of the form

$$W_t = AX_t + y_t,$$  \hspace{1cm} (1.1)

where the $X_t$ are column vectors, and the $y_t$ are real variates, asymptotically orthogonal to $X_t$ in a sense to be defined, whose lagged sample second moments converge as $T \rightarrow \infty$. 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 or stochastic
variables. We examine the situation in which the modeler considers a model
\[ W_t = A^M X_t^M + y_t^M. \]  

whose regressor \( X_t^M \) is not able to reproduce \( AX_t \) for all \( t \), due to known or unknown omissions, approximations, simplifications, etc. The modeler proceeds as though, for some coefficient vector \( A \) to be estimated, the residual process \( y_t^M = W_t - A^M X_t^M \) has the autocovariance sequence of an autoregressive moving average (ARMA) model (which need not be correct). The resulting model for \( W_t \) is called a regARMA model.

Given a proposed invertible ARMA model \( \phi(L)y_t^M = \alpha(L)a_t \) for \( y_t^M \), with \( L \) denoting the lag (backshift) operator, let \( \theta = (1, \theta_1, \theta_2, \ldots) \) denote the coefficient sequence of the power series expansion \( \phi(L)/\alpha(L) = \sum_{j=0}^{\infty} \theta_j L^j \) (\( \theta_0 = 1 \)). Then \( y_{t|t-1}^M(\theta) = -\sum_{j=1}^{\infty} \theta_j y_{t-j}^M \) is the model’s linear forecast of \( y_t^M \) from \( y_s^M \), \( -\infty < s < t - 1 \) and \( u_t = y_t^M - y_{t|t-1}^M(\theta) = \sum_{j=0}^{\infty} \theta_j y_{t-j}^M \) is the associated forecast error (innovations) sequence; see Section 5.3.3 of Box and Jenkins (1976). All ARMA model quantities of interest in this paper depend only on \( \theta \), and so are not subject to the parameter-estimate convergence problems caused by common factors in \( \phi(L) \) and \( \alpha(L) \); see Appendix A of Pötscher (1991). With finitely many observations \( W_t, 1 \leq t \leq T \), setting \( W_t[\theta] = \sum_{j=0}^{t-1} \theta_j W_{t-j} \) and \( X_t[\theta] = \sum_{j=0}^{t-1} \theta_j X_{t-j} \), where \( ' \) denotes transpose, one can define a Generalized Least Squares (GLS) estimate of \( A^M \) for the \( \theta \) model by
\begin{equation}
A_T^M(\theta) = \left( \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} \right) , 
\end{equation}

see Pierce (1971). (We discuss another GLS estimate in Section 7.) With \( A_T^M(\theta) X_t^M \) providing a candidate model for the mean function, the ARMA coefficients leading to \( \theta \) can be determined by conditional or unconditional maximum (Gaussian) likelihood estimation (MLE). For the conditional estimates, on which we focus for simplicity, for given \( 1 \leq t \leq T \), one defines the \( \theta \)-model’s forecast of \( W_t \) from \( W_s, 1 \leq s \leq t - 1 \) to be \( W_{t|t-1}^M(\theta, \theta, T) = A_T^M(\theta) X_t^M + \sum_{j=0}^{t-2} (-\theta_{j+1}) \left( W_{t-1-j} - A_T^M(\theta) X_{t-1-j}^M \right) \), with \( \sum_{j=0}^{t-1} = 0 \). We focus on the case when conditional MLE specifies a \( \theta^T \) for which
\begin{equation}
\frac{1}{T} \sum_{t=1}^{T} \left( W_t - W_{t|t-1}^M(\theta, \theta, T) \right)^2 
\end{equation}
is minimized over a compact set \( \Theta \) specified by ARMA \((p, q)\) models whose AR and MA polynomial have all zeroes in \( \{|z| \geq 1 + \varepsilon\} \) for some \( \varepsilon > 0 \). Because of the extensive literature comparing GLS with OLS, we also focus on the Ordinary Least Squares estimate of \( A^M \),
\begin{equation}
A_T^M = \sum_{t=1}^{T} W_t X_t^M \left( \sum_{t=1}^{T} X_t^M X_t^M \right)^{-1} , 
\end{equation}

2
The regressors \( X \) are required to satisfy the conditions of Grenander (1954), which define a property we call scalable asymptotic stationarity (S.A.S.); 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, including polynomials, periodic functions, and stationary stochastic regressors. We will indicate in Subsection 6.2 why efficiency in Grenander’s sense is rarely applicable in the context of misspecified nonstochastic regressors.

For the misspecified models we consider, the regressor \( X_t \) in (1.2), which can be stochastic, is taken to be a proper subvector of \( X_t \). The remaining coordinate 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 \( A_M^N \) and \( A_M^N \), the regression component of \( W_t \) is given by \( A_M^N X_t^M + A_M^N X_t^N \). Thus, in (1.2), we have

\[
y_t^M = A_N^N X_t^N + y_t. \tag{1.7}
\]

Our additional requirements for \( X_t^M \), \( X_t^N \) and \( y_t \) are given in Section 2. and are verified for some important classes of models in Subsection 2.1. 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. Theorem 4.1 describes uniform limiting properties over compact sets \( \Theta \) of the error of GLS estimates, \( A_T^M(\theta) - A^M \). Section 5 obtains the limits functions of the sample second moments of the forecast errors of \( W_{t_{t-1}}^M(\theta, \theta, T) \) and \( W_{t_{t-1}}^M(\theta, \theta^*, T) \) uniformly over compact sets and also the analogous results for regARIMA-type nonstationary models. The latter cover situations in which the disturbance process requires differencing or similar transformations prior to fitting an ARMA model. When the sequences \( X_t^N \) and \( X_t^M \) are asymptotically correlated, we describe, in Theorem 6.1 of Section 6, how the

\[
W_{t_{t-1}}^M(\theta, \theta^*, T) = A_T^M(\theta^*) X_t^M + \sum_{j=0}^{t-2} (-\theta_{j+1}) (W_{t_{t-1}} - A_T^M(\theta^*) X_{t_{t-1}}). \tag{1.6}
\]
optimality property of GLS mentioned above arises: the better performance of GLS is shown to occur when the OLS estimate has an asymptotic bias characteristic different from the GLS estimate. These results provide support for the application by Statistics Netherlands (Aelen, 2004) of regARIMA models with stochastic regressors to forecast/impute the net contribution of late-reporting firms to monthly economic time series from surveys see Subsection 5.1. Subsection 6.1 provides elementary expressions for some asymptotic quantities associated with GLS and OLS estimation for the case in which \( y_t^M \) is modeled as a first-order autoregression. These are used to illustrate how generally GLS is uniquely optimal.

Proofs of the Theorems are presented in Appendix E and utilize a Proposition obtained mainly from Findley, Pötscher and Wei (2001) (hereafter FPW 2001) and a related Lemma. These auxiliary results are in Appendix D. Appendix A provides the precise characterization of the compact sets \( \tilde{\Theta} \).

2. THE DATA AND MISSPECIFIED REGRESSOR ASSUMPTIONS

In (1.1), we require \( y_t \) to be asymptotically stationary (A.S.), meaning that for each \( k = 0, \pm 1, \ldots \), lag \( k \) sample second moments have asymptotic limits either almost surely (i.e., with probability one) or in probability. That is, the limits

\[
\gamma^y_k = \lim_{T \to \infty} \frac{1}{T} \sum_{t=|k|+1}^{T-|k|} y_t y_{t+k} \text{ a.s. [i.p.]} \tag{2.1}
\]

exist (with \( \gamma^{-k}_y = \gamma^y_k \)). The sequence of asymptotic lag \( k \) second moments sequence \( \gamma^y_k \) has an asymptotic spectral distribution, denoted \( G_y(\lambda) \): \( \gamma^y_k = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_y(\lambda) \) for each \( k \).

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

\[
\Gamma^X_k = \lim_{T \to \infty} D_{X,T}^{-k} \sum_{t=1}^{T-k} X_{t+k} X_t' D_{X,T} \text{ a.s., } k \geq 0, \tag{2.2}
\]

with positive definite diagonal scaling matrices \( D_{X,T} = \text{diag}(d_{1,T}, \ldots, d_{\dim X,T}) \) that decrease to zero, \( D_{X,T} \searrow 0 \), and satisfy \( \lim_{T \to \infty} D_{T+k}^{-1} D_T = I_X \) for each \( k \geq 0 \), where \( I_X \) is the identity matrix of order \( \dim X \). Negatively lagged scaled sample second moments also converge: for \( k > 0 \), \( \Gamma^{-k}_X = \lim_{T \to \infty} D_{X,T} \sum_{t=k+1}^{T} X_{t-k} X_t' D_{X,T} = (\Gamma^X_k)' \text{ a.s.} \) Ordinary convergence is meant if no coordinate of \( X_t \) is stochastic. Partition \( X_t \) as

\[
X_t = \begin{bmatrix} X^M_t \\ X^N_t \end{bmatrix}, \tag{2.3}
\]

where, as in the Introduction, the superscript \( N \) designates the regressors not in the model. Let the corresponding partition of \( A \) in (1.1) be \( A = \begin{bmatrix} A^M & A^N \end{bmatrix} \), and those of \( D_{X,T} \), the
sequence $\Gamma^X_k$, and its asymptotic spectral distribution matrix $G_X(\lambda)$ be, respectively,

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

$$\Gamma^X_k = \begin{bmatrix} \Gamma_{k}^{MM} & \Gamma_{k}^{MN} \\ \Gamma_{k}^{NM} & \Gamma_{k}^{NN} \end{bmatrix}, \quad G_X(\lambda) = \begin{bmatrix} G^{MM}(\lambda) & G^{MN}(\lambda) \\ G^{NM}(\lambda) & G^{NN}(\lambda) \end{bmatrix}. \quad (2.4)$$

See Appendix B for more details and examples. From $D_{X,T} \setminus 0$, we have

$$D_{M,T} \setminus 0. \quad (2.5)$$

In addition to the S.A.S. property of $X_t$, we require $\Gamma_0^{MM}$ to be positive definite,

$$\Gamma_0^{MM} > 0, \quad (2.6)$$

and $X_t^N$ to be A.S.,

$$D_{N,T} = T^{-1/2}I_N, \quad (2.7)$$

where $I_N$ is the identity matrix of order $\dim X^N$. Omitted regressor variables of larger order, e.g. $t^p$ with $p > 1$, would yield $y_t^M$ that would be recognized as unsuitable for ARMA modeling with large enough $T$.

Our last requirement is that the two series $y_t$ and $X_t$ be asymptotically orthogonal, meaning

$$\lim_{T \to \infty} T^{-\frac{1}{2}} \sum_{t=1}^{T-k} y_t X_{t+k} D_{X,T} = 0 \text{ a.s. [i.p.], } k = 0, \pm 1, \ldots. \quad (2.8)$$

In statements like this, the applicable mode of convergence is to be understood as the mode applicable in (2.1), which is required to be a.s. when $X_t$ is stochastic.

In summary, our assumptions concerning (1.1) are (2.1)–(2.2) and (2.5)–(2.8). For any matrix $M$, define $\|M\| = \lambda_\text{max}^{1/2}(MM^t)$, with $\lambda_\text{max}(\cdot)$ denoting the maximum eigenvalue. If $M$ is a vector with real coordinates $m_1, \ldots, m_n$, then $\|M\| = (\sum_{i=1}^{n} m_i^2)^{1/2}$.

### 2.1. Discussion of (2.1) and (2.8)

Condition (2.1) is verified by Theorem IV.3.6 and Section 3.1 of FPW (2001) for a broad range of weakly stationary processes. This property is then inherited by any process that differs from such a weakly stationary process by a transient process whose asymptotic lag zero second moment is zero.

Turning to (2.8), we note first that if $X_t$ 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 (2.8) yields $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} y_t = 0 \text{ a.s. [i.p.]}$. In this sense, $y_t$ in (1.1) can be thought of as an asymptotically mean zero process. When $X_t$ is nonstochastic and $\Gamma^X_0 > 0$, the results of Appendix D of FPW (2001) readily
show that (2.8) holds a.s. when \( y_t \) is a weakly stationary linear processes, \( y_t = \sum b_j \varepsilon_{t-j} \), with independent white noise process \( \varepsilon_t \) such that \( \sup_t E |\varepsilon_t|^r < \infty \). It suffices that \( r > 2 \) if the spectral density of \( y_t \) is bounded, or that \( r > 4 \) if the spectral density is unbounded but square integrable. When \( y_t \) lacks such a representation but has a bounded spectral density, then i.p. convergence applies in (2.8); see Subsection 3.2 of FPW (2001).

For a fundamental example of (2.8) with stochastic \( X_t \) in our context, suppose \( W_t \) and \( Z_t \) are jointly covariance stationary variates with zero means such that the spectral density matrix of \( Z_t \) is Hermitian positive definite at all frequencies. Then \( W_t = \sum_{k=-\infty}^{\infty} A_k Z_{t-k} + y_t \) with \( E y_t Z_{t-k}' = 0 \), \( k = 0, \pm 1, \ldots \), with \( \sum_{k=-\infty}^{\infty} ||A_k|| < \infty \) when the autocovariance sequence \( \Gamma_k^Y \) of \( V_t = \begin{bmatrix} W_t & Z_t' \end{bmatrix}' \) satisfies \( \sum_{k=-\infty}^{\infty} ||\Gamma_k^Y|| < \infty \), see Theorem 8.3.1 of Brillinger (1975). (\( E \) denotes expectation.) For any \( m \geq 0 \), setting \( X_t^M = \begin{bmatrix} Z_t' & \cdots & Z_{t-m}' \end{bmatrix}' \), \( X_t^N = \sum_{k \neq 0, \ldots, m} A_k Z_{t-k} \), \( A^M = \begin{bmatrix} A_0 & \cdots & A_m \end{bmatrix} \) and \( A^N = 1 \) leads to (1.2) having the form of a \textit{distributed lag model}, see Stock and Watson (2002), and to (1.7) with (2.8) holding a.s. under a variety of assumptions on \( V_t \), e.g. Gaussianity; see Theorem IV.3.6 of Hannan (1970). An application with \( m = 0 \) is described in Subsection 5.1.

When the process \( Z_t \) in \( X_t^M = \begin{bmatrix} Z_t' & \cdots & Z_{t-m}' \end{bmatrix}' \) contains lagged values of \( W_t \), then the disturbance process \( y_t \) is zero, but \( y_t^M = \sum_{k \neq 0, \ldots, m} A_k Z_{t-k} \) is nonzero if \( X_t^M \) is misspecified, i.e, if \( A_k \neq 0 \) for some \( k \neq 0, \ldots, m \).

**2.2. Properties of the Misspecified Regression Model’s Disturbance**

With (2.7), the properties (2.1) and (2.8) determine the properties of the disturbance process (1.7) of the model (1.2) being estimated: \( y_t^M \) is A.S. with \( \gamma_k^M = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} y_{t+k} y_t^M \), \( k = 0, \pm 1, \ldots \) given by

\[
\gamma_k^M = A^N \Gamma_k^{NN} A^{N'} + \gamma_k^Y = \int_{-\pi}^{\pi} e^{-ik\lambda} dG_{y^M}(\lambda), \ a.s. \ [i.p.],
\]

where \( G_{y^M}(\lambda) = A^N G_{NN}(\lambda) A^{N'} + G_y(\lambda) \). Our assumption that \( X_t^M \) is misspecified means that \( A^N \Gamma_k^{NN} A^{N'} \neq 0 \) for some \( k \).

**2.3. An Example with Periodic \( X_t \)**

We introduce an example that will be analyzed further in Subsections 6.1 and 6.2. Its basic motivation is the idea that when \( X_t^M \) is a periodic regressor which inadequately models an unknown regressor with the same period, then \( X_t^N \) will include a compensating regressor with the same period. Consider the simple period-four regressor defined by \( X_t^M = a^M \cos \frac{\pi}{2} t + c^M (-1)^t \), with \( a^M, c^M \neq 0 \). Suppose that \( X_t^N = a^N \cos \frac{\pi}{2} t + b^N \sin \frac{\pi}{2} t \) with \( |a^N| + |b^N| \neq 0 \). Then \( X_t = \begin{bmatrix} X_t^M & X_t^N \end{bmatrix}' = a \cos \frac{\pi}{2} t + b \sin \frac{\pi}{2} t + c(-1)^t \), where \( a = \begin{bmatrix} a^M & a^N \end{bmatrix}, b = \begin{bmatrix} 0 & b^N \end{bmatrix} \) and \( c = \begin{bmatrix} c^M & 0 \end{bmatrix} \), resulting in \( \Gamma_k^X = \frac{1}{2} a'a \cos \frac{\pi}{2} k + \frac{1}{2} b'b \sin \frac{\pi}{2} k + c'c (-1)^k \), \( k = 0, \pm 1, \ldots \); see
Anderson (1971, p. 581). $G_X(\lambda)$ is piecewise constant with upward jumps at \( \lambda = \pm \pi/2, \pi \). For this example, even if \( y_t \) is weakly stationary, this is not true of \( y_t^M = A_N X_t^N + y_t \) for \( A_N \neq 0 \).

3. ARMA MODELING OF \( y_t^M \)

As was indicated in the Introduction, for our analyses, the AR and MA polynomials of a candidate ARMA model \( \phi(L) y_t^M = \alpha(L) a_t \) are transformed into the innovations filter \( \theta(L) = \phi(L)/\alpha(L) \) whose coefficient sequence \( \theta = (\theta_1, \theta_2, \ldots) \) is such that \( a_t = \sum_{j=0}^{\infty} \theta_j y_{t-j}^M \) will be white noise if the model is correct and \( y_t^M \) is covariance stationary. Under our weaker assumptions on \( y_t^M \), we would call the model correct when its autocovariance sequence is proportional to \( \gamma_k^M, k = 0, \pm 1, \ldots \), or equivalently, if \( dG_{y_t^M}(\lambda)/d\lambda \) is proportional to \( |\theta(e^{i\lambda})|^{-2} \) for \(-\pi \leq \lambda \leq \pi \). But our theorems do not require any model to be correct. For a white noise model, \( \theta = (1, \theta_1, \theta_2, \ldots) \) has the form \( \theta = (1, 0, 0, \ldots) \). 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 any \( \phi \), this sequence reduces to that of white noise when \( \alpha = \phi \). For an AR (1) model, \( \theta = (1, -\phi, 0, 0, \ldots) \).

Model “parameterization” by \( \theta \) is useful because the \( \theta \)’s determined by likelihood-maximizing ARMA coefficients have large sample limits in situations where the ARMA coefficients themselves do not, due to common zeroes in AR and MA polynomials. Convergence of a sequence of ARMA \((p, q)\) coefficients implies coordinatewise convergence of the associated \( \theta \)-sequence (as the ARMA (1, 1) formula illustrates). Consequently, the \( \theta \)’s associated with a continuously parameterized ARMA model family depend continuously on the family’s ARMA coefficient vector. However, convergence of a sequence \( \theta^T, T = 1, 2, \ldots \) from ARMA \((p, q)\) models implies convergence of the ARMA coefficients only when \( \theta = \lim_{T \to \infty} \theta^T \) is identifiable within the ARMA \((p, q)\) models, that is, when one of the minimal degree polynomials such that \( \theta(L) = \phi(L)/\alpha(L) \) is of full degree: either \( \deg \phi(L) = p \) or \( \deg \alpha(L) = q \). See the Appendix of Pötscher (1991) for more details and references. (Pötscher’s parameter is the coefficient sequence of \( \bar{\theta}(L) = \alpha(L)/\phi(L) \). The relationship between \( \theta \) and \( \bar{\theta} \) is continuous and invertible; see Section 3 of FPW, 2004.) It will be convenient to call every \( \theta \) a model.

To obtain the uniform convergence and continuity properties needed to establish the results indicated in the Introduction, ARMA \((p, q)\), model coefficient-vector estimation is restricted to compact sets of coefficient vectors whose AR and MA polynomials have all zeroes in \( \{|z| \geq 1 + \varepsilon\} \) for some \( \varepsilon > 0 \). These sets specify compact sets \( \Theta \) of the sort 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). To avoid confusion, we note that, for compatibility with the finite-past prediction error definition of GLS
estimates mentioned later in Section 7, quantities with \([\theta]\) like \(X_t^M[\theta]\) always have the definition

\[
X_t^M[\theta] = \sum_{j=0}^{t-1} \theta_j X_{t-j}^M.
\]  

That is, variates \(X_t^M[j]\) at time-index values \(t-j \leq 0\) are effectively treated as being zero even when known nonzero values are available at such times. In (1.3) and elsewhere, a generalized inverse is to be understood whenever the inverse matrix fails to exist. This can only happen for a fixed finite number of \(T\) values, due to (2.6) and (d) of Proposition D.1 in Appendix D.

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 any \(\theta\) from an invertible ARMA model, define

\[
C^{NM}(\theta) = \Gamma_0^{NM}(\theta) \Gamma_0^{MM}(\theta)^{-1}.
\]

In Appendix E, we prove

**Theorem 4.1.** Under the assumptions (2.1)–(2.2) and (2.5)–(2.8), let \(\Theta\) be a compact set of invertible ARMA models in the sense of Appendix A. Then

\[
\sup_{\theta \in \Theta} \left\| (A_T^M(\theta) - A^M) T^{-1/2} D_{M,T}^{-1} - A^N C^{NM}(\theta) \right\| \to 0 \text{ a.s. [i.p.]} \tag{4.3}
\]

holds, and \(A^N C^{NM}(\theta)\) is continuous on \(\Theta\) and thus bounded there, \(\sup_{\theta \in \Theta} \| A^N C^{NM}(\theta) \| < \infty\).

The proof shows that \(A^N C^{NM}(\theta)\) is the limiting value as \(T \to \infty\) of the \(\theta\)-model’s GLS coefficient estimate of the regression of \(A^N X_t^N, 1 \leq t \leq T\) on the array \(X_t^M T^{-1/2} D_{M,T}^{-1}, 1 \leq t \leq T\). We call \(A^N C^{NM}(\theta)\) the asymptotic bias characteristic of GLS estimation with the model \(\theta\) for the complementary regressor \(X_t^N\). When \(X_t^M\) is A.S., \(A^N C^{NM}(\theta)\) is the asymptotic bias of \(A_t^M(\theta)\). Generally, if \(A^N X_t^N\) and \(X_t^M\) are asymptotically correlated, i.e., not asymptotically orthogonal, then \(A^N C^{NM}(\theta) \neq 0\) for some \(\theta\). Omitted variable bias is a fundamental modeling issue; see e.g., Stock and Watson (2002, pp. 143-149). Its important role in forecast accuracy will be made clear in Section 6.

## 5. UNIFORM ASYMPTOTIC STATIONARITY OF FORECAST ERRORS

We consider sample second moments of the errors of one-step-ahead forecasts \(W_t^M(\theta, \theta^*, T)\) of (1.6). The forecast errors \(W_t - W_t^M(\theta, \theta^*, T), 1 \leq t \leq T\) are observable and equal to \(W_t[\theta] - A_T^M(\theta^*) X_t^M[\theta] = y_t[\theta] + \{AX_t - A_T^M(\theta^*) X_t^M\}[\theta]\). Here filtered quantities are truncated
as in (4.1). Thus, setting \( U_t(T) = \left[ y_t \ T_t^1 D_{M,T} X^M_t \ X^N_t \right]' \), \( 1 \leq t \leq T \), and \( \beta_T(\theta^*) = \left[ 1 \ (A^M - A^M_T(\theta^*)) T^{-1/2} D_{M,T}^{-1} \ A^N \right] \), these forecast errors are given by

\[
W_t - W_{t-1}^M (\theta, \theta^*, T) = \beta_T(\theta^*) U_t [\theta] (T), \ 1 \leq t \leq T.
\]

(5.1)

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

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

(5.2)

This fact and the properties of the \( U_t(T) \) array described in Appendix C lead to the following theorem, proved in Appendix E, establishing that, uniformly on relatively compact parameter sets, the limiting sample second moments of the forecast errors (5.1) are the same as those of the \( \theta \)-model forecast errors of the A.S. array

\[
y_t^M(\theta^*, T) = y_t + A^N \left( X_t^N - C^{NM}(\theta^*) T^1 D_{M,T} X^M_t \right), \ 1 \leq t \leq T.
\]

This asymptotic coincidence occurs because, with

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

(5.3)

this array has limiting sample second moments with spectral distribution function

\[
G_{M, \theta^*}(\lambda) = G_y(\lambda) + B^{NM}(\theta^*) G_X(\lambda) B^{NM}(\theta^*)_t.
\]

(5.4)

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 \( \theta^T \to \theta \) and \( \theta^{T*} \to \theta^* \) coordinatewise.

**Theorem 5.1.** Under the assumptions (2.1)–(2.2) and (2.5)–(2.8), let \( \bar{\Theta} \) and \( \bar{\Theta}^* \) be compact sets in the sense of Appendix A. Then the forecast-error arrays \( W_t - W_{t-1}^M (\theta, \theta^*, T), 1 \leq t \leq T \) 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

\[
\Gamma_k^M(\theta, \theta^*) = \int_{-\pi}^{\pi} e^{-ik\lambda} \left| \theta \left( e^{i\lambda} \right) \right|^2 dG_{M, \theta^*}(\lambda),
\]

(5.5)

the limits

\[
\frac{1}{T} \sum_{t=1}^{T-k} \left( W_{t+k} - W_{t+k|t+k-1}(\theta, \theta^*, T) \right) \left( W_t - W_{t-1}^M (\theta, \theta^*, T) \right) \to \Gamma_k^M(\theta, \theta^*)
\]

(5.6)

hold uniformly a.s. [i.p.] on \( \bar{\Theta} \times \bar{\Theta}^* \). Further, the functions \( \Gamma_k^M(\theta, \theta^*) \) are continuous and thus bounded on \( \bar{\Theta} \times \bar{\Theta}^* \).
Theorem 5.1 shows that the quantities $\Gamma_0^M (\theta, \theta^*)$ are of special interest because they describe limiting average squared one-step-ahead forecast errors. With

$$\gamma_0^M(\theta) = \int_{-\pi}^{\pi} \theta \left( e^{i\lambda} \right)^2 dG_{\theta} (\lambda), \quad (5.7)$$

(5.4) yields the decomposition

$$\Gamma_0^M(\theta, \theta^*) = \gamma_0^y(\theta) + B^{NM}(\theta^*) \Gamma_0^X(\theta) B^{NM}(\theta^*)' . \quad (5.8)$$

By specializing the argument used to establish Theorem 5.1, $\gamma_0^y(\theta)$ is seen to be the limiting average squared error of the one-step-ahead forecast of $W_t$ when $X_t^M = X_t$. Similarly, using (4.3), the second quantity on the right in (5.8) is seen to be the limit of the average of the squares of one-step-ahead forecast errors of the mean function error array $AX_t - A_T^M(\theta^*)X_t^M, 1 \leq t \leq T$,

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \left( \{ AX_t - A_T^M(\theta^*)X_t^M \} [\theta] \right)^2 \quad (5.9)$$

This formula reveals that $\Gamma_0^M(\theta, \theta^*)$ does not depend on the specific $X_t^N$ defining $X_t$. That is, if $\tilde{X}_t^N$ is A.S. and such that $\tilde{X}_t = \left[ X_t^{M'} \tilde{X}_t^N \right]'$ is S.A.S. and has the property that $A\tilde{X}_t = AX_t$ holds for all $t \geq 1$ for some $A$, then because the value of the l.h.s. of (5.9) is unchanged when $AX_t$ is replaced by $A\tilde{X}_t$, so also the value of the r.h.s. is unchanged if all quantities determined by $X_t$ are replaced by their analogues for $\tilde{X}_t$. For example, if the $i$-th coordinate $X_{t,i}^M$ of $X_t^M$ is constant with value one, then $\tilde{X}_t^N = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} X_t^N$ can be assumed to be zero, because by modifying $A^M$ by changing $A_i^M$ to $\tilde{A}_i^M = A_i^M + A^N \tilde{X}_t^N$ and defining $\tilde{X}_t^N = X_t^N - \tilde{X}_t^N$, one obtains $A\tilde{X}_t = AX_t$ with $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \tilde{X}_t^N = 0$.

5.1. Extension to ARIMA-Type Disturbance Models

Now suppose the observed data are $\tilde{W}_t, 1 - d \leq t \leq T$ from a time series of the form $\tilde{W}_t = A\tilde{X}_t + \tilde{y}_t$ to which a model of the form $\tilde{W}_t = A^M\tilde{X}_t^M + \tilde{y}_t^M$ is being fit. Suppose also that it has been correctly determined that the disturbances $\tilde{y}_t^M$ require "differencing" with a certain operator $\delta (L) = \sum_{j=0}^{d} \delta_j L^j$ (with $\delta_0 = 1$), whose zeroes are on the unit circle, in order to obtain residuals for which an ARMA model can be considered, resulting in the fitting of what is called a regARIMA model to $\tilde{W}_t$. Such models are extensively used for seasonal time series in the context of seasonal adjustment; see Findley et al. (1998) and 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 $W_t = \delta (L)\tilde{W}_t, y_t = \delta (L)\tilde{y}_t, X_t = \delta (L)\tilde{X}_t$, and $X_t^M = \delta (L)\tilde{X}_t^M$, the latter being a subvector of $X_t$. For any $1 \leq t \leq T$, since $\tilde{W}_t = W_t - \sum_{j=1}^{d} \delta_j \tilde{W}_{t-j}$, using $\theta$ from a candidate ARMA $(p, q)$ model for $y_t$ and $A_T^M(\theta^*)$ defined as above, a natural one-step-ahead forecast for $\tilde{W}_t$ is
\( \tilde{W}_{t|t-1}^M (\theta, \theta^*, T) = W_{t|t-1}^M (\theta, \theta^*, T) - \sum_{j=1}^{d} \delta_j \tilde{W}_{t-j} \), with \( W_{t|t-1}^M (\theta, \theta^*, T) \) defined by (1.6). This leads to \( \tilde{W}_t - \tilde{W}_{t|t-1}^M (\theta, \theta^*, T) = W_t - W_{t|t-1}^M (\theta, \theta^*, T) \) for \( 1 \leq t \leq T \), for the forecast errors and therefore to limiting results as in Theorem 5.1 with the same functions \( \Gamma_0^M (\theta, \theta^*) \).

Subsection 3.3 of Aelen (2004) provides an interesting one-step-ahead forecasting application involving a variety of seasonal time series \( \tilde{W}_t \) whose values come from enterprises that report economic data to Statistics Netherlands a month late, and \( \tilde{X}_{t}^M \) includes the series from enterprises of the same kind that report in the desired month, and so is stochastic. The models for \( W_t = \delta (L) \tilde{W}_t \) resemble that of Subsection 2.1 with \( m = 0 \), so our next theorem supports Aelen’s use of GLS estimation of the regARIMA models, done with X-12-ARIMA (Findley et al., 1998).

### 6. OPTIMALITY OF GLS

Because of the uniform convergence and continuity results established in Theorem 5.1, for any compact \( \Theta \) in the sense of Appendix A we have

\[
\min_{\theta \in \Theta} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left( W_t - W_{t|t-1}^M (\theta, \theta^*, T) \right)^2 \right\} \rightarrow \min_{\theta \in \Theta} \Gamma_0^M (\theta, \theta) , \tag{6.1}
\]

and, for any fixed \( \theta^* \in \Theta \),

\[
\min_{\theta \in \Theta} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left( W_t - W_{t|t-1}^M (\theta, \theta^*, T) \right)^2 \right\} \rightarrow \min_{\theta \in \Theta} \Gamma_0^M (\theta, \theta^*) . \tag{6.2}
\]

In Appendix E, we establish

**Theorem 6.1.** For any compact \( \Theta \) in the sense of Appendix A, let \( \tilde{\theta} \) denote a minimizer of \( \Gamma_0^M (\theta, \theta) \) over \( \Theta \) and, for a given \( \theta^* \in \Theta \), let \( \tilde{\theta}^* \) denote a minimizer of \( \Gamma_0^M (\theta, \theta^*) \) over \( \Theta \). Then

\[
\Gamma_0^M (\tilde{\theta}, \tilde{\theta}) \leq \Gamma_0^M (\tilde{\theta}^*, \theta^*) , \tag{6.3}
\]

with equality holding if and only if \( \tilde{\theta}^* \) is also a minimizer of \( \Gamma_0^M (\theta, \theta) \),

\[
\Gamma_0^M (\tilde{\theta}^*, \theta^*) = \Gamma_0^M (\tilde{\theta}, \tilde{\theta}) , \tag{6.4}
\]

and, simultaneously, the asymptotic bias characteristic of \( \theta^* \) coincides with that of \( \tilde{\theta}^* \),

\[
A^N C^{N, M} (\theta^*) = A^N C^{N, M} (\tilde{\theta}^*) . \tag{6.5}
\]

When strict inequality obtains in (6.3), then

\[
A^N C^{N, M} (\theta^*) \neq A^N C^{N, M} (\tilde{\theta}) \tag{6.6}
\]

11
holds for every minimizer $\bar{\theta}$ of $\Gamma_0^M(\theta, \theta)$. Therefore, when $\Gamma_0^M(\theta, \theta)$ has a unique minimizer $\bar{\theta}$, then (6.6) is both necessary and sufficient for strict inequality in (6.3).

We expect that both $\Gamma_0^M(\bar{\theta}, \bar{\theta}) < \Gamma_0^M(\theta^*, \theta^*)$ and $A^{NCNM}(\theta^*) \neq A^{NCNM}(\bar{\theta}^*)$ will hold except in quite special situations, the only one known to us being when $C^{NM}(\theta^*)$, and therefore also $\Gamma_0^M(\theta, \theta^*)$, does not depend on $\theta^*$. This is shown to occur for AR(1) models in Subsection 6.1 below only in a singular situation.

It follows from Theorem 5.1 by standard arguments (see Chapter 3 and Lemma 4.2 of Pötscher and Prucha, 1997) that the sequence $\theta^T = \arg\min_{\theta \in \Theta} \left\{ \sum_{t=1}^T \left( W_t - W^M_{t|i{-}1}(\theta, \theta, T) \right)^2 \right\}$ converges to the set of minimizers of $\Gamma_0^M(\theta, \theta)$ and therefore to the unique minimizer $\bar{\theta}$ when there is one. These $\theta^T$ are the GLS generalization of what are called conditional maximum likelihood estimates in the case of no regressors; see Box and Jenkins (1976, pp. 210–211). In case $\theta^T \to \bar{\theta}$, it follows from Theorem 4.1 that $(A^M_T(\theta^T) - A^M) T^{-1/2} D_{M,T}^1 \to A^{NCNM}(\bar{\theta})$ a.s. [i.p.], in which case (6.6) has the interpretation that $A^M_T(\theta^T)$ and $A^M_T(\theta^*)$ have different asymptotic bias characteristics. The same conclusions apply when unconditional (exact) maximum likelihood estimation is used; see Section 7. See Pötscher (1991) for results and references for the uniqueness or lack thereof for minimizers of the function (5.7) to which $\Gamma_0^M(\theta, \theta)$ reduces when the regressor is correct, but the ARMA model for the disturbances $y_t$ is not.

Model sets $\Theta$ usually include the white noise model $\theta^* = (1, 0, 0, \ldots)$ as a degenerate case making the conclusions of Theorem 6.1 generally applicable to the consideration of OLS as an alternative to GLS. The conclusions yield the following optimality property of GLS: In conjunction with maximum likelihood estimation of $\theta$ and in the asymptotic sense being considered, OLS estimation is never better than GLS estimation for one-step-ahead forecasting. When the mean function is misspecified and the limiting model $\bar{\theta}$ is not white noise, $\bar{\theta} \neq \theta^*$, then if $A^{NCNM}(\theta)$ is nonconstant over $\bar{\Theta}$, OLS is typically worse due to its asymptotic bias characteristic being different from that of the GLS estimator of $A^M$.

Thursby (1987) provides comparisons of OLS and GLS biases when $y_t$ is known to be i.i.d. (white noise), $\dim X^M_t = 2$, $\dim X^N_t = 1$, the coordinates of $X_t$ are correlated first-order autoregressive processes, and the loss function is the posterior mean squared bias associated with a prior for the parameters that determine the covariance structure between $X^N_t$ and $X^M_t$. With the aid of numerical integrations for the GLS quantities, he establishes that, depending on the choice of the autocovariance structure of $X^M_t$, the mean squared asymptotic bias of GLS is sometimes less and sometimes greater than that of OLS. Theorem 6.1 shows that, for either outcome, GLS has an asymptotic advantage over OLS for one-step-ahead forecasting.
6.1. Examples with AR(1) Models and \( \dim X_t^M = \dim X_t^N = 1 \)

We consider the case in which \( \dim X_t^M = \dim X_t^N = 1 \) and a first-order autoregressive model, with \( \theta = \theta (\phi) = (1, -\phi, 0, 0, \ldots) \) is used for disturbance series \( y_t^M \) in (1.2). This leads to

\[
\Gamma_0^M (\theta, \theta^*) = \int_{-\pi}^{\pi} \left[ 1 - \phi e^{i\lambda} \right]^2 dG_y (\lambda) + B_{NM}^M (\theta^*) \int_{-\pi}^{\pi} \left[ 1 - \phi e^{i\lambda} \right]^2 dG_X (\lambda) B_{NM}^M (\theta^*)',
\]

with \( \int_{-\pi}^{\pi} \left[ 1 - \phi e^{i\lambda} \right]^2 dG_y (\lambda) = (1 + \phi^2) \gamma_0^y - 2\phi \gamma_1^y \) and \( \int_{-\pi}^{\pi} \left[ 1 - \phi e^{i\lambda} \right]^2 dG_X (\lambda) = (1 + \phi^2) \Gamma_0^X - \phi (\Gamma_1^X + \Gamma_-^X) \). Also, with \( \theta^* = (1, -\phi^*, 0, \ldots) \), the \( C_{NM} (\theta^*) \) component of \( B_{NM}^M (\theta^*) \) is

\[
C_{NM}^M (\theta^*) = \frac{(1 + \phi^{*2}) \Gamma_0^N - \phi^* (\Gamma_1^N + \Gamma_-^N)}{(1 + \phi^{*2}) \Gamma_0^{MM} - 2\phi^* \Gamma_1^{MM}}.
\]

It is easily seen that \( C_{NM}^M (\theta^*) \) is equal to \( C_{NM}^N = \Gamma_0^N / \Gamma_0^{MM} \) for all \( \theta^* \), and equality holds in (6.3), unless

\[
2\Gamma_0^N \Gamma_1^{MM} - (\Gamma_1^N + \Gamma_-^N) \Gamma_0^{MM} \neq 0.
\]

Under (6.8), the derivative of \( C_{NM}^N (\theta^*) \) is nonzero and \( C_{NM}^M (\theta^*) \) is strictly monotonic on \(-1 < \phi^* < 1\); see Subsection 6.3 of Findley (2005), whose derivation shows that the unique \( \theta^* = (1, -\phi^*, 0, \ldots) \) minimizing (6.7) is

\[
\bar{\phi}^* = \gamma_1^y + (A^N)^2 \left\{ \Gamma_1^{NN} + (C_{NM} (\theta^*))^2 \Gamma_1^{MM} - C_{NM} (\theta^*) (\Gamma_1^N + \Gamma_-^N) \right\} / \gamma_0^y + (A^N)^2 \left\{ \Gamma_0^{NN} - (C_{NM} (\theta^*))^2 \Gamma_0^{MM} \right\}.
\]

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 (6.5) fails, and therefore strict inequality holds in (6.3) by Theorem 6.1. For the OLS choice, \( \phi^* = 0 \), when \( C_{NM}^M (\theta^*) = C_{NM}^N \), (6.9) shows that \( \bar{\phi}^* \neq 0 \) (except possibly at a single value of \( (A^N)^2 \)), when either \( \gamma_1^y \) or \( \Delta_{NM} = \Gamma_1^{NN} + (C_{NM})^2 \Gamma_1^{MM} - C_{NM} (\Gamma_1^N + \Gamma_-^N) \) is nonzero, which will usually be the case.

For example, for the period-four \( X_t \) of Subsection 2.3, the l.h.s. of (6.8) has the value \(-a^M a^N (c^M)^2 \), so (6.8) holds if and only if \( a^N \neq 0 \). Further, \( C_{NM} = a^M a^N \left\{ (a^M)^2 + 2 (c^M)^2 \right\}^{-1} \) and \( \Delta_{NM} = - (c^M C_{NM})^2 \). Hence strict inequality holds in (6.3) for OLS estimation except when \( a^N = 0 \), or, if \( \gamma_1^y > 0 \), when \( (A^N)^2 = \gamma_1^y (c^M C_{NM})^{-2} \).

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

For the correct regression model case (1.1) in which \( X_t \) consists of polynomials (including constants) and products of polynomials and sinusoids, or slightly more complex functions, Grenander (1954) showed that, when \( y_t \) is weakly stationary with mean zero and with a positive,
continuous spectral density, then the OLS estimate $A_T^M = \sum_{t=1}^T W_t X_t' \left( \sum_{t=1}^T X_t X_t' \right)^{-1}$ of $A$ is asymptotically efficient in the sense that $\lim_{T \to \infty} D_T^{-1} E \left\{ (A_T - A)' (A_T - A) \right\} D_T^{-1}$ is minimal (in the ordering of symmetric matrices) among all linear, unbiased estimates of $A$. However, OLS is not efficient for periodic regressors whose representation in terms of sines and cosines has more than $\dim X_t$ components, such as the trading day regressors with period 336 months described in Findley et al. (1998) or the Easter regressors of Bell and Hillmer (1983) with much longer periods. Nor is it efficient for the example $X_t$ of Subsection 2.3 when $b^N \neq 0$. See example (1) on p. 253 and Section 7.7 of Grenander and Rosenblatt (1984).

In the misspecified regression case (1.2), to apply Grenander’s efficiency result, $y_t^M$ must be weakly stationary with mean zero and with a positive, continuous spectral density. Further, $G^{NN} (\lambda)$ must be continuously differentiable, thereby excluding periodic regressors from $X_t^N$. Applied to (nonstochastic) $X_t^M$, Grenander’s regressor assumptions require $G^{MM} (\lambda)$ to be piecewise constant (and the sum of the ranks of the jumps $G^{MM} (\lambda^+) - G^{MM} (\lambda^+)$, $0 \leq \lambda \leq \pi$ cannot exceed $\dim X_t^M$; see Theorem 10.2.10 of Anderson, 1971). Section 6.1 of Findley (2005) gives a simple argument showing that these properties of $G^{MM} (\lambda)$ and $G^{NN} (\lambda)$ force $X_t^N$ and $X_t^M$ to be asymptotically orthogonal, $\Gamma_k^{NM} = 0$, $k = 0, \pm 1, \ldots$, resulting in $A^N C^{NM} (\theta^*) = 0$ for all $\theta^*$. In this case, equality holds in (6.3) because $\Gamma_0^M (\theta, \theta^*)$ does not depend on $\theta^*$. (There will still be regressor misspecification effects because $\Gamma_0^M (\theta, \theta^*)$ and $\gamma_0^m (\theta)$ will differ in a nonconstant way. Further, they will generally have different minimizers.)

When the asymptotic bias characteristic $A^N C^{NM} (\theta^*)$ is nonzero, as in the examples of Subsections 2.1 and 6.1, the analogue for $A_T^M (\theta^*)$ of Grenander’s efficiency measure fails because some entries of $(A_T^M (\theta^*) - A^M) D_{M,T}^{-1}$ have order $T^{1/2}$; see (4.3). In other words, the bias of $A_T^M (\theta^*)$ has a larger order of magnitude than its statistical variability as measured by $\lim_{T \to \infty} D_{M,T}^{-1} E \left\{ (A_T^M (\theta^*) - EA_T^M (\theta^*))' (A_T^M (\theta^*) - EA_T^M (\theta^*)) \right\} D_{M,T}^{-1}$ under Grenander’s assumptions on $y_t$.

7. EXTENSIONS AND RELATED RESULTS

From their definitions, it is not surprising that GLS estimates of regARMA and regARIMA models have an optimality property for one-step-ahead forecasting, but a systematic investigation of the topic has been lacking. A pleasingly simple result, such as Theorem 6.1’s connection of optimality with asymptotic bias characteristics, does not seem likely even for the correct regressor case. Indeed, if asymptotic efficiency results are indicative, the correct regressor case itself 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 Subsection 6.2 show that OLS is asymptotically efficient only for a limited range of mostly simple regressors in the deterministic case.

For any fixed $\theta^*$, 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 could be shown that the limit as \( T \to \infty \) of the variance of \( T^{-1/2} \sum_{t=1}^{T} \left( W_t - W_{t|t-1}^{M}(\theta^*, \theta^*, T) \right) \) does not depend on \( \theta^* \).

So far, we have only provided asymptotic results for the most simply defined GLS estimates, obtained by truncating the infinite-past forecast error filters. Section 2.4 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, as in Amemiya (1973), in conjunction with unconditional maximum likelihood estimation. See Section 9 of Findley (2003) for additional details, and also for details about weakening the assumptions on \( X_t^M \) to include the frequently used intervention variables of Box and Tiao (1975), which decay exponentially to zero, and so have weight one in \( D_{M,T} \), causing (2.5) to fail. 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 \( \theta \)-parameter estimates that minimize average squared multistep-ahead forecast errors (using for \( y_t^M \) the more comprehensive model classes of FPW, 2004). With a.s convergence in (2.1), the asymptotic second moments in (2.1) and (2.2) can be random, from nonergodic \( y_t \) and \( X_t \). In this case, the limits \( A^{NC^{NM}}(\theta^*) \) and \( \Gamma_0^M(\theta, \theta^*) \) in Theorems (4.1) and (5.1) are random.

Findley (2005) uses the results of Theorem 4.1 and 6.1 to obtain formulas 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.

Acknowledgement. We are grateful to John Aston and Tucker McElroy and two referees for comments that stimulated substantial improvements.

Disclaimer. Any views expressed are the author’s and not necessarily those of the U. S. Census Bureau.

References


APPENDICES

A. COMPACT $\theta$–SETS FOR ESTIMATION

For each $\varepsilon > 0$ and integer pair $p, q \geq 0$, 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 $\phi(z)/\alpha(z) = \hat{\theta}(z)$ all belong to $\{ |z| \geq 1 + \varepsilon \}$. This set has the property that 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., $\theta^{S(T)}_j \to \theta_j$, $j \geq 1$. Thus $\Theta_{p,q,\varepsilon}$ is compact for coordinatewise convergence. It also has the property that whenever $0 \leq \varepsilon_0 < \varepsilon$, the sums $\sum_{j=0}^{\infty}(1 + \varepsilon_0)^j |\theta_j| \leq \infty$ and $\lim_{j \to \infty} \sup_{\theta \in \Theta_{p,q,\varepsilon}} \sum_{j=0}^{\infty}(1 + \varepsilon_0)^j |\theta_j| = 0$. See Lemmas 2 and 10 of Findley (2005) for these and other properties mentioned. Our uniform convergence results below follow from these facts as do some other important properties. First, the functions $\theta \left( e^{i\lambda} \right) = \sum_{j=0}^{\infty} \theta_j e^{i\lambda}$ are continuous on $-\pi \leq \lambda \leq \pi$ and uniformly bounded and bounded away from zero:

$$m_{p,q,\varepsilon} = \min_{-\pi \leq \lambda \leq \pi, \theta \in \Theta_{p,q,\varepsilon}} \left| \theta \left( e^{i\lambda} \right) \right| > 0, \quad M_{p,q,\varepsilon} = \max_{-\pi \leq \lambda \leq \pi, \theta \in \Theta_{p,q,\varepsilon}} \left| \theta \left( e^{i\lambda} \right) \right| < \infty. \quad (A.1)$$

Second, if a sequence $\theta^T$, $T = 1, 2, \ldots$ in $\Theta_{p,q,\varepsilon}$ converges coordinatewise to some $\theta$, then it also converges in the stronger sense that $\lim_{T \to \infty} \sum_{j=0}^{\infty}(1 + \varepsilon_0)^j |\theta^T_j - \theta_j| = 0$ whenever $0 \leq \varepsilon_0 < \varepsilon$. (Thus $\Theta_{p,q,\varepsilon}$ is compact for these stronger forms of convergence, i.e., topologies.)
Our theorems apply to any \( \tilde{\Theta} \) for which \( \tilde{\Theta} \subseteq \Theta_{p,q,\varepsilon} \) holds for some \( \varepsilon > 0 \) and \( p, q \geq 0 \) that are closed, i.e., the limit of every coordinatewise convergent sequence in \( \tilde{\Theta} \) belongs to \( \tilde{\Theta} \). The \( \Theta \) of interest are specified by compact sets of ARMA coefficient vectors.

**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 scalably asymptotically stationary, a property we now consider in some detail. Let \( U_t, t \geq 1 \) be a real-valued column vector sequence that is SAS, and let \( I_U \) denote the identity matrix of order \( \dim U \), the dimension of \( U_t \). Thus there is a decreasing sequence \( D_1 \geq D_2 \geq \ldots \) of positive definite diagonal matrices that satisfy

\[
\lim_{T \to \infty} D_{T+k}^{-1} D_T = I_U \quad (k \geq 1)
\]

and are such that, for each \( k \geq 0 \), the limits

\[
\Gamma_k^U = \lim_{T \to \infty} D_T \sum_{t=1}^{T-k} U_{t+k} U_t^T D_T \text{ a.s. \([i.p.]\)}
\]

exist (finitely). The properties (B.1)–(B.2) yield \( \lim_{T \to \infty} D_T U_{T-j} = 0 \text{ a.s. } [i.p.], \ j \geq 0 \). For example, when \( j = 0 \), as \( T \to \infty \),

\[
D_T U_T U_T^T D_T = D_T \sum_{t=1}^{T} U_t U_t^T D_T - (D_T D_T^{-1} D_T - 1) \sum_{t=1}^{T-1} U_t U_t^T D_T^{-1} (D_T^{-1} D_T)
\]

converges to \( \Gamma_0^U - \Gamma_0^U = 0 \text{ a.s. } [i.p.]. \) Also \( D_T \ \downarrow \ 0 \), which leads to \( \lim_{T \to \infty} D_T U_{1+j} = 0 \text{ a.s. } [i.p.] \) for all \( j \geq 0 \).

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 requires the diagonal elements of \( \Gamma_k^U \) to be positive. Our requirement (2.5) is stronger for \( X_1^M \).) He shows that, with \( \Gamma_k^U = (\Gamma_k^U)' \), the real matrix sequence \( \Gamma_k^U, k = 0, \pm 1, \ldots \) has a representation

\[
\Gamma_k^U = \int_{-\pi}^{\pi} e^{-it\lambda} dG_U(\lambda)
\]

in which \( G_U(\lambda) \) is an Hermitian-matrix-valued function such that the eigenvalues of increments \( G_U(\lambda_2) - G_U(\lambda_1), \lambda_2 \geq \lambda_1 \), are non-negative, 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 \( G_U(\lambda) \) 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 S.A.S. property for regressors whose entries \( X_{i,t} \) are polynomials, linear combinations (perhaps infinite) of sinusoids, \( \cos \lambda t \) or \( \sin \lambda t \), with \( -\pi \leq \lambda \leq \pi \), (scaling sequence \( T^{-1/2} \)) and products of polynomials \( t^p \) and sinusoids (scaling sequence \( T^{-p-1/2} \)). For
exponentially increasing regressors, e.g. \( U_t = e^{bt} \) with \( b > 0 \), (B.1) is incompatible with the requirement \( \Gamma_0^U > 0 \); see Hannan (1970, p. 77), so such regressors are not S.A.S.

C. VECTOR ARRAY REFORMULATION OF ASSUMPTIONS

The following reformulation of our assumptions (2.1)–(2.2) and (2.5)–(2.8) concerning \( y_t \) and \( X_t \) will enable us to make use of the results of FPW (2001, 2004). The vector array

\[
U_t(T) = \begin{bmatrix} y_t \\ T^{1/2} D_{X,T} X_t^N \end{bmatrix}, \quad 1 \leq t \leq T, \quad (C.1)
\]

is A.S. More specifically, for each \( k \geq 0 \),

\[
\Gamma_k^U = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T-k} U_{t+k}(T)U_t(T)' = \begin{bmatrix} \gamma_k^y & 0 & 0 \\ 0 & \Gamma_k^{MM} & \Gamma_k^{MN} \\ 0 & \Gamma_k^{NM} & \Gamma_k^{NN} \end{bmatrix}, \text{ a.s. [i.p.],} \quad (C.2)
\]

with \( \Gamma_0^{MM} > 0 \). Further, from Appendix B,

\[
\lim_{T \to \infty} T^{-1/2} U_{1+j}(T) = 0 = \lim_{T \to \infty} T^{-1/2} U_{T-j}(T) \text{ a.s. [i.p.], } j \geq 0. \quad (C.3)
\]

Due to (C.2), the asymptotic spectral distribution matrix of \( U_t(T) \) has the block diagonal form \( G_U(\lambda) = \text{diag}(G_y(\lambda), G_X(\lambda)) \).

D. UNIFORM CONVERGENCE RESULTS FOR FILTERED A.S. ARRAYS

The results below are formulated to encompass the more general results indicated in Section 7.

**Proposition D.1.** Let \( U_t(T), \quad 1 \leq t \leq T \) be an A.S. \( n \)-dimensional column vector array satisfying (C.3) whose asymptotic spectral distribution matrix is denoted by \( G_U(\lambda) \). Let \( H \) and \( Z \) be sets of filters \( \eta = (\eta_0, \eta_1, \ldots) \) and \( \zeta = (\zeta_0, \zeta_1, \ldots) \) whose absolute coefficient sums converge uniformly on \( H \) resp. \( Z \). Then the filter output arrays \( U_t[\eta](T), \ U_t[\zeta](T), \ 1 \leq t \leq T, \eta \in H, \zeta \in Z \) defined in analogy with (4.1) have the following properties:

(a) \( \lim_{T \to \infty} \sup_{\eta \in H} \|T^{-1/2} U_{1+j,T}[\eta]\| = \lim_{T \to \infty} \sup_{\eta \in H} \|T^{-1/2} U_{T-j}[\eta]\| = 0, \text{ a.s. [i.p.]} \)

\[ j \geq 0, \text{ and analogously for } U_t[\zeta](T). \]

(b) For any \( k \geq 0 \), as \( T \to \infty, \sup_{\eta \in H, \zeta \in Z} \left\| T^{-1} \sum_{t=1}^{T-k} U_{t+k}[\eta](T)U_t[\zeta](T)' - \Gamma_k^U(\eta, \zeta) \right\| \to 0 \text{ a.s. [i.p.], with } \Gamma_k^U(\eta, \zeta) = \int_{-\pi}^{\pi} e^{-ik\lambda} \eta(e^{i\lambda}) \zeta(e^{-i\lambda}) dG_U(\lambda). \]
(c) The functions $\Gamma_k^U (\eta, \zeta)$ are bounded on $H \times Z$, \[ \| \Gamma_k^U (\eta, \zeta) \| \leq \| \Gamma_0^U \| \sup_{\eta \in H} |\eta (e^{i\lambda})| \sup_{\zeta \in Z} |\zeta (e^{i\lambda})| < \infty, \]
and are jointly continuous in $\eta, \zeta$ 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 $\inf_{\eta \in H, -\pi \leq \lambda \leq \pi} |\eta (e^{i\lambda})|^2 \Gamma_0^U (\eta, \eta) \leq \sup_{\eta \in H, -\pi \leq \lambda \leq \pi} |\eta (e^{i\lambda})|^2 \Gamma_0^U (\eta, \eta) \leq \sup_{\eta \in H} |\eta (e^{i\lambda})|^2 \Gamma_0^U (\eta, \eta)$. 

(d) Let $H$ be an index set for a family of arrays $U_t (\eta, T), \ 1 \leq t \leq T, \eta \in H$ such that, as $T \to \infty$, \[ \sup_{\eta \in H} \left\| \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) - \Gamma_0 (\eta) \right\| \to 0 \ a.s., \tag{D.1} \]
where the $\Gamma_0 (\eta)$ are positive definite matrices whose minimum eigenvalues are bounded away from zero, that is \[ \inf_{\eta \in H} \lambda_{\min} (\Gamma_0 (\eta)) \geq m_H \tag{D.2} \]
holds for some $m_H > 0$. Then \[ \sup_{\eta \in H} \left\| \left( \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) \right)^{-1} - \Gamma_0 (\eta)^{-1} \right\| \to 0 \ a.s. \tag{D.3} \]

**Proof.** Parts (a)–(c) are straightforward vector extensions of special cases of Theorem 2.1 and Proposition 2.1 of FPW (2001). For (d), 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 $T_\varepsilon$ such that for $T \geq T_\varepsilon$ the inequalities $\sup_{\eta \in H} \left\| \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) - \Gamma_0 (\eta) \right\| < \frac{\varepsilon m_H^2}{2}$ and $\inf_{\eta \in H} \lambda_{\min} \left( \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) \right) \geq \frac{1}{2} m_H$ hold. Hence for these $T$ and all $\eta \in H$, 

\[
\sup_{\eta \in H} \left\| \left( \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) \right)^{-1} - \Gamma_0 (\eta)^{-1} \right\|
\leq \sup_{\eta \in H} \left\{ \left\| \left( \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) \right)^{-1} \right\| \left\| \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) - \Gamma_0 (\eta) \right\| \right\}
\leq \frac{1}{m_H} \sup_{\eta \in H} \left\{ \lambda_{\min}^{-1} \left( \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) \right) \right\} \sup_{\eta \in H} \left\{ \left\| \frac{1}{T} \sum_{t=1}^T U_t (\eta, T) - \Gamma_0 (\eta) \right\| \right\}
\leq \varepsilon,
\]
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 FPW (2004).

**Lemma D.2.** Suppose that, on a set \( \hat{\Theta}^* \), the sequence \( \beta_T (\theta^*) \), \( T = 1, 2, \ldots \) of row vector functions converges uniformly to a bounded function \( \beta (\theta^*) \), i.e., (5.2) 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

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

and \( \sup_{\eta \in H, \zeta \in Z} \| \Gamma_0 (\eta, \zeta) \| < \infty \) hold. Then, as \( T \to \infty \),

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

\[
\to 0 \text{ a.s. [i.p.], } k \geq 0.
\]

**E. PROOFS**

**Proof of Theorem 4.1.** We have

\[
(A_T^M (\theta) - A^M) T^{-1/2} D_{M,T}^{-1}
\]

\[
= T^{-1/2} \sum_{t=1}^{T} y_t \chi_t^M [\theta] X_t^M [\theta]' D_{M,T} \left( D_{M,T} \sum_{t=1}^{T} X_t^M [\theta] X_t^M [\theta]' D_{M,T} \right)^{-1}
\]

\[
= T^{-1/2} \sum_{t=1}^{T} y_t \chi_t^M [\theta] X_t^M [\theta]' D_{M,T} \left( D_{M,T} \sum_{t=1}^{T} X_t^M [\theta] X_t^M [\theta]' D_{M,T} \right)^{-1}
\]

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

By (b) and (c) of Proposition D.1, \( T^{-1/2} \sum_{t=1}^{T} y_t \chi_t^M [\theta] X_t^M [\theta]' D_{M,T} \) converges uniformly a.s. [i.p.] to 0 and \( T^{-1/2} \sum_{t=1}^{T} \chi_t^N [\theta] X_t^M [\theta]' D_{M,T} \) 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 \( m_\Theta^2 \Gamma_0^{MM} \), where \( m_\Theta = \min_{\pi \leq \lambda \leq \pi, \theta \in \Theta} | \theta (e^{i\lambda}) | > 0 \); see Appendix A. It follows from (d) 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 \( m_\Theta^{-2} (\Gamma_0^{MM})^{-1} \)). Hence \( (A_T^M (\theta) - A^M) T^{-1/2} D_{M,T}^{-1} \) converges uniformly a.s. [i.p.] to \( A^N C^{NM} (\theta) \), which is continuous on \( \Theta \) as well as bounded.

**Proof of Theorem 5.1.** The assertions follows from Lemma D.2 with \( \tau_T (\theta^*) = \beta_T (\theta^*) \),
$H = Z = \bar{\Theta}$ and $U_t(\theta, T) = W_t(\theta, T) = U_t[\theta](T) = \sum_{j=0}^{T-1} \theta_j U_{t-j}(T)$ defined by (C.1), because the uniform convergence of $T^{-1} \sum_{t=1}^{T-k} U_{t+k}[\theta](T) U_t[\theta](T)'$ to $\Gamma^U_0(\theta) = \int_{-\pi}^{\pi} e^{-ik\lambda} \theta(e^{i\lambda})^2 dG_U(\lambda)$ and the boundedness of $||\Gamma^U_0(\theta)||$ on $\bar{\Theta}$ that are required to apply Lemma D.2 follow from (b) and (c) 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 $G_U(\lambda) = diag(G_y(\lambda), G_\bar{x}(\lambda))$ yields the two component form of $G_{M, \theta^*}(\lambda)$ in (5.4).

**Proof of Theorem 6.1.** We start by establishing that, for any invertible $\theta$ and $\theta^*$, we have $\Gamma^M_0(\theta, \theta) \leq \Gamma^M_0(\theta, \theta^*)$ with equality holding if and only if $A^N C^{NM}(\theta^*) = A^N C^{NM}(\theta)$. Indeed, the component of $\Gamma^M_0(\theta, \theta^*)$ that depends on $\theta^*$ can be reexpressed in terms of the analogues of $C^{NM}(\theta^*)$ and $\Gamma^X_0(\theta)$ obtained by replacing $X^N_t$ with $\bar{X}^N_t = A^N X^N_t$, which we denote by $\bar{C}^{NM}(\theta^*)$ and $\bar{\Gamma}^X_0(\theta)$:

\[
A^N \begin{bmatrix} -C^{NM}(\theta^*) & I_N \end{bmatrix} \bar{\Gamma}^X_0(\theta) \begin{bmatrix} -C^{NM}(\theta^*) & I_N \end{bmatrix}^T = \begin{bmatrix} -\bar{C}^{NM}(\theta^*) & 1 \end{bmatrix} \bar{\Gamma}^X_0(\theta) \begin{bmatrix} -\bar{C}^{NM}(\theta^*) & 1 \end{bmatrix}^T.
\]

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

\[
\begin{bmatrix} -\bar{C}^{NM}(\theta^*) & 1 \end{bmatrix} \bar{\Gamma}^X_0(\theta) \begin{bmatrix} -\bar{C}^{NM}(\theta^*) & 1 \end{bmatrix}^T \leq \begin{bmatrix} -C & 1 \end{bmatrix} \bar{\Gamma}^X_0(\theta) \begin{bmatrix} -C & 1 \end{bmatrix}^T,
\]

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

Next, note that because $\Gamma^X_0(\theta, \theta)$ and $\bar{\Gamma}^X_0(\theta, \theta^*)$ are continuous functions of $\theta$ on $\bar{\Theta}$, they have minimizers $\bar{\theta}$, resp. $\bar{\theta}^*$ over $\bar{\Theta}$. From Lemma ??, we have

\[
\Gamma^M_0(\bar{\theta}, \bar{\theta}) \leq \Gamma^M_0(\bar{\theta}^*, \bar{\theta}^*) \leq \Gamma^M_0(\bar{\theta}^*, \theta^*).
\]

Thus $\Gamma^M_0(\bar{\theta}, \bar{\theta}) = \Gamma^M_0(\bar{\theta}^*, \theta^*)$ holds if and only if (6.4) and $\Gamma^M_0(\bar{\theta}^*, \bar{\theta}^*) = \Gamma^M_0(\bar{\theta}^*, \theta^*)$ do, and the latter is equivalent to (6.5), as was just shown.

Finally, if (6.6) fails, i.e., if $A^N C^{NM}(\theta^*) = A^N C^{NM}(\bar{\theta})$, then $\Gamma^M_0(\bar{\theta}^*, \theta^*) \leq \Gamma^M_0(\bar{\theta}, \theta^*) = \Gamma^M_0(\bar{\theta}, \bar{\theta})$, which, from (E.2), yields $\Gamma^M_0(\bar{\theta}^*, \theta^*) = \Gamma^M_0(\bar{\theta}, \bar{\theta}) = \Gamma^M_0(\bar{\theta}^*, \bar{\theta}^*)$, contradicting strict inequality in (6.3).