Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions

LARS PETER HANSEN, JOHN C. HEATON, and MASAO OGAKI*

In this article we study a class of econometric models that imply a set of multiperiod conditional moment restrictions. These restrictions depend on an unknown parameter vector. We construct an extensive class of consistent, asymptotically normal estimators of this parameter vector and calculate the greatest lower bound for the asymptotic covariance matrices of estimators in this class. In so doing, we extend results reported by Hansen (1985) and Stoica, Söderström, and Friedlander (1985), by allowing for more general forms of nonlinearities and temporal dependence. Many dynamic econometric models imply that the expectation of a function of a currently observed data vector and an unknown parameter vector conditioned on information available at some point in the past is 0. We focus on models in which the conditioning information is lagged more than one time period, as in the models considered by Barro (1981), Dunn and Singleton (1986), Eichenbaum and Hansen (1987), Eichenbaum, Hansen, and Singleton (1988), Hansen and Hodrick (1983), Hansen and Singleton (1988), and Hall (1988). Hence we consider econometric models that imply multiperiod conditional moment restrictions that depend on an unknown parameter vector. Within the context of these models, it is possible to estimate the parameter vector without simultaneously estimating the law of motion for the entire set of observable variables. The basic idea is to use the conditional moment restrictions to deduce a set of unconditional moment restrictions. Then estimators of the parameter vector can be obtained by using sample counterparts to the unconditional moment restrictions as described by Sargan (1958) and Hansen (1982). Such estimators are referred to as *generalized method of moments* (GMM) estimators. For most applications the conditional moment restrictions imply an extensive set of unconditional moment restrictions. As a consequence, there is a vast array of GMM estimators that can be used to estimate consistently the parameter vector of interest. Each member of this set of estimators is constructed using a distinct collection of the unconditional moment restrictions. Hence it is of interest to compare the performances of the alternative GMM estimators. For tractability we investigate only the asymptotic distributions of the estimators in question. More precisely, we use a method suggested by Hansen (1985) for calculating a greatest lower bound for the asymptotic covariance matrices of the alternative GMM estimators, that is, an efficiency bound. We compute the efficiency bound for a rich collection of time series models that imply multiperiod conditional moment restrictions. Hansen (1985) illustrated this method for a time series model with conditionally homoscedastic moving-average disturbance terms for which the moving-average polynomial is invertible. Stoica et al. (1985) calculated efficiency bounds for GMM estimators for autoregressive parameters in *autoregressive moving-average* models without unit roots. They established that the efficiency bound for GMM estimators of the autoregressive parameters coincides with the asymptotic covariance matrix of the Gaussian maximum likelihood estimators. The models considered by Hansen (1985) in his illustrative example and by Stoica et al. (1885) can be viewed as special cases of the models considered in this article. Although we do not make any direct comparisons to maximum likelihood, we do allow for moving-average disturbances that are conditionally heteroscedastic and moving-average lag polynomials that cannot be inverted.

KEY WORDS: Generalized method of moments; Heteroscedasticity; Martingale approximation; Stationary time series.

1. INTRODUCTION

The plan of the article is as follows: In Section 2 we describe the admissible data-generation processes and the martingale approximation that is used to obtain central limit results. In Section 3 we specify the econometric model under consideration and construct a class of *generalized method of moments* (GMM) estimators. In Section 4 we characterize a lower bound for the asymptotic covariance matrices of the GMM estimators and give conditions under which this bound is the greatest lower bound. In Section 5 we provide two examples of the calculation of this efficiency bound. The first displays the calculation in the presence of an explicit form of conditional heteroscedasticity; the second displays the calculation when the moving-average polynomial of an *autoregressive moving-average* (ARMA) model can have a unit root. Finally, we make some concluding remarks in Section 6.

2. DATA GENERATION AND MARTINGALE APPROXIMATION

Let $(\Omega, A, \Pr)$ denote an underlying probability space. Associated with this space is a transformation S mapping Ω onto Ω that determines the law of motion for states of the world over time.

Assumption 1. The transformation S is one-to-one, measurable, measure-preserving, and ergodic, and S^{-1} is measurable.

Suppose that x is a random vector and $x(\omega)$ is a measurement vector when the state of the world is ω. Then x in conjunction with S generates a stochastic process via $x_t(\omega) \equiv x[S^t(\omega)]$, where S^t is interpreted as the transformation S applied t times. Since S is one-to-one, x_t is also well-defined for negative values of t. The stochastic process so generated is strictly stationary and ergodic; conversely,

* Lars Peter Hansen is Professor and John C. Heaton is Student, Department of Economics, University of Chicago, Chicago, IL 60637, and Economics Research Center, National Opinion Research Center, 1155 East 60th Street, Chicago, IL 60637. Masao Ogaki is Assistant Professor, Department of Economics, University of Rochester, Rochester, NY 14627. Hansen's work was partially supported by National Science Foundation Grant SES-8512371, Heaton's work was supported by the Greentree Foundation and the Social Science and Humanities Research Council of Canada, and Ogaki's work was supported by a Sloan Foundation Dissertation Fellowship and a Ford Motor Company Fellowship. The authors thank Ricardo Barros, Bo Honore, Whitney Newey, and George Tauchen for helpful comments.

Journal of the American Statistical Association
September 1988, Vol. 83, No. 403, Theory and Methods

any stationary stochastic process can be constructed in this manner (e.g., see Doob 1953, pp. 452–457). Throughout this article we find it convenient to index stochastic processes constructed in this fashion by the random vector x used in conjunction with S to generate the stochastic process. Equivalently, x can be viewed as the time-zero component of the stochastic process, since S^0 is the identity transformation.

Let y be a particular random vector that is observable at time 0 and B_t be the sigma algebra generated by $y_t, y_{t-1}, \ldots$. Then $\{B_t : -\infty < t < \infty\}$ is a nondecreasing sequence of sigma algebras. The sigma algebra B_t summarizes the information available at date t.

Following Gordin (1969) we approximate stochastic processes by martingale difference sequences. The stochastic processes to be approximated are multiperiod forecast errors relative to $\{B_t : -\infty < t < +\infty\}$. More precisely, we consider random vectors in the space X, where

$$X \equiv \{x : x \text{ is a } k\text{-dimensional random vector that is measurable with respect to } B_0,\ \|x\| < \infty, \text{ and } E(x \mid B_{-s}) = 0\}, \tag{2.1}$$

$\|x\| \equiv [E(x'x)]^{1/2}$, and s is a fixed positive integer. For each x in X, there is a corresponding random vector $M(x) \equiv \sum_{\tau=0}^{s-1} [E(x_\tau \mid B_0) - E(x_\tau \mid B_{-1})]$ that generates a martingale difference sequence. The random vector $M(x)$ is measurable with respect to B_0 and satisfies $\|M(x)\| < \infty$ and $E[M(x) \mid B_{-1}] = 0$. Define $P_T(x) \equiv (1/T)^{1/2} \sum_{t=1}^{T} x_t$. It is straightforward to show that $\lim_{T\to\infty} \|P_T[x - M(x)]\| = 0$ (e.g., see Gordin 1969; Hansen 1985). We apply Billingsley's (1961) central limit theorem for martingales to show that $\{P_T[M(x)] : T \geq 1\}$ converges in distribution to a normally distributed random vector with mean 0 and covariance matrix $E[M(x)M(x)']$. It then follows that $\{P_T(x) : T \geq 1\}$ has the same limiting distribution. Furthermore, for any x and x^* in X,

$$\begin{aligned} E[M(x)M(x^*)'] &= \lim_{T\to\infty} E[P(x)P(x^*)'] \\ &= \sum_{\tau=-s+1}^{s-1} E(xx_\tau^{*\prime}). \end{aligned} \tag{2.2}$$

3. ECONOMETRIC MODEL AND GMM ESTIMATORS

Let ϕ denote a function specified a priori, mapping the random vector y and the unknown k-dimensional parameter vector β_0 into an n-dimensional disturbance term.

Assumption 2. $\phi(\cdot, \beta)$ is Borel measurable for all β in a neighborhood of β_0 and $\phi(r, \cdot)$ is continuously differentiable in the same neighborhood of β_0 for all vectors of real numbers r in the support set of y.

Assumption 2 ensures that the disturbance $e \equiv \phi(y, \beta_0)$ is a vector of random variables, the entries of the $n \times k$ matrix $d \equiv \partial\phi(y, \beta_0)/\partial\beta$ are random variables, and

$$\operatorname{mod}(\delta) \equiv \sup_{|\beta-\beta_0|<\delta} |\partial\phi(y, \beta)/\partial\beta - \partial\phi(y, \beta_0)/\partial\beta| \tag{3.1}$$

is a random variable for sufficiently small values of δ. The matrix norm $|c|$ in (3.1) is defined as $\operatorname{tr}(cc')^{1/2}$.

The remaining assumptions in this section are stated in terms of positive numbers ε and η in the interval $[2, \infty)$. The number ε is used to restrict moments of the disturbance vector; η is used in the construction of a class of GMM estimators. Different values of (ε, η) are associated with different sets of assumptions.

The disturbance vector is assumed to have finite εth moments and to be a multiperiod forecast error.

Assumption 3. $E[|e|^\varepsilon] < \infty$ and $E(e \mid B_{-s}) = 0$, where $s > 0$ is fixed.

The second requirement given in Assumption 3 is a conditional moment restriction that is used to identify and estimate the parameter vector β_0. This type of conditional moment restriction is often implied by a variety of economic models (e.g., see Barro 1981; Dunn and Singleton 1986; Eichenbaum and Hansen 1987; Eichenbaum, Hansen, and Singleton 1988; Hall 1988; Hansen and Hodrick 1983; Hansen and Singleton 1988).

We construct estimators of β_0 as follows: Let z be an $n \times k$ matrix of random variables that are measurable with respect to the time $-s$ information set, B_{-s}. In addition, suppose that $E|z|^\eta$ is finite; then $E(z'e) = 0$. We refer to a consistent estimator $\{b_T : T \geq 1\}$ of β_0 for which $\{P_T[z'\phi(y, b_T)] : T \geq 1\}$ is $o_p(1)$ (converges in probability to 0) as a GMM estimator with index z. Such an estimator can also be viewed as an M estimator and as an instrumental-variables estimator, where z' is a matrix of instrumental variables.

Since there is great flexibility in the selection of z, a rich class of estimators of β_0 is at our disposal. We find it convenient to introduce an index set for a family of such estimators. Let h be a random vector that is measurable with respect to B_{-s}, B^τ be the subsigma algebra of B_{-s} that is generated by $h, h_{-1}, \ldots, h_{-\tau}$, and

$$Z^\tau \equiv \{z : z \text{ is } n \times k \text{ matrix of random variables that are measurable with respect to } B^\tau \text{ and } E(|z|^\eta) < \infty\}. \tag{3.2}$$

Indexes in the space Z^τ are constructed using functions of the current and τ lags of the vector h.

For each τ the index set is a closed linear subspace of the Banach space consisting of all $n \times k$ matrices z of random variables for which $E(|z|^\eta)$ is finite. The norm on the Banach space is given by $E(|z|^\eta)^{1/\eta}$, and we let $L^\eta(\Pr)$ denote this Banach space for general specifications of η, k, and n.

The specification of the finite lag τ is both arbitrary and inconvenient. For this reason, we consider the larger index set $Z \equiv \bigcup_{\tau=1}^{\infty} Z^\tau$. This index set is also a linear subspace of $L^\eta(\Pr)$, but it is not necessarily closed. Associated with this index set is the sigma algebra B^*, where $B^* \equiv \bigvee_{\tau=1}^{\infty} B^\tau$. Note that $B^* \subseteq B_{-s}$. Analogously we define $B_j^* \equiv \bigvee_{\tau=1}^{\infty} B_j^\tau$, where B_j^τ is the sigma algebra generated by $h_j, h_{j-1}, \ldots, h_{j-\tau}$.

It is convenient to introduce an even larger index set:

$$Z^+ \equiv \{z : z \text{ is an } n \times k \text{ matrix of random variables that are measurable with respect to } B^* \text{ and } E(|z|^\eta) < \infty\}. \tag{3.3}$$

The space Z^+ is just the $L^\eta(\Pr)$ closure of Z. In this article we do not formally consider estimators with indexes in Z^+ that are not simultaneously in Z. Such estimators are more problematic to construct in practice because they must depend, at least asymptotically, on the infinite past of h. By not considering such estimators, the efficiency bound we calculate often is not attained, but can only be approximated by members of the index set Z.

In deriving the asymptotic distribution of the estimators indexed by the set Z, we require $z'e$ to have a finite second moment. One way to ensure this is to restict ε and η as follows:

Assumption 4. $1/\varepsilon + 1/\eta = 1/2$.

When Assumption 4 is imposed, we can use the Hölder inequality to verify that $E(z'ee'z)$ is well-defined and finite.

An alternative way to guarantee that $z'e$ has a finite second moment is to restrict e to be conditionally homoscedastic.

Assumption 4′. $E(ee_j \mid B^*)$ is constant almost surely for all j.

In this case $E(|z'ee'z|) \le E(|z|^2|e|^2) = E[E(|z|^2|e|^2 \mid B^*)] = E(|z|^2)E(|e|^2) < \infty$. Assumption 4′ restricts all of the conditional autocovariances of e to be constant.

Our next assumption restricts the matrix $d^* \equiv E(d \mid B^*)$ of random variables.

Assumption 5. $E(|d^*|^\sigma) < \infty$ and $E[1_{\{|d^*|<\delta\}}d^{*\prime}d^*]$ is nonsingular for $1/\sigma + 1/\eta = 1$ and some $\delta > 0$.

The first restriction imposed in Assumption 5 guarantees via the Hölder inequality that $|z'd^*|$ has a finite moment for any index z in Z^+. By the law of iterated expectations, it follows that $E(z'd^*) = E(z'd)$ for any z in Z^+ because the entries of z are measurable with respect to B^*. The second restriction imposed by Assumption 5 ensures the existence of at least one consistent estimator in Z. When $\eta = 2$, Assumption 5 implies that $|d^*|$ has a finite second moment. In this case, the additional restriction on d^* is equivalent to a requirement that $E(d^{*\prime}d^*)$ be nonsingular.

The derivation of the asymptotic distribution of the GMM estimators in Z uses the familiar approach relying on the mean-value theorem. To apply this approach we impose a local domination condition.

Assumption 6. $E[\text{mod}(\delta)^\sigma] < \infty$ for $1/\sigma + 1/\eta = 1$ and for some $\delta > 0$.

This assumption implies that

$$E[\sup_{|\beta-\beta_0|<\delta} |z\partial\phi(y, \beta)/\partial\beta - z\partial\phi(y, \beta_0)/\partial\beta|] < \infty \quad (3.4)$$

for some $\delta > 0$, which in turn implies the first-moment continuity restriction used by Hansen (1982) to derive the asymptotic distributions for GMM estimators.

For each z in Z^+ either Assumption 4 or 4′ implies that $z'e$ is in X. It is convenient to define a matrix inner product:

$$\langle z \mid z^* \rangle \equiv \sum_{\tau=-s+1}^{s-1} E(z'ee_\tau' z_\tau^*). \quad (3.5)$$

Using this notation, $\{P_T(z'e) : T \ge 1\}$ converges in distribution to a normally distributed random vector with mean 0 and covariance matrix $\langle z \mid z \rangle$. Lemma 3.1 follows from analysis in Hansen (1982).

Lemma 3.1. Suppose that Assumptions 1–3, 6, and either 4 or 4′ are satisfied, $\{b_T : T \ge 1\}$ converges in probability to β_0, and $\{P_T[z'\phi(y, b_T)] : T \ge 1\}$ is $o_p(1)$. Then $\{E(z'd^*)[T^{1/2}(b_T - \beta_0)] - P_T(ze) : T \ge 1\}$ is $o_p(1)$.

When $E(z'd^*)$ is nonsingular, Lemma 3.1 implies that $\{T^{1/2}(b_T - \beta_0) : T \ge 1\}$ is asymptotically equivalent to $\{[E(z'd^*)]^{-1}P_T(ze) : T \ge 1\}$. In this case $\{T^{1/2}(b_T - \beta_0) : T \ge 1\}$ has a limiting normal distribution with asymptotic covariance matrix

$$\text{cov}(z) = [E(z'd^*)]^{-1}\langle z \mid z \rangle [E(z'd^*)']^{-1}. \quad (3.6)$$

4. EFFICIENCY BOUND

4.1 General Analysis

In Section 3 we described a class of GMM estimators that can be infinite dimensional. The estimators in this class are indexed by members of the set Z. Relation (3.6) gives a mapping cov from the index set Z into the collection, PSD, of $k \times k$ positive semidefinite matrices augmented by the point infinity. The set PSD can be partially ordered as follows: The inequality $c \le c^*$ is satisfied for c and c^* in PSD if $c^* - c$ is in PSD. When c^* is infinite, the inequality $c \le c^*$ is satisfied for any c in PSD. Let LB be the subset of PSD containing all matrices c that satisfy

$$c \le \text{cov}(z) \quad \text{for all } z \text{ in } Z. \quad (4.1)$$

The efficiency bound inf(Z) is defined to be the maximal element of LB assuming such a maximal element exists. Hansen (1985) derived a set of sufficient conditions for inf(Z) to exist and suggested a general method for calculating this matrix. In this section we apply that method to the estimation problem described in Section 3.

Our first step in calculating inf(Z) is to find an alternative representation for $\langle \cdot \mid \cdot \rangle$ by using a conditional counterpart to the forward filter suggested by Hayashi and Sims (1983). Let

$$U \equiv \{u : u = (v_1^1) \cdot e_1 + (v_2^2) \cdot e_2 + \cdots + (v_\tau^\tau) \cdot e_\tau \text{ for some positive integer } \tau \text{ and some random vectors } v^j \text{ that are measurable with respect to } B^* \text{ and bounded}\}. \quad (4.2)$$

The members of U have finite εth moments and satisfy the conditional moment restriction

$$E(u \mid B^*) = 0, \quad (4.3)$$

since the elements of v^j are measurable with respect to B^*. Let U^* denote the mean-square (L^2) closure of U, and let

$$e^* \equiv e - \text{proj}(e \mid U^*), \quad (4.4)$$

where proj($\cdot \mid U^*$) denotes the least squares projection operator onto U^*.

The forward filtering in (4.4) has two desirable features. First, it preserves the orthogonality of the resulting disturbance vector with the members of Z, since (4.3) and Assumption 3 guarantee that $E(e^* \mid B^*) = 0$. Second, forward filtering results in a disturbance vector e^* that generates a conditionally serially uncorrelated stochastic process (as will be shown in Lemma 4.1). In general, the process generated by e^* will not be a martingale difference sequence. Thus forward filtering is used in characterizing $\inf(Z)$, but it cannot be used to establish central limit results via martingale difference approximations. The projection error e^* is restricted as follows:

Assumption 7. $E(e^*e^{*\prime} \mid B^*) \equiv \Phi^*$ is nonsingular almost surely.

Let Φ^+ be an $n \times n$ matrix of random variables that are measurable with respect to B^* and satisfy $(\Phi^*)^{-1} = \Phi^{+\prime}\Phi^+$. We define $e^+ \equiv \Phi^+e^*$. Then e^+ is a conditional linear function of e^*, where the conditional weighting matrix Φ^+ is chosen so that $E(e^+e^{+\prime} \mid B^*) = I$. Since Φ^+ is measurable with respect to B^*, $E(e^+ \mid B^*) = 0$. A conditional moving-average representation for e in terms of e^+ is given by the following lemma.

Lemma 4.1. Suppose that Assumptions 1–3 and 7 are satisfied. Then

$$e = \sum_{j=0}^{s-1} (\lambda_j^j)(e_j^+), \tag{4.5}$$

where λ^j is an $n \times n$ matrix of random variables that are measurable with respect to B^* and have finite second moments for $0 \leq j \leq s - 1$, λ^0 is nonsingular almost surely, and $E(e^+e_j^{+\prime} \mid B^*) = 0$ for all $j \neq 0$ and $\lambda_k^j(\omega) \equiv \lambda^j[S^k(\omega)]$. If in addition Assumption 4′ is satisfied, $\lambda^0, \lambda^1, \ldots, \lambda^{s-1}$ can be chosen to be constant (nonrandom).

All proofs of lemmas and theorems not reported in the text are given in the Appendix.

We now construct two operators using $\{\lambda^j : 0 \leq j \leq s - 1\}$. Let D^+ be the set of all $n \times k$ matrices of random variables that are measurable with respect to B^*. Define

$$\Lambda(z) \equiv \sum_{j=0}^{s-1} E[(\lambda_j^j)z_j \mid B^*] \quad \text{and} \quad \Lambda^-(z) \equiv \sum_{j=0}^{s-1} (\lambda^j)'z_{-j} \tag{4.6}$$

for any z in D^+ with finite second moments. Notice that $E[(\lambda_j^j)z_j \mid B^*]$ is well defined because both λ_j^j and z_j have finite second moments and that $\Lambda(z)$ and $\gamma^-(z)$ are in D^+.

We use the operators Λ and Λ^- to represent $\langle \cdot \mid \cdot \rangle$.

Lemma 4.2. Suppose that Assumptions 1–4 and 7 are satisfied. Then for any z and z^* in Z^+, $\langle z \mid z^* \rangle = E[\Lambda^-(z)'\Lambda^-(z^*)] = E\{z'\Lambda[\Lambda^-(z^*)]\}$. The same implications are obtained when Assumption 4′ is used in place of Assumption 4.

Our strategy for calculating the bound is essentially the same as that used for solving minimum norm problems in Hilbert spaces. The basic idea is to find an $n \times k$ matrix d^+ of random variables with finite second moments that satisfies the *first-order* conditions

$$E(z'd^*) = E[\Lambda^-(z)'d^+] \quad \text{for all } z \text{ in } Z. \tag{4.7}$$

Lemma 4.3. Suppose that Assumptions 1–3 and 5 are satisfied. If $E[|d^+|^2] < \infty$ and d^+ satisfies (4.7), then $E(d^{+\prime}d^+)$ is nonsingular.

A lower bound for $\inf(Z)$ turns out to be $[E(d^{+\prime}d^+)]^{-1}$. This fact can be established as follows: Consider any z in Z such that $E(z'd^*)$ is nonsingular. Let $a = [E(z'd^*)]^{-1} \Lambda^-(z)' - [E(d^{+\prime}d^+)]^{-1}d^{+\prime}$. Then it follows from (3.6), Lemma 4.2, and $E(aa') \geq 0$ that $[E(d^{+\prime}d^+)]^{-1} \leq \operatorname{cov}(z)$.

To obtain a candidate for d^+, we construct two operators whose inverses are Λ and Λ^-, respectively. Let $\psi^0 \equiv (\lambda^0)^{-1}$ and $\psi^j = -(\lambda_{-j}^0)^{-1} \sum_{\tau=1}^{j} (\lambda_{\tau-j}^\tau)(\psi^{j-\tau})$ for $j \geq 1$, where $\lambda^j = 0$ for $j \geq s$. Notice that ψ^j is measurable with respect to B^* for all $j \geq 0$. Let D be the subset of D^+ for which $E[(\psi_j^j)z_j \mid B^*]$ is well defined for each j and

$$\Psi(z) = \sum_{j=0}^{\infty} E[\psi_j^j)z_j \mid B^*] \tag{4.8}$$

converges in $L^2(\Pr)$. Similarly, let D^- be the subset of D^+ for which

$$\Psi^-(z) \equiv \sum_{j=0}^{\infty} (\psi^j)'z_{-j} \tag{4.9}$$

converges in $L^\eta(\Pr)$. Then D is the domain of the operator Ψ, and D^- is the domain of the operator Ψ^-. By construction, Λ and Λ^- are inverses of Ψ and Ψ^-, respectively.

Lemma 4.4. Suppose that Assumptions 1–4 and 7 are satisfied. If z is in D, then $\Lambda[\Psi(z)] = z$. If z is in D^-, then $\Lambda^-[\Psi^-(z)] = z$. The same implications are obtained when Assumption 4′ is used in place of Assumption 4.

With the following extra assumption, $\Psi(d^*)$ is a candidate for d^+.

Assumption 8. d^* is in D.

Theorem 4.1. Suppose that Assumptions 1–8 are satisfied. Then $\inf(Z)$ exists and $\{E[\Psi(d^*)'\Psi(d^*)]\}^{-1} \leq \inf(Z)$. The same implications are obtained when Assumption 4′ is used in place of Assumption 4.

This theorem gives a nontrival lower bound for the asymptotic covariance matrices. For this bound to be sharp, there must exist a sequence $\{z^j : j \geq 1\}$ in Z such that $\{\Lambda^-(z^j) : j \geq 1\}$ converges in $L^2(\Pr)$ to $\Psi(d^*)$. The following additional restriction guarantees the convergence.

Assumption 9. $\Psi(d^*)$ is in the $L^2(\Pr)$ closure of $\Lambda^-(Z^+)$.

Theorem 4.2. Suppose that Assumptions 1–9 are satisfied. Then $\{E[\Psi(d^*)'\Psi(d^*)]\}^{-1} = \inf(Z)$. The same implication is obtained when Assumption 4′ is used in place of Assumption 4.

Proof. By Assumption 9, there exists a sequence $\{z^j : j \geq 1\}$ in Z^+ such that $\{\Lambda^-(z^j) : j \geq 1\}$ converges to $\Psi(d^*)$ in $L^2(\Pr)$. The sequence $\{z^j : j \geq 1\}$, however, is not necessarily in Z. In the Appendix we show how to approximate $[L^n(\Pr)]$ members of this sequence by elements in Z. This approximation entails taking expectations conditioned on B^τ for sufficiently large values of τ. In this manner we can construct GMM estimators with asymptotic covariance matrices that are arbitrarily close to the efficiency bound.

A convenient sufficient condition for Assumption 9 is that $\psi(d^*)$ be in D^-. In this case, $\Psi^-[\Psi(d^*)]$ is in Z^+, implying that $\Psi(d^*)$ is in $\Lambda^-(Z^+)$. Hence the following corollary to Theorem 4.2 is immediate.

Corollary 4.1. Suppose that Assumptions 1–8 are satisfied and $\Psi(d^*)$ is in D^-. Then $\{E[\Psi(d^*)'\Psi(d^*)]\}^{-1} = \inf(Z)$. The same implication is obtained when Assumption 4′ is used in place of Assumption 4.

When $s = 1$, there is additional flexibility in the construction of the index set. For instance, Z^τ could be used in place of Z for any nonnegative integer τ and B^* modified accordingly. In this case, $\Psi(d^*) = \psi^0 E(d \mid B^*)$ so that the bound is just $[E(d^{*\prime}\psi^{0\prime}\psi^0 d^*)]^{-1} = \{E[d^{*\prime}(\Phi^*)^{-1}d^*]\}^{-1}$, where Φ^* is now the covariance matrix of e conditioned on B^*. This bound is the time series counterpart to the bound reported by Amemiya (1977) and Jorgenson and Laffont (1974), modified to accommodate conditional heteroscedasticity.

4.2 Conditional Homoscedasticity With Unit Roots

In this section we explore further implications when Assumption 4′ is used instead of Assumption 4. Assumption 4′ was imposed in an illustration of Hansen (1985, sec. 5), Hansen and Sargent (1982), Hayashi and Sims (1983), and Stoica et al. (1985). [All references to Hansen (1985) in this section refer only to the analysis in sec. 5 of that article and not to the more general analysis in earlier sections of that article.]

Under Assumption 4′, the coefficients λ^j can be chosen to be constant. In fact, the representation $e = \sum_{j=0}^{s-1}(\lambda_j^j)e^+$ implied by Lemma 4.1 is a forward version of a Wold decomposition.

It is of interest to study the spectral density of the process generated by e. Consider the function $\Xi(\zeta) \equiv \sum_{j=0}^{s-1} \lambda^j(\zeta)^j$ of a complex variable ζ. When Assumption 7 is satisfied, it is necessarily true that the rank of $\Xi(\zeta)$ is n for all $|\zeta| < 1$ (e.g., see Rozanov 1967, p. 63). Hansen (1985), Hayashi and Sims (1983), and Stoica et al. (1985) maintained the assumption that the spectral density function of the process generated by e has rank n at all frequencies. This implies that $\Xi(\zeta)$ is nonsingular on $\{\zeta : |\zeta| = 1\}$ as well. We relax that assumption by allowing $\Xi(\zeta)$ to be singular at isolated points in $\{\zeta : |\zeta| = 1\}$ or, equivalently, the spectral density function of e to be singular at a finite number of frequencies. Hence we allow for (possibly complex) unit roots in the moving-average representation for the disturbance vector.

In Hansen (1985) Assumption 8 was implied by the more primitive assumptions imposed. This is no longer true when unit roots are permitted. The presence of unit roots means that the sequence of coefficients $\{\psi^j : j \geq 0\}$ will not converge to 0 and in fact may display polynomial growth. It is still possible, however, for the sequence $\{E(d_j^* \mid B^*) : j \geq 1\}$ to converge in $L^2(\Pr)$ to 0 sufficiently rapid so that d^* is in D. Hence by only imposing Assumption 8, we are able to extend the results of Hansen (1985).

When there are unit roots present in the moving-average representation, $\Psi(d^*)$ cannot be in D^- because the sequence of coefficients $\{\psi^j : j \geq 0\}$ does not converge to 0 and $E[\Psi(d^*)'\Psi(d^*)]$ is nonsingular. Consequently, Corollary 4.1 is not applicable; however, Assumption 9 can still be satisfied. To see this suppose that the following restriction is satisfied.

Assumption 9′. The spectral density for each column of $\Psi(d^*)$ is essentially bounded.

Note that the Wold decomposition theorem for covariance stationary processes implies that there exists a sequence of $n \times n$ matrix polynomials $\{\Gamma^j : j \geq 1\}$

$$\Gamma^j(\zeta) \equiv \sum_{\tau=0}^{N(j)} \gamma^{j,\tau}(\zeta)^\tau \tag{4.10}$$

such that

$$\lim_{j\to\infty}(1/2\pi)\int_{-\pi}^{\pi} |\Gamma^j[\exp(i\theta)]\Xi[\exp(i\theta)] - I|^2\, d\theta = 0, \tag{4.11}$$

because e^+ is the mean-square limit of a sequence of finite linear combinations of current and future values of e. Let

$$\Delta^j(z) \equiv \sum_{\tau=0}^{N(j)} \gamma^{j,\tau\prime} z_{-\tau}. \tag{4.12}$$

Since Assumption 9′ is satisfied, it follows that $\{\Lambda^-(\Delta^j[\Psi(d^*)]) : j \geq 1\}$ converges in $L^2(\Pr)$ to $\Psi(d)$. Hence we have the following second corollary to Theorem 4.2.

Corollary 4.2. Suppose that Assumptions 1–3, 4′, 5–8, and 9′ are satisfied. Then $\{E[\Psi(d^*)'\Psi(d^*)]\}^{-1} = \inf(Z)$.

To accommodate conditional heteroscedasticity in Section 2 we used a sigma algebra to build Z^τ and hence Z. When the forecast error e is conditionally homoscedastic, we can use an alternative linear space construction for the index set. For example, we can use the space, J^*, which is the mean-square closure of the space of random variables that are finite linear combinations of current and past values of h. We then replace expectations conditional on B^* with least squares projections onto J^*. All of our analysis applies to this alternative index set as well.

5. APPLICATIONS

5.1 Application 1

The first example is derived from a continuous-time financial economics model examined by Grossman, Melino, and Shiller (1987), Hall (1988), and Hansen and Singleton (1988) and from the martingale taxation model examined by Barro (1981). In this model there is a two-period conditional moment restriction that results from time averaging the underlying continuous-time processes and then sampling these averages.

Consider a q-dimensional random vector x that generates a Gaussian process. Let w denote a q-dimensional random vector that generates a process that is fundamental for x in the sense of linear prediction theory (e.g., see Rozanov 1967). For notational convenience, we define $y' \equiv [x', x'_{-1}]$. The two-period conditional moment restrictions apply to a linear combination $u \equiv [1\ \beta_0\ 0]x$, where β_0 is a scalar parameter to be estimated. In particular, $E(u \mid B_{-2}) = 0$ and $E(\rho u^2 - uu_{-1} \mid B_{-2}) = 0$, where ρ is known a priori and determined by the assumed form of time averaging. These two conditional moment restrictions can then be used to estimate the parameter β_0.

To map this model into the notation of Section 2, let

$$\phi(y, \beta) = \begin{bmatrix} [1\ \beta\ 0]x \\ \rho([1\ \beta\ 0]x)^2 - ([1\ \beta\ 0]x)([1\ \beta\ 0]x_{-1}) \end{bmatrix} \tag{5.1}$$

so that $n = 2$ in this example. The conditional moment restrictions take the form $E(e \mid B_{-2}) = 0$, where $e \equiv \phi(y, \beta_0)$. Notice that the model is nonlinear in both the parameters and the variables even though the underlying process generated by x is Gaussian. Finally, let $B^* = B_{-2}$ and η be any positive number greater than 2. The magnitude of η affects the size of the index set Z; however, the efficiency bound turns out to be insensitive to the choice of η.

Although u is conditionally homoscedastic, it turns out that the second entry of e induces conditional heteroscedasticity of a known form. It is straightforward to characterize the form of this heteroscedasticity, λ^0 and λ^1, and ψ^j for $j = 0, 1, \ldots$ (see Heaton and Ogaki 1988).

Calculation of $\Psi(d^*)$ requires evaluating expectations of $\psi_j^j d_j$ conditioned on B^* for $j = 0, 1, \ldots$. Heaton and Ogaki (1988) show how to perform these computations in the special case in which x has a state-space representation: $Y = AY_{-1} + Cw$ and $x = H'Y$, where Y is a $p \times 1$ state vector, A is a $p \times p$ matrix with eigenvalues that have moduli that are less 1, and C and H are $p \times q$ matrices of real numbers. In this case $\Psi(d^*)$ is given by

$$\Psi(d^*) = \begin{bmatrix} Q \cdot Y_{-2} \\ R_1 + w'_{-2}R_2Y_{-2} + Y'_{-2}R_3Y_{-2} \end{bmatrix}, \tag{5.2}$$

where Q is a p-dimensional vector of real numbers, R_1 is a scalar real number, and R_2 and R_3 are $p \times p$ matrices of real numbers. The efficiency bound is then given by $1/E[\Psi(d^*)'\Psi(d^*)]$.

In this example it is also possible to estimate β_0 using just the conditional moment restriction $E(u \mid B^*) = 0$ and not $E(\rho u^2 - uu_{-1} \mid B^*) = 0$. There is a corresponding loss in asymptotic efficiency, since the resulting efficiency bound is the reciprocal of the second moment of the first entry of $\Psi(d^*)$.

5.2 Application 2

The second example is an ARMA model:

$$y = \beta_0 y_{-1} + v_0 w + v_1 w_{-1}, \tag{5.3}$$

where y and w are scalar random variables, $|\beta_0| < 1$, and v_0 and v_1 are real numbers. We assume that w is fundamental for y, which in turn implies that $|v_0| \geq |v_1|$ (e.g., see Rozanov 1967, p. 63). Let $B^* = B_{-2}$. We assume that $E(w \mid B_{-1}) = 0$ and $E(w^2 \mid B_{-1}) = 1$.

We focus exclusively on estimating β_0 via GMM (or instrumental variables) as suggested by Stoica et al. (1985). In terms of the notation given in Section 3, $e = v_0 w + v_1 w_{-1}$, $\phi(y, \beta) = y - \beta y_{-1}$, $d = -y_{-1}$, $n = 1$, and $s = 2$. Let η be any positive number greater than 2. Again, the efficiency bound will not depend on the choice of η.

A forward factorization for e, as given in Lemma 4.2, is $e = v_0 e^+ + v_1 e_1^+$. Thus in this example, $\Lambda(z) = v_0 z + v_1 E[z_1 \mid B^*]$ and $\Psi(z) = \sum_{j=0}^{\infty} (v_0)^{-1}(-v_1/v_0)^j E[z_j \mid B^*]$. It is straightforward to show that $\Psi(d^*) = -[1/(v_0 + v_1\beta_0)](\beta_0 y_{-2} + v_1 w_{-2})$. This calculation is applicable even when $|v_0| = |v_1|$ and there is a unit root in the moving-average polynomial. In other words, Assumption 8 is always satisfied as long as $|\beta_0| < 1$. It is also straightforward to show that the spectral density function of $\Psi(d^*)$ is continuous and hence bounded, so Assumption 9′ is also satisfied. Consequently, Corollary 4.2 is applicable. The efficiency bound in this case is just

$$\inf(Z) = (v_0 + \beta_0 v_1)^2(1 - \beta_0^2)/(\beta_0 v_0 + v_1)^2. \tag{5.4}$$

This efficiency bound agrees with that calculated by Hansen (1985) and Stoica et al. (1985) when the ratio $|v_1/v_0|$ is less than 1. Bound (5.4), however, also applies to cases in which $|v_1/v_0|$ is equal to 1. Note that for a fixed $|v_0| > 0$ and $|\beta_0| < 1$, the efficiency bound is continuous as v_1/v_0 tends to 1 or -1.

If y has a nonzero mean and a constant term appears on the right side of (5.3), then the column of $\Psi(d^*)$ corresponding to the constant term will not converge in $L^2(\Pr)$ when $|v_0| = |v_1|$. Even in this case, the efficiency bound for the GMM estimators of β_0 can be calculated by the same method. It is straightforward to show that estimation of a constant term has no impact on the estimation of β_0 as long as an additional moment restriction that $E(e) = 0$ is used in estimation.

6. CONCLUDING REMARKS

An advantage of the GMM estimators is that they allow for the estimation of the parameter vector of interest without simultaneously estimating the auxiliary parameters that determine the law of motion for the vector of state variables. Calculation of the bound reported here requires knowledge of the auxiliary parameters as well; however, given estimates of these parameters, it is often straightforward to compute an estimate of the efficiency bound. Efficiency losses can be assessed by computing the estimated efficiency bound to the estimated covariance matrix from a GMM estimator that uses an ad hoc choice of an index. [For examples of these types of calculations, see Hansen and Singleton (1988).] If the loss in asymptotic efficiency is quite small and an ad hoc index estimator has reasonable small-sample properties, then it may not be desirable to construct more complicated estimators that are asymptotically efficient [see Tauchen (1986) for a Monte Carlo analysis of GMM estimators].

As in Chamberlain (1987), when $L^\eta(\Pr)$ is separable it is possible to approximate the efficiency bound as follows: First, find a basis sequence $\{z^j : j \geq 1\}$ for Z^+ that is in Z. All members of Z^+ are then $L^\eta(\Pr)$ limits of finite combinations of this basis. Then for each j, use the methods of Hansen (1982) and Sargan (1958) to attain a bound for the finite-dimensional set of indexes of the form $\sum_{j=1}^{\tau} c_j z^j$, where $c_1, c_2, \ldots, c_\tau$ are $k \times k$ matrices of real numbers. The sequence of finite-dimensional efficiency bounds indexed by τ will converge to $\inf(Z)$. This conceptual exercise illustrates how to approximate the efficiency bound, but it allows for too much flexibility to be of use in empirical work.

Although we delineated the sense of approximation required for sequences of GMM estimators to get arbitrarily close to the efficiency bound, we did not show how to construct estimators that actually attain the efficiency bound. Procedures along the lines of Newey (1987) and Robinson (1987a,b) that accommodate conditional heteroscedasticity could possibly be extended to the model considered here. Stoica et al. (1985) suggested a method for attaining the bound for linear, conditional homoscedastic examples like the one considered in Section 5.2. Their approach relies on filtering some of the time series data using the inverse of the moving-average filter. When there are unit roots in the moving-average polynomial, their approach breaks down. It would be of interest to devise methods for attaining the bound that apply more generally for models like those considered in Sections 4.2 and 5.2.

When the index set Z is constructed using a sigma algegra $B^* = B_{-s}$, we conjecture that the efficiency bound calculated in this article will apply to a much richer class of estimators. Demonstrating this entails extending Chamberlain's (1987) analysis of efficiency bounds implied by conditional moment restrictions to time series contexts. Such an extension could exploit methods developed by Begun, Hall, Huang, and Wellner (1983), Levit (1975), and Stein (1956). Following Stein (1956) we could examine the set F of alternative finite-dimensional parameterizations of the true law of motion for the dynamic system subject to the conditional moment restrictions. For each parameterization in F, there corresponds an asymptotic covariance matrix for a maximum likelihood estimator of β_0. Let $\sup(F)$ be the *least upper bound* of these asymptotic covariance matrices. In light of results in Chamberlain (1987) and Levit (1975), we conjecture that $\inf(Z) = \sup(F)$ in these circumstances. Hence our methods may provide an alternative way to calculate the semiparametric efficiency bound $\sup(F)$.

APPENDIX: PROOFS OF LEMMAS AND THEOREMS

Proof of Lemma 4.1. Let U^c denote the orthogonal complement of U^* relative to U^*_{-1}. Our first task is to obtain a convenient representation of U^c. Let u be any element of U_{-1}. Then,

$$u = (v_0^1) \cdot e_0 + (v_1^2) \cdot e_1 + \cdots + (v_{\tau-1}^\tau) \cdot e_{\tau-1} \quad \text{(A.1)}$$

for some τ and some random vectors v^j that are bounded and measurable with respect to B^*. Since $(v_1^2) \cdot e_1 + \cdots + (v_{\tau-1}^\tau) \cdot e_{\tau-1}$ is in U^*,

$$\operatorname{proj}(u \mid U^*) = \operatorname{proj}[(v_0^1) \cdot e_0 \mid U^*] + (v_1^2) \cdot e_1 + \cdots + (v_{\tau-1}^\tau) \cdot e_{\tau-1}. \quad \text{(A.2)}$$

Hence $u - \operatorname{proj}(u \mid U^*) = (v^1 \cdot e) - \operatorname{proj}[(v^1) \cdot e \mid U^*]$.

Using results of Hansen and Richard (1987) (see their lemmas A.2 and A.3 and especially the proof of theorem A.2), it can be shown that $\operatorname{proj}[(v^1) \cdot e \mid U^*] = (v^1) \cdot \operatorname{proj}(e \mid U^*)$. Therefore, U^c is the mean-square closure of the set of all random variables that can be expressed as $(v) \cdot e^*$, where v is a random vector that is bounded and measurable with respect to B^*. Taking a mean-squared Cauchy sequence of the form $\{(v^j) \cdot e^*: j \geq 1\}$ and using the definition of Φ^+, it can be shown that

$$U^c \subseteq \{v \cdot e^+: v \text{ is an } n\text{-dimensional random vector that is measurable with respect to } B^* \text{ and satisfies } E\,|v|^2 < \infty\}. \quad \text{(A.3)}$$

To show that the right side of (A.3) is a subset of U^c, let $u^+ \equiv v \cdot e^+ = v'(\Phi^+)^{-1}e^*$ be an element in the set on the right side of (A.3). Then $u^j \equiv 1_{\{|v'(\Phi^+)^{-1}|<j\}} v'(\Phi^+)^{-1}e^*$ is an element of U^c. Note that $|u^j - u^+|^2 \leq (u^+)^2$ and $E[(u^+)^2] < \infty$. Hence by the dominated convergence theorem, $\lim_{j\to\infty} E(|u^j - u^+|^2) = 0$ and u^+ is in U^c.

Our second task is to use this characterization of U^c to represent e. Note that each entry of e is in U^*_{-1} and that

$$U^*_{-1} = U^c \oplus U_1^c \oplus \cdots \oplus U_{s-1}^c \oplus U_s^*. \quad \text{(A.4)}$$

Since (A.4) gives an orthogonal decomposition of U^*_{-1} and e is in U^*_{-1},

$$e = \sum_{j=0}^{s-1} \operatorname{proj}(e \mid U_j^c) + \operatorname{proj}(e \mid U_s^*). \quad \text{(A.5)}$$

Also, note that $\operatorname{proj}(e \mid U_s^*) = 0$, since

$$\begin{aligned} E(v_\tau' e_\tau e') &= E[E(v_\tau' e_\tau e' \mid B_{\tau-s})] \\ &= E[v_\tau' E(e_\tau \mid B_{\tau-s}) e'] = 0 \end{aligned} \quad \text{(A.6)}$$

for any $\tau \geq s$ and any random vector v that is bounded and measurable with respect to B^*. Representation (A.5) combined with (A.4) implies (4.5).

Our third task is to prove that λ^0 is nonsingular. Since assumption 7 is satisfied, $\lambda^0 e^+ = \operatorname{proj}(e \mid U^c) = e^* = (\Phi^+)^{-1}e^+$. Since $E(e^+e^+ \mid B^*) = I$ is satisfied, $\lambda^0\lambda^{0\prime} = (\Phi^+)^{-1}(\Phi^{+1\prime})^{-1}$. Therefore, λ^0 is nonsingular.

Our fourth task is to prove that $E(e^+e_j^{+\prime} \mid B^*) = 0$ for all $j \neq 0$. The arguments of Hansen and Richard (1987) show that $E(v'e^+e_{-\tau}^{+\prime} \mid B_{-\tau}^*) = 0$ for any random vector v that is measurable with respect to B^* and has a finite second moment and any $\tau > 0$. The claim follows immediately for $j > 0$. For the case when $j < 0$, suppose to the contrary that $E(e^+e_j^{+\prime} \mid B^*) \neq 0$. Then there exists a bounded random vector v^+ that is measurable with respect to B^* such that at least one entry of $v^{+\prime}E(e^+e_j^{+\prime} \mid B^*)$ is nonnegative and strictly positive with positive probability. Hence $E[v^{+\prime}E(e^+e_j^{+\prime} \mid B^*) \mid B_j^*] = E(v^{+\prime}e^+e_j^{+\prime} \mid B_j^*) \neq 0$, which is a contradiction.

Our final task is to show that the λ^j's can be chosen to be constant when Assumption 4′ is satisfied. Arguments of Hansen and Richard (1987) can be used to demonstrate that the forward Wold representation for e is a conditional forward moving-average representation for e.

Proof of Lemma 4.2. Using repeated applications of the law of iterated expectations and Lemma 4.1, it can be shown that

$$E[|z'e|^2] = E\left[\left|z' \sum_{j=0}^{s-1} \lambda_j^j e_j^+\right|^2\right] = \sum_{j=0}^{s-1} E[|z'\lambda_j^j|^2]. \quad \text{(A.7)}$$

Relation (A.7) in turn implies that $|z'\lambda_j^j|$ has a finite second moment for any z in Z^+ and any $j = 0, 1, \ldots, s-1$. Substituting for e and e_j from (4.5), if follows that

$$E[(z'e)(e_j'z_j^*)] = E\left[z' \left(\sum_{\tau=0}^{s-1} \lambda_\tau^\tau e_\tau^+\right)\left(\sum_{\tau=0}^{s-1} \lambda_{\tau+j}^\tau e_{\tau+j}^+\right)' z_j^*\right]. \quad \text{(A.8)}$$

Now $z'(\lambda_\tau^\tau)e_\tau^+$ is in U_τ^c and $\{U_\tau^c : -\infty < \tau < +\infty\}$ is a sequence of mutually orthogonal linear spaces. Hence for $0 \le j \le s-1$,

$$\begin{aligned} E[(z'e)(e_j'z_j^*)] &= E\left[z' \sum_{\tau=j}^{s-1} \lambda_\tau^\tau(\lambda_\tau^{\tau-j})'z_j^*\right] \\ &= E\left\{z' E\left[\sum_{\tau=j}^{s-1} \lambda_\tau^\tau(\lambda_\tau^{\tau-j})'z_j^* \mid B^*\right]\right\}, \quad \text{(A.9)} \end{aligned}$$

since $E(e^+e^+ \mid B^*) = I$ is satisfied. Using the fact that S is measure-preserving gives

$$E[(z'e)(e_j'z_j^*)] = E\left[\sum_{\tau=j}^{s-1} (z'_{-\tau})\lambda^\tau(\lambda^{\tau-j})'z_{j-\tau}^*\right]. \quad \text{(A.10)}$$

Performing a similar calculation for $s - 1 \le j < 0$ and summing (A.9) and (A.10) across j gives the result.

Next we establish an intermediate result that is used in proving Lemma 4.3 and Theorem 4.1.

Lemma A.1. Assumption 3 implies that there exists a z in Z for which $E(z'\ d^*)$ is nonsingular.

Proof. Define $d^l \equiv 1_{\{|d^*|<l\}}d^*$. Then $|d^l|$ is dominated by l and $E(d^{l\prime}d^l) = E(d^{l\prime}d^*)$ is nonsingular. Now $|E(d^l \mid B^\tau)|$ is dominated by l and $E(d^l \mid B^\tau)$ is a member of the index set Z^τ for any positive integer τ. Also, $\{E(d^l \mid B^\tau) : \tau \ge 1\}$ converges in $L^2(\text{Pr})$ to d^l. Hence a subsequence $\{E(d^l \mid B^{\tau(j)}) : j \ge 1\}$ must converge almost surely to d^l. Since $|d|$ has a finite first moment, $|d^*|$ also has a finite first moment. It follows that $\{E[E(d^l \mid B^{\tau(j)})'d^*] : j \ge 1\}$ converges to $E(d^{l\prime}d^*)$. Consequently, there exists a j such that $E[E(d^l \mid B^{\tau(j)})'d^*]$ is nonsingular. Note that $E(d^l \mid B^{\tau(j)})$ is in Z.

Proof of Lemma 4.3. Choose z^* in Z such that $E(z^{*\prime}d^*)$ is nonsingular (see Lemma A.1). Then $z^+ \equiv z^*[E(d^{*\prime}z^*)]^{-1}$ is also in Z. To prove that $E(d^{+\prime}d^+)$ is nonsingular, we suppose to the contrary that this matrix is singular. Then there is a $k \times k$ matrix of real numbers, not all 0, such that $cd^{+\prime}$ is 0 almost surely. Now $E(z^{+\prime}d^*c') = c'$. Substituting z^+ for z into the right side of (4.7) gives $E(z^{+\prime}d^*c') = E[\Lambda(z^+)'d^*c'] = 0$, which is a contradiction.

Proof of Lemma 4.4. First, we prove that $\Lambda[\Psi(z)] = z$ for any z in D. Note that the form of the ψ^j's implies that

$$\begin{aligned} \sum_{\tau=0}^{j} \lambda_{\tau-j}^\tau \psi^{j-\tau} &= I \quad \text{if } j = 0 \\ &= 0 \quad \text{if } j > 0. \quad \text{(A.11)} \end{aligned}$$

Recall that $\{\sum_{j=0}^{l} E(\psi_j^j z_j \mid B^*) : l \ge 1\}$ converges to $\Psi(z)$ in $L^2(\text{Pr})$ and that each of the λ^τ's has finite second moments. Hence for each τ, $\{\lambda_\tau^\tau \sum_{j=0}^{l} E(\psi_{j+\tau}^j z_{j+\tau} \mid B_\tau^*) : l \le 1\}$ converges in $L^1(\text{Pr})$ to $\lambda_\tau^\tau[\Psi(z)]_\tau$, implying that $\{E[\lambda_\tau^\tau \sum_{j=0}^{l} E(\psi_{j+\tau}^j z_{j+\tau} \mid B_\tau^*) \mid B^*] : l \ge 1\}$ converges in $L^1(\text{Pr})$ to $E\{\lambda_\tau^\tau[\Psi(z)]_\tau \mid B^*\}$. Therefore, $\{\Lambda(\sum_{j=0}^{l} \psi_j^j z_j) : l \ge 1\}$ converges in $L^1(\text{Pr})$ to $\Lambda[\Psi(z)]$. Now,

$$\begin{aligned} \Lambda\left[\sum_{j=0}^{l} \psi_j^j z_j\right] &= \sum_{\tau=0}^{s-1} E\left\{\lambda_\tau^\tau \sum_{j=\tau}^{\tau+l} E[\psi_j^{j-\tau} z_j \mid B_\tau^*] \mid B^*\right\} \\ &= \sum_{j=0}^{l} E\left\{\sum_{\tau=0}^{j} \lambda_\tau^\tau \psi_j^{j-\tau} z_j + \sum_{\tau=1}^{s-1} \lambda_\tau^\tau \sum_{j=l+1}^{l+\tau} \psi_j^{j-\tau} z_j \mid B^*\right\} \\ &= z + \sum_{\tau=1}^{s-1} E\left\{\lambda_\tau^\tau \sum_{j=l+1}^{l+\tau} \psi_j^{j-\tau} z_j \mid B^*\right\}, \quad \text{(A.12)} \end{aligned}$$

where the second term on the right side converges in $L^1(\text{Pr})$ to 0 as $l \to \infty$.

Second, we can prove that $\Lambda^-[\Psi^-(z)] = z$ for any z in D^- by an analogous argument, once we find the relation that parallels (A.11). Fix any $j \ge 0$, and let

$$F = \begin{bmatrix} \lambda_0^0 & \lambda_1^1 & \cdots & \lambda_j^j \\ 0 & \lambda_1^0 & \cdots & \lambda_j^{j-1} \\ 0 & 0 & & \lambda_j^0 \end{bmatrix}, \qquad G = \begin{bmatrix} \psi_0^0 & \psi_1^1 & \cdots & \psi_j^j \\ 0 & \psi_1^0 & \cdots & \psi_j^{j-1} \\ 0 & 0 & & \psi_j^0 \end{bmatrix}. \quad \text{(A.13)}$$

Then (A.11) implies that $FG = I$. Hence $F = G^{-1}$ and $GF = I$. Multiplying the first row partition of G by the last column partition of F gives the relation we need:

$$\begin{aligned} \sum_{\tau=0}^{j} \psi_\tau^\tau \lambda_j^{j-\tau} &= I \quad \text{if } j = 0 \\ &= 0 \quad \text{if } j > 0. \quad \text{(A.14)} \end{aligned}$$

Proof of Theorem 4.1. Lemma 4.4 guarantees that $E(z'd^*) = E\{z'\Lambda[\Psi(d^*)]\}$ for any z in Z. By mimicking the proof of Lemma 4.2, we can show that $E\{z'\Lambda[\Psi(d^*)]\} = E[\Lambda^-(z)'\Psi(d^*)]$. It follows that $[E(\Psi(d^*)'\Psi(d^*))]^{-1}$ is a lower bound as shown in Section 4.1.

Next, we prove the existence of $\inf(Z)$ by applying lemma 4.3 of Hansen (1985). To apply this lemma we must verify that Z satisfies Hansen's properties 3.1–3.5. It is straightforward to show that Z satisfies properties 3.1, 3.2, and 3.4, and Lemma A.1 shows that Z satisfies property 3.3. To show that property 3.5 is satisfied, let $\{z^j : j \ge 1\}$ be any sequence in Z for which $\{\langle z^j \mid z^j\rangle : j \ge 1\}$ converges to 0. Then Lemma 4.2 implies that $\{E[|\Lambda^-(z^j)|^2] : j \ge 1\}$ converges to 0. By the Cauchy–Schwarz inequality, $|E[\Lambda^-(z^j)'d^*]| \le E[|d^*|^2]^{1/2}E[|\Lambda^-(z^j)|^2]^{1/2}$. It follows from (4.7) that $\{E(z^{j\prime}d^*) : j \ge 1\}$ converges to 0, which establishes that Z satisfies property 3.5.

Proof of Theorem 4.2. The proof is complete once we show that for any z in Z^+, there is a sequence $\{z^l : l \ge 1\}$ in Z that converges in $L^\eta(\text{Pr})$ to z. The sequence $\{E(z \mid B^\tau) : \tau \ge 1\}$ is in Z and converges in $L^2(\text{Pr})$ to z. Hence a subsequence converges almost surely to z. In the special case in which $|z|$ is essentially bounded, the same bound applies to $E(z \mid B^\tau)$ for $\tau \ge 1$. Hence the dominated convergence theorem guarantees that a subsequence of $\{E(z \mid B^\tau) : \tau \ge 1\}$ converges in $L^\eta(\text{Pr})$ to z. In the general case we can first construct a sequence $\{z^j : j \ge 1\}$ in Z^+ such that every member of the sequence is essentially bounded and the sequence converges in $L^\eta(\text{Pr})$ to z. For instance, let $z^j = 1_{\{|z|<j\}}z$. Then we can approximate the z^j's as previously suggested.

[*Received December 1985. Revised February 1988.*]

REFERENCES

Amemiya, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955–968.

Barro, R. J. (1981), "On the Predictability of Tax-Rate Changes," unpublished manuscript.

Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. (1983), "Information and Asymptotic Efficiency in Parametric–Nonparametric Models," *The Annals of Statistics*, 11, 432–452.

Billingsley, P. (1961), "The Lindeberg–Levy Theorem for Martingales," *Bulletin: American Mathematical Society*, 12, 788–792.

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

Doob, J. L. (1953), *Stochastic Processes*, New York: John Wiley.

Dunn, K. B., and Singleton, K. J. (1986), "Modeling the Term Structure of Interest Rates Under Nonseparable Utility and Durability of Goods," *Journal of Financial Economics*, 17, 27–55.

Eichenbaum, M. S., and Hansen, L. P. (1987), "Estimating Models With Intertemporal Substitution Using Aggregate Time Series Data," Working Paper 2181, National Bureau of Economic Research, Cambridge, MA.

Eichenbaum, M. S., Hansen, L. P., and Singleton, K. J. (1988), "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

Gordin, M. I. (1969), "The Central Limit Theorem for Stationary Processes," *Soviet Mathematics Doklady*, 10, 1174–1176.

Grossman, S., Melino, A., and Shiller, A. (1987), "Estimating the Continuous-Time Consumption-Based Asset-Pricing Model," *Journal of Business & Economic Statistics*, 5, 315–327.

Hall, R. (1988), "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–357.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

——— (1985), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30, 203–238.

Hansen, L. P., and Hodrick, R. J. (1983), "Risk Averse Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models," in *Exchange Rates and International Macroeconomics*, ed. J. A. Frenkel, Chicago: University of Chicago Press, pp. 113–142.

Hansen, L. P., and Richard, S. F. (1987), "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

Hansen, L. P., and Sargent, T. J. (1982), "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models," *Journal of Monetary Economics*, 9, 263–296.

Hansen, L. P., and Singleton, K. J. (1988), "Efficient Estimation of Linear Asset Pricing Models With Moving Average Errors," unpublished manuscript.

Hayashi, F., and Sims, C. A. (1983), "Nearly Efficient Estimation of Time Series Models With Predetermined, but Not Exogenous, Instruments," *Econometrica*, 51, 783–798.

Heaton, J. C., and Ogaki, M. (1988), "Efficiency Bound Calculations for a Simple Model With Conditional Heteroskedasticity," unpublished manuscript.

Jorgenson, D. W., and Laffont, J. (1974), "Efficient Estimation of Non-Linear Simultaneous Equations With Additive Disturbances," *Annals of Economic and Social Measurement*, 3, 615–640.

Levit, B. Y. (1975), "On the Efficiency of a Class of Non-Parametric Estimates," *Theory of Probability and Its Applications*, 2, 723–740.

Newey, W. K. (1987), "Efficient Estimation of Models With Conditional Moment Restrictions," working paper, Princeton University, Dept. of Economics.

Robinson, P. M. (1987a), "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica*, 55, 875–891.

——— (1987b), "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," unpublished manuscript.

Rozanov, Y. A. (1967), *Stationary Random Processes*, San Francisco: Holden-Day.

Sargan, J. D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Stein, C. (1956), "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), Berkeley: University of California Press, pp. 187–195.

Stoica, P., Söderström, T., and Friedlander, B. (1985), "Optimal Instrumental Variable Estimates of the AR Parameters of an ARMA Process," *IEEE Transactions of Automatic Control*, 30, 1066–1074.

Tauchen, G. (1986), "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data," *Journal of Business & Economic Statistics*, 4, 397–416.