

On the Asymptotic Efficiency of GMM*

Marine Carrasco[†] Jean-Pierre Florens[‡]
The University of Rochester Université de Toulouse

Preliminary and Incomplete: March 2003

Abstract

This paper derives conditions under which the generalized method of moments (GMM) estimator is as efficient as the maximum likelihood estimator (MLE). The data are supposed to be drawn from a parametric family and to be stationary Markov. We study the efficiency of GMM in a general framework where the set of moment conditions may be finite, countable infinite, or a continuum. Our main result is the following. GMM estimator is efficient if and only if the true score belongs to the closure of the linear space spanned by the moment conditions. This result extends former ones in two dimensions: (a) the moments may be correlated, (b) the number of moment restrictions may be infinite. It suggests a way to construct estimators that are as efficient as MLE. In the last part of this paper, we show how to calculate the greatest lower bound of instrumental variable estimators.

KEY WORDS: Asymptotic efficiency, GMM, infinity of moment conditions, reproducing kernel Hilbert space, efficiency bound.

*Carrasco gratefully acknowledges financial support from the National Science Foundation, grant # SES-0211418.

[†]University of Rochester, Department of Economics, Harkness 226, Rochester, NY 14627.
Email: cscoc@troi.cc.rochester.edu.

[‡]Université de Toulouse, IDEI, GREMAQ, 21 allée de Brienne, 31000 Toulouse, France.
Email: florens@cict.fr.

1. Introduction

Since the seminal paper by Hansen (1982), the Generalized Method of Moments (GMM) has been widely applied in econometrics. However, GMM estimator is usually inefficient relative to the Maximum Likelihood estimator (MLE) due to the arbitrary choice of moment conditions. In this paper, we shall study the efficiency of GMM estimators. For this purpose, we assume that the observations are drawn from a parametric family and are stationary Markov of arbitrary order L . Suppose that moment restrictions indexed by an index parameter τ in I are available such that

$$E^{\theta_0} [h(\tau; \theta_0)] = 0, \tau \in I.$$

If I is the set $\{1, 2, \dots, M\}$, there is a finite number of moments (M finite) or a countable infinity of moments (M infinite). If I is an interval of R or R itself, there is a continuum of moments. When I is finite, $h(\theta)$ denotes the vector where the moment functions are stacked. In this case, Hansen (1982) shows that the optimal GMM estimator is obtained by minimizing the quadratic form

$$h_T(\theta)' K_T^{-1} h_T(\theta)$$

where $h_T(\theta)$ is the sample estimator of $E^{\theta_0} [h(\theta)]$ and K_T is a consistent estimator of K , the covariance matrix of $h(\theta_0)$. Under some regularity conditions, the GMM estimator is \sqrt{T} -consistent and asymptotically normal with asymptotic variance $\left[E^{\theta_0} [\nabla_{\theta} h]' K^{-1} E^{\theta_0} [\nabla_{\theta} h] \right]^{-1}$. These results can be generalized to the case where I is infinite. We propose a general framework that encompasses both the finite and infinite cases. Let $h(\theta)$ denote $\{h(\tau; \theta) : \tau \in I\}$ and $h_T(\theta)$ be the sample counterpart of $E^{\theta_0} [h(\theta)]$. It is shown that the optimal GMM estimator is solution of

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \|h_T(\theta)\|_K^2$$

where $\|\cdot\|_K^2$ denotes the norm in the Reproducing Kernel Hilbert Space (RKHS) associated with the covariance operator, K , of $h(\theta_0)$. Its asymptotic variance is $\left[\left\| E^{\theta} [\nabla_{\theta} h(\theta)] \right\|_K \right]^{-1} \Big|_{\theta=\theta_0}$. We do not discuss the implementation of the GMM estimator. In the case of a continuum of moment conditions, the GMM estimation is detailed in Carrasco and Florens (2000, 2001) and Carrasco, Chernov, Ghysels, and Florens (2002). In the case of a countable infinity of moment restrictions, it is customary to truncate the sequence of moment conditions and to let the number

of moments grow with the sample size, see Chamberlain (1985) and Gallant and Long (1997).

The question we address in this paper is a simple one. Under which conditions on the set of moment restrictions is the (optimal) GMM estimator asymptotically as efficient as the MLE? To answer this question, it suffices to compare the asymptotic variance of $\hat{\theta}$, $\left[\left\|E^{\theta_0}[\nabla_{\theta}h]\right\|_K\right]^{-1}$ with the Cramer Rao efficiency bound. To do so, we use a characterization of the norm in a RKHS proposed by Parzen (1970). The main result of this paper is the following. The GMM estimator reaches the Cramer Rao bound if and only if the Data Generating Process (DGP) score belongs to the closure of the set of moment conditions. Although this result may sound familiar, it generalizes former results in two dimensions: (a) the moment conditions may be autocorrelated, (b) the number of moment conditions may be infinite since we allow for both a continuum and a countable infinity of moment conditions.

Our result can be particularly useful to devise an efficient GMM estimator in the cases where MLE is either intractable or burdensome. For instance, it permits to establish that GMM estimator based on power functions or on the characteristic function is asymptotically efficient. In particular, it implies that the GMM estimator based on the joint characteristic function of $(X_t, X_{t-1}, \dots, X_{t-L})$ is asymptotically efficient when X_t is Markov of order L . To our knowledge, we are the first to give a formal proof of this result¹. As the expectation of power functions or the value of the characteristic function is not always known, it may be necessary to estimate them by simulations. The simulated method of moments (see Gourieroux and Monfort, 1996 and the survey by Carrasco and Florens, 2002) consists in matching the sample moment with an estimated moment based on simulated data. As the simulated method of moments is asymptotically equivalent to the usual GMM when the number of simulations goes to infinity, our results also apply to these estimators.

Our approach is strictly parametric and we do not give semiparametric efficiency bound as e.g. in Chamberlain (1987). However, Chamberlain shows that his bound can be approached using an increasing sequence of moment restrictions if the set of moment restrictions is “complete”, that is, if it spans the whole space L^2 . This condition is similar to our sufficient condition in Corollary 4.2. Hansen (1985) gives the greatest lower bound for the asymptotic covariance matrices of GMM estimators when an infinity of moment conditions is available. He does

¹Feuerverger (1990) claims this result but does not give a formal proof.

not impose enough structure on the estimation problem to discuss efficiency. He assumes that the moment $\{h\}$ can be approximated by martingale difference sequences (m.d.s.). The lower bound is obtained by solving an equation similar to our condition $C(g)$ in Section 3. Our result does not use an approximation of $\{h\}$ by a m.d.s.. Tauchen (1997) has the same efficiency result as us in an iid environment with a finite number of moments. Gallant and Long (1997) discuss the efficiency of the efficient method of moments (EMM) and show that if the auxiliary model is sufficiently rich to nest the DGP then efficiency follows. Note that our results show that it is sufficient that the set of moment restrictions encompasses the DGP score.

The rest of the paper is organized as follows. Section 2 presents a general framework for GMM estimation with a finite or infinite number of moments. Section 3 characterizes the norm in a RKHS. Section 4 gives the efficiency results and illustrates their usefulness with simple examples. Section 5 gives a method for deriving the greatest lower bound of instrumental variable estimators in the spirit of Hansen (1985). The proofs of the main propositions are in appendix.

2. A general framework for GMM estimation

Assume that the process $\{X_t\}$ is stationary, ergodic, and Markov of order L . X may be multivariate so that $X_t \in R^p$. The observations are given by $\{x_1, \dots, x_T\}$. The distribution of X_t is indexed by θ , its conditional pdf is denoted

$f_\theta(x_t|x_{t-1}, x_{t-2}, \dots, x_{t-L}; \theta)$. Let $\theta \in \Theta \subset R^q$. Let $Y_t = (X_t, X_{t-1}, \dots, X_{t-N})$ for some finite N and let $f_\theta(y)$ be its pdf. E^{θ_0} denotes the expectation with respect to $f_{\theta_0}(y)$. Define $L^2(Y)$ as the set $\{G(Y) | E^{\theta_0}[G(Y)^2] < \infty\}$.

Inference is based on a set of moment conditions

$$E^\theta [h(\tau, Y, \theta)] = 0, \tau \in I \tag{2.1}$$

where $h(\tau, Y, \theta)$ is a function from $I \times R$ to R or C . Assume that $h(\tau, Y, \theta) \in L^2(Y)$, for any $\tau \in I$ and $\theta \in \Theta$.

Identification condition:

$$E^{\theta_0} [h(\tau, Y, \theta)] = 0, \forall \tau \in I \Rightarrow \theta = \theta_0. \tag{2.2}$$

For efficiency consideration, we might want to base the estimation on a full set of moment conditions indexed by $\tau \in I$ even if identification is achieved from a finite subset of I . Let I be a measurable space provided with a measure π . The

space of reference is $L^2(I, \pi)$ the Hilbert space of complex valued functions such that

$$L^2(I, \pi) = \left\{ g : I \rightarrow \mathbb{C} \mid \int_I |g(\tau)|^2 \pi(d\tau) < \infty \right\}.$$

The inner product on this space is

$$\langle f, g \rangle = \int_I f(\tau) \overline{g(\tau)} \pi(d\tau).$$

where \bar{g} denotes the complex conjugate of g . The term $\|g\|^2$ denotes the norm of g in $L^2(I, \pi)$. When g is a vector, we define

$$\|g\|^2 = \int_I g(\tau)' \overline{g(\tau)} \pi(d\tau)$$

It is assumed that $h(\tau, Y, \theta) \in L^2(I, \pi)$. The norm in $L^2(I, \pi)$ may take different forms.

Examples of spaces $L^2(I, \pi)$:

- Finite number of restrictions: $I = \{1, \dots, M\}$, π is a uniform probability measure. The norm is the euclidean norm $\|g\|^2 = \frac{1}{M} \sum_{\tau=1}^M |g(\tau)|^2$.
- Countable infinity of restrictions: $I = \mathbb{N}^*$, π is a Poisson probability measure such that $\pi\{\tau\} = e^{-1}/\tau!$, $\tau = 1, 2, \dots$. The norm is $\|g\|^2 = \sum_{\tau=1}^{\infty} |g(\tau)|^2 \pi(\tau)$.
- Continuum of restrictions: $I = [0, 1]$, π is Lebesgue measure on $[0, 1]$. The norm is $\|g\|^2 = \int_0^1 |g(\tau)|^2 d\tau$.
- Continuum of restrictions on R : $I = R$, π is a probability measure on R . The norm is $\|g\|^2 = \int |g(\tau)|^2 \pi(d\tau)$.

In the following, we use the generic notation $\int |g(\tau)|^2 \pi(d\tau)$ even in the discrete case. Let K be the covariance operator with kernel $k(\tau, \lambda) = E^{\theta_0} \left[h(\tau, Y, \theta_0) \overline{h(\lambda, Y, \theta_0)} \right]$. It associates to a function g of $L^2(I, \pi)$ the element of $L^2(I, \pi)$:

$$Kg(\tau) = \int_I k(\tau, \lambda) g(\lambda) \pi(d\lambda).$$

For instance, in the case of a countable infinity of moments $Kg(\tau) = \sum_{s=1}^{\infty} k(\tau, \lambda) g(\lambda) \pi(\lambda)$. Hence K is an infinite-dimensional matrix.

Let $h_T(\cdot, \theta) = \frac{1}{T} \sum_{t=1}^T h(\cdot, y_t, \theta)$. The GMM estimators are solutions of

$$\min_{\theta} \langle Bh_T(\cdot, \theta), Bh_T(\cdot, \theta) \rangle$$

where B is an operator from $L^2(I, \pi)$ in $L^2(I, \pi)$. Denote $K^{-1}f$ the solution g (when it exists) of the equation $Kg = f$ and $K^{-1/2} = (K^{-1})^{1/2}$. The optimal GMM estimator in the class of GMM estimators is obtained for $B = K^{-1/2}$:

$$\hat{\theta} = \arg \min_{\theta} \langle K^{-1/2}h_T(\cdot, \theta), K^{-1/2}h_T(\cdot, \theta) \rangle. \quad (2.3)$$

In the case of a finite number of moments, we have the usual GMM objective function $h_T'K^{-1}h_T$ where $h_T = (h_T(1, \theta), \dots, h_T(M, \theta))'$ and K^{-1} is the inverse of the covariance matrix of h . When I is infinite dimensional, $K^{-1/2}$ is not defined on the whole space $L^2(I, \pi)$ but on a subset denoted $\mathcal{H}(K)$ and called the Reproducing Kernel Hilbert Space (RKHS). We use the notation $(g, f)_K \equiv \langle K^{-1/2}g, K^{-1/2}f \rangle$ for the inner product in the RKHS associated with K .

Under some regularity conditions, $\hat{\theta}$ satisfies

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1}) \quad (2.4)$$

where Σ is the $q \times q$ -matrix with (i, j) element equal to

$$\left(E^{\theta}(\nabla_{\theta_i} h(\theta)), E^{\theta}(\nabla_{\theta_j} h(\theta)) \right)_K \Big|_{\theta=\theta_0}. \quad (2.5)$$

For I finite, the asymptotic variance can be rewritten as

$$\left\{ E^{\theta_0}(\nabla_{\theta} h) K^{-1} E^{\theta_0}(\nabla_{\theta} h)' \right\}^{-1}$$

The result of (2.4) is shown in Carrasco and Florens (2000) for $I = [0, 1]$, in Carrasco and Florens (2001) and Carrasco and al. (2002) for $I = R$. In presence of a countable infinity of moments, the result can be established using the same approach as in Carrasco and Florens (2000). As our focus is on efficiency, we do not discuss the implementation of the GMM estimator. In the sequel, we assume that the result (2.4) holds.

3. Norm in a RKHS

Our aim is to compute $\langle f, g \rangle_K$ that show up in the inverse of the asymptotic variance of the GMM estimator. Calculation of this variance permits to establish

whether the estimator is efficient or not. This section gives results on the norm in a RKHS that appear in Parzen (1970) and Saitoh (1997). The setting is very general, and goes far beyond the GMM context.

Besides the space $L^2(I, \pi)$ introduced in the previous section (on which the norm $\|\cdot\|^2$ has been defined), we need to define several spaces:

- Let k be a Hermitian kernel from $I \times I$ into C that is

$$k(\tau, \lambda) = \overline{k(\lambda, \tau)}.$$

Assume moreover that k is a L^2 -kernel that is

$$\int \int |k(\tau, \lambda)|^2 \pi(d\tau) \pi(d\lambda) < \infty.$$

Therefore k defines a Hilbert-Schmidt operator K :

$$\begin{aligned} K & : L^2(I, \pi) \rightarrow L^2(I, \pi) \\ f & \rightarrow g(\tau) = \int_I k(\tau, \lambda) f(\lambda) \pi(d\lambda). \end{aligned}$$

The operator K has a countable spectrum. Let $\{\phi_j, \mu_j\}$ be the eigenfunctions and eigenvalues of K .

- We define the RKHS on $L^2(I, \pi)$ associated with k as the Hilbert space:

$$\mathcal{H}(K) = \left\{ f \in L^2(I, \pi) \mid \sum_j \frac{|\langle f, \phi_j \rangle|^2}{\mu_j} < \infty \right\}.$$

The inner product on this space is defined as

$$(f, g)_K = \langle K^{-1/2} f, K^{-1/2} g \rangle = \sum_j \frac{\langle f, \phi_j \rangle \overline{\langle g, \phi_j \rangle}}{\mu_j}.$$

The properties of $\mathcal{H}(K)$ are (i) $k(\cdot, \tau) \in \mathcal{H}(K)$, (ii) (reproducing property) $f(\tau) = (f, k(\cdot, \tau))_K$.

- Let H^0 be a complex Hilbert space with inner product $(\cdot, \cdot)_{H^0}$.
- Assume there exists $\{h(\tau), \tau \in I\}$ a family of vectors in H^0 such that k takes the form:

$$k(\tau, \lambda) = (h(\tau), h(\lambda))_{H^0}.$$

- Define $L^2(h)$ the closure of the subspace of H^0 spanned by $\{h(\tau), \tau \in I\}$, it consists of combinations (for some n , real $\omega_1, \dots, \omega_n$ and indices τ_1, \dots, τ_n) of the form

$$v = \sum_{l=1}^n \omega_l h(\tau_l).$$

and limits in mean square of such finite linear combinations.

Examples of kernels k include various covariance kernels. If there is no auto-correlation and h is real, we have

$$(h(\tau), h(\lambda))_{H^0} = E^{\theta_0} (h(\tau, Y_t, \theta_0) h(\lambda, Y_t, \theta_0))$$

and H^0 is the Hilbert space of real functions $\varphi(Y)$ such that $E^{\theta_0} (\varphi(Y)^2) < \infty$. If h is complex, we have

$$(h(\tau), h(\lambda))_{H^0} = E^{\theta_0} (h(\tau, Y_t, \theta_0) \overline{h(\lambda, Y_t, \theta_0)})$$

where $\overline{h(\lambda)}$ denotes the complex conjugate of $h(\lambda)$. In that case, H^0 is the Hilbert space of complex-valued functions $\varphi(Y)$ such that $E^{\theta_0} (|\varphi(Y)|^2) < \infty$. Finally, if there are autocorrelations, we have

$$(h(\tau), h(\lambda))_{H^0} = \sum_{j=-\infty}^{\infty} E^{\theta_0} (h(\tau, Y_0, \theta_0) \overline{h(\lambda, Y_j, \theta_0)})$$

and H^0 is the Hilbert space of complex-valued processes $\{\varphi(Y_t), -\infty < t < \infty\}$ such that $\sum_{j=-\infty}^{\infty} E^{\theta_0} (\varphi(Y_0) \overline{\varphi(Y_j)}) < \infty$.

The question of interest is to characterize $(f, g)_K$ for any $f, g \in \mathcal{H}(K)$. To do so, we need the following assumption.

Assumption 3.1. $\{h(\tau) : \tau \in I\}$ is complete in $L^2(h)$ that is for all $G \in L^2(h)$, $(G, h(\tau))_{H^0} = 0$ for all $\tau \in I \Rightarrow G = 0$.

Assumption 3.1 implies that $\{k(\lambda, \tau) : \tau \in I\}$ are linearly independent for all $\lambda \in I$. Note that the reproducing property implies that $\{k(\cdot, \tau) : \tau \in I\}$ spans $\mathcal{H}(K)$. As a result of the linear independence, each element of $\mathcal{H}(K)$ has a unique representation as $\sum_j \alpha_j k(\cdot, \tau_j)$ for some constants $\alpha_1, \alpha_2, \dots$

Let

$$C(g) = \left\{ G \in H^0 : g(\tau) = (G, h(\tau))_{H^0}, \quad \forall \tau \in I \right\}.$$

Proposition 3.1. *Suppose that Assumption 3.1 holds.*

(a) $\mathcal{H}(K)$ and $L^2(h)$ are isometrically isomorphic, i.e. there exists a one-to-one linear mapping \tilde{J} (not necessarily unique) from $\mathcal{H}(K)$ to $L^2(h)$ such that

$$(f, g)_K = \left(\tilde{J}(f), \tilde{J}(g) \right)_{H^0}$$

for all $f, g \in \mathcal{H}(K)$. \tilde{J} is referred to as Hilbert space isomorphism from $\mathcal{H}(K)$ to $L^2(h)$.

(b) Let J be the Hilbert space isomorphism from $\mathcal{H}(K)$ to $L^2(h)$ that transforms $k(\cdot, \tau)$ into $h(\tau)$, i.e. $J(k(\cdot, \tau)) = h(\tau)$. We have $J(g) = G$ and $\|g\|_K^2 = \|G\|_{H^0}^2$ where G is the unique element of $C(g) \cap L^2(h)$.

(c) Another characterization of J is the following: $J(g) = \arg \min_{G \in C(g)} \|G\|_{H^0}^2$ and hence $\|g\|_K^2 = \min_{G \in C(g)} \|G\|_{H^0}^2$.

Although the results of Proposition 3.1 are not new, they have not been presented under this compact form before. In appendix, we provide a personal proof of this proposition that we think is useful to understand the role played by the reproducing property. Note that (a) and (c) imply that for any $g_1, g_2 \in \mathcal{H}(K)$ and $G_i = \arg \min_{G \in C(g_i)} \|G\|_{H^0}^2$, $i = 1, 2$, we have

$$(g_1, g_2)_K = (G_1, G_2)_{H^0}.$$

The condition that is satisfied by the elements of $C(g)$ is basically the same condition as in Equation (4.2) of Hansen (1985). Hansen uses this condition to calculate the greatest lower bound for the asymptotic covariance matrices of instrumental variable estimators. He does not discuss (parametric) efficiency.

4. Asymptotic Efficiency of GMM estimators

We consider arbitrary functions $h(\tau, X_t^L, \theta_0)$ that satisfy $E^{\theta_0} h(\tau, X_t^L, \theta_0) = 0$ where X_t^L is the $(L+1)$ -vector of r.v.: $X_t^L = (X_t, X_{t-1}, \dots, X_{t-L})$. We drop θ_0 in $h(\tau, X_t^L, \theta_0)$ in the sequel to simplify notation. The question is: Under which conditions on h is the GMM estimator asymptotically efficient?

Let $L^2(h)$ be the closure of the linear space spanned by $\{h(\tau, x_t^L)\}$, $L^2(h)$ is composed of terms of the form $\sum_{j=1}^n \omega_j h(\tau_j, x_t^L)$ and their limits in L^2 norm.

Assumption 4.1. (i) X_t is stationary, ergodic, and Markov of order L .

- (ii) $f_\theta(x_1, x_2, \dots, x_L | x_{-T+1}, x_{-T+2}, \dots, x_{-T+L}) \xrightarrow{T \rightarrow \infty} f_\theta(x_1, x_2, \dots, x_L)$.
- (iii) $\int \sup_{\theta \in \Theta} \left| h(\tau, x_t^L; \theta) f_\theta(x_0, x_1, \dots, x_L) \right| dx_0 dx_1 \dots dx_L < \infty$

Proposition 4.1. *Assume that Assumption 4.1 holds. The GMM estimator based on $h(\tau, X_t^L; \theta)$, $\tau \in I$ is asymptotically as efficient as MLE if and only if*

$$\frac{\partial \ln f_\theta}{\partial \theta_j}(x_t | x_{t-1}^{L-1}) \in L^2(h) \text{ for all } j = 1, 2, \dots, q.$$

The assumption 4.1 (ii) is equivalent to requiring that $(X_t, X_{t-1}, \dots, X_{t-L+1})$ is strong mixing, see Rosenblatt (1971, page 195). If X_t is geometrically ergodic, this condition will be automatically satisfied. Proposition 4.1 is a generalization to the dynamic case of the efficiency result by Carrasco and Florens (2001) which was proved in an iid environment. A similar result (for a countable sequence of moments) is discussed by Tauchen (1997) in the static case and by Gallant and Long (1997) in the context of moment conditions based on the score of a semi-nonparametric auxiliary model, in the dynamic case. The proof uses the results of Proposition 3.1. It is proved in the two opposite cases, the case where $\{h(\tau, X_t^L)\}$ is a martingale difference sequence:

$$E^\theta \left[h(\tau, X_t^L, \theta) | X_{t-1}^{L-1} \right] = 0$$

and the general case where $\{h(\tau, X_t^L, \theta)\}$ is autocorrelated. Note that the MLE efficiency obtains only for the optimal GMM estimator, that is, if the optimal weighting matrix/operator is used.

The following corollary to Proposition 4.1 gives a sufficient condition for efficiency. $L^2(Y)$ denotes the space of functions of X_t^L that are square integrable.

Corollary 4.2. *Assume that Assumption 4.1 holds. If $\{h(\tau) : \tau \in I\}$ spans $L^2(Y)$ then the GMM estimator based on $\{h(\tau)\}$ is asymptotically efficient.*

Remark: $\{h(\tau) : \tau \in I\}$ spans $L^2(Y)$ is equivalent to say that the family of moment functions $\{h(\tau) : \tau \in I\}$ is complete in $L^2(Y)$, see Chamberlain (1987). It is trivially satisfied if $L^2(h)$ coincides with $L^2(Y)$.

Example 1. (Power functions)

Assume that the data are iid with distribution for which the moment generating function exists. One can reach efficiency using a countable infinity of moments indexed by τ

$$h(\tau, x, \theta) = x^\tau - E^\theta(X^\tau), \tau = 1, \dots$$

The efficiency follows from the properties of Hermite polynomials, under mild conditions on the score, see Sansone page 367-368. The moments $E^\theta(X^\tau)$ are usually unknown and need to be estimated using simulations.

Example 2. (EMM)

We now revert to the case where $\{X_t\}$ is a stationary Markov process of order L . The indirect inference method proposed by Gourieroux, Monfort, and Renault (1993) and the Efficient Method of Moments (EMM) proposed by Gallant and Tauchen (1996) are methods designed to handle cases for which the MLE is intractable, these methods use an auxiliary model (easy to estimate) indexed by a parameter λ . The auxiliary model defines a conditional pdf $f_M(y_t|y_{t-1}, \dots, y_{t-L}; \lambda)$ where M denotes the dimension of λ . Denote $\lambda(\theta_0)$ the pseudo-true value of λ (the limit of the QMLE of λ) and denote h_{Mt} the pseudo score:

$$h_{Mt}(\theta_0) = \frac{\partial \ln f_M}{\partial \lambda}(y_t|y_{t-1}, \dots, y_{t-L}; \lambda(\theta_0)).$$

The asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta_0)$ is given by

$$\left\{ E^{\theta_0} \left[\frac{\partial h_{Mt}}{\partial \theta'} \right] \left[\sum_{j=-\infty}^{\infty} E^{\theta_0} (h_{Mt} h'_{Mt-j}) \right]^{-1} E^{\theta_0} \left[\frac{\partial h_{Mt}}{\partial \theta} \right] \right\}^{-1}.$$

By Proposition 4.1, efficiency is achieved if the DGP score belongs to the closure of the set $\{h_{Mt}\}$. This will be satisfied in general if the auxiliary model is based on a seminonparametric model using a Hermite expansion suggested by Gallant and Tauchen (1996). Gallant and Long (1997) show that the variance of the EMM estimator converges to the Cramer Rao efficiency bound when M goes to infinity (see also Gallant and Nychka 1987).

Example 3. (Characteristic function)

Consider moment functions based on the joint characteristic function of $X_t^L = (X_t, X_{t-1}, \dots, X_{t-L})$

$$h(\tau, X_t^L, \theta) = e^{\tau' X_t^L} - E^\theta [e^{\tau' X_t^L}]. \quad (4.1)$$

or the conditional characteristic function

$$h(\tau, x_t^L, \theta) = e^{is' X_{t-1}^{L-1}} [e^{ir X_t} - E^\theta [e^{ir X_t} | X_{t-1}^{L-1}]] \quad (4.2)$$

where $\tau = (s', r)'$. In both cases a continuum of moment conditions is available. Corollary 4.2 implies that GMM estimators based on either moment functions are

asymptotically efficient. This result was shown by Singleton (2001) for (4.2) but is new for (4.1)². Note that the first set of moment conditions are autocorrelated unless the data are iid. On the other hand, the moment conditions (4.2) are martingale difference sequences and hence are uncorrelated. Therefore the second set of moment conditions may be easier to handle than the first one. If the conditional c.f. is unknown, it may be estimated via simulations. This requires drawing from the conditional distribution which is not always possible. An application of the characteristic function to the estimation of iid mixture models is presented by Carrasco and Florens (2001) and of diffusions is given by Singleton (2001) and Carrasco, Chernov, Florens and Ghysels (2002). The last papers allow for the possibility that the characteristic function is unknown and estimated via simulations.

5. Efficiency bound for instrumental variables (incomplete)

We use the results on RKHS to derive a greatest lower bound for instrumental variable estimators in a time series context. Our approach is closely related to that of Hansen (1985), Hansen, Heaton and Ogaki (1988), and West (2001). We specify the same model as in Hansen (1985):

$$e = m(x, \theta)$$

where θ is the parameter of interest. Let Z be the set of instruments so that

$$Eze = 0 \tag{5.1}$$

for all $z \in Z$. The greatest lower bound for instrumental variable estimators based on any subsets of moments of type (5.1) is the variance of the GMM estimator based on all the moment conditions as the variance of the GMM estimator can never decrease when the number of moments increase. So Formula (2.5) and Proposition 3.1 provide a way to compute the efficiency bound of GMM estimators. When $\{h\}$ are martingale difference sequences (m.d.s.), our method is very similar to that of Hansen (1985) as his condition (4.8) corresponds to the condition satisfied by the elements of $C(g)$. When $\{h\}$ are not m.d.s., both approaches are different as Hansen follows Gordin by approximating stationary ergodic processes by m.d.s. Below we give an illustrative example on how to apply our approach to study moment redundancy.

²Feuerverger (1990) claims the result but his proof is heuristic.

Example of application. There have been recently several papers treating of the moment redundancy (Kim, Qian, and Schmidt, 1999, Broze, Francq, and Zakoian, 2001, Hall and Inoue, 2003). We show here how our bound permits to give a necessary and sufficient condition for moment redundancy in a simple framework. Consider the following autoregressive model

$$y_t = \theta y_{t-1} + \varepsilon_t$$

where ε_t satisfies

$$\begin{aligned} E[\varepsilon_t | y_{t-1}, y_{t-2}, \dots] &= 0, \\ E[\varepsilon_t^2 | y_{t-1}, y_{t-2}, \dots] &= a + by_{t-1}^2. \end{aligned}$$

with $a > 0$ and $b \geq 0$. Here $e = \varepsilon$, $Z = \{y_{t-1}, y_{t-2}, \dots\}$, $h(i) = y_{t-i}(y_t - \theta y_{t-1})$, $\nabla_{\theta} h(i) = -y_{t-i}y_{t-1}$, $g(i) = -E(y_{t-i}y_{t-1})$. To compute the bound, we look for an element G of the linear space spanned by $\{h(i), i = 1, 2, \dots\}$ which belongs to $C(g)$. Hence $G = \left(\sum_{j=1}^{\infty} \omega_j y_{t-j}\right) \varepsilon_t$ where $\{\omega_j : j = 1, 2, \dots\}$ are the solutions of

$$\begin{aligned} g(i) &= E(h(i)G), \text{ for all } i = 1, 2, \dots \\ \Leftrightarrow -E(y_{t-i}y_{t-1}) &= E\left(y_{t-i} \left(\sum_{j=1}^{\infty} \omega_j y_{t-j}\right) \varepsilon_t^2\right), \text{ for all } i = 1, 2, \dots \\ \Leftrightarrow -E(y_{t-i}y_{t-1}) &= E\left(y_{t-i} \left(\sum_{j=1}^{\infty} \omega_j y_{t-j}\right) (a + by_{t-1}^2)\right), \text{ for all } i = 1, 2, \dots \end{aligned}$$

The solution if it exists is unique. The efficiency bound is then given by

$$\begin{aligned} (EG^2)^{-1} &= \left\{ E \left[\varepsilon_t^2 \left(\sum_{j=1}^{\infty} \omega_j y_{t-j} \right)^2 \right] \right\}^{-1} \\ &= \left\{ E \left[(a + by_{t-1}^2) \left(\sum_{j=1}^{\infty} \omega_j y_{t-j} \right) \right] \right\}^{-1}. \end{aligned}$$

The OLS estimator will be as efficient as any GMM estimator based on moments $\{h(i), i = 1, 2, \dots\}$ if and only if this bound coincides with the asymptotic variance of OLS that is

$$\left\{ E \left[(a + by_{t-1}^2) \left(\sum_{j=1}^{\infty} \omega_j y_{t-j} \right) \right] \right\}^{-1} = \frac{E(y_{t-1}^2 (a + by_{t-1}^2))}{E(y_{t-1}^2)^2}. \quad (5.2)$$

This condition is satisfied if ε_t is homoskedastic (case where $b = 0$, $\omega_1 = 1/a$, and $\omega_i = 0$ for $i > 1$) and may or may not be satisfied when ε_t is heteroskedastic. Broze, Francq, and Zakoian (2001) give an example where OLS is not efficient. If y_{t-1} is replaced by an exogenous variable x_t and x_t is iid normal $(0, \sigma^2)$, then condition (5.2) is satisfied in spite of the presence of heteroskedasticity because $\omega_1 = -1/(a + 3b\sigma^2)$ and $\omega_i = 0$ for $i > 1$ ³.

A. Proofs

Proof of Proposition 3.1. 1 - Proof of (a) and (b). As $\{k(\cdot, \tau) : \tau \in I\}$ span $\mathcal{H}(K)$ and are linearly independent and $\{h(\tau) : \tau \in I\}$ are complete in $L^2(h)$, we have that any element of $\mathcal{H}(K)$ can be written as

$$g = \sum_j \alpha_j k(\cdot, \tau_j)$$

and any element of $L^2(h)$ can be written as

$$G = \sum_j \beta_j h(\tau_j).$$

The fact that there are at most a countable infinity of terms in the two sums above follows from Theorem 10 of Dunford and Schwartz (1988, p. 251). Note moreover that by the reproducing property, we have

$$(k(\cdot, \tau), k(\cdot, \lambda))_K = k(\tau, \lambda) = (h(\tau), h(\lambda))_{H^0}. \quad (\text{A.1})$$

Let $g = \sum_j \alpha_j k(\cdot, \tau_j)$ be an element of $\mathcal{H}(K)$, it follows from (A.1) that

$$\begin{aligned} \|g\|_K^2 &= \left(\sum_l \alpha_l k(\cdot, \tau_l), \sum_j \alpha_j k(\cdot, \tau_j) \right)_K \\ &= \sum_{l,j} \alpha_l \bar{\alpha}_j k(\tau_l, \tau_j) \\ &= \sum_{l,j} \alpha_l \bar{\alpha}_j (h(\tau_l), h(\tau_j))_{H^0} \\ &= \|G\|_{H^0}^2 \end{aligned}$$

³This remark is inspired from Anatolyev (2003).

for $G = \sum_l \alpha_l h(\tau_l)$. G is unique by Assumption 3.1 and belongs to $C(g)$. Hence J is the one-to-one mapping that transforms $\sum_j \alpha_j k(\cdot, \tau_j)$ into $\sum_j \alpha_j h(\tau_j)$.

Proof of (c) - Now we prove (b) \implies (c). First note that an element of $C(g)$ can not be orthogonal to $h(\tau)$ for all τ . Define a general element of $C(g)$ $G = G_0 + G_1$ where $G_0 = \sum_l \omega_l h(\tau_l)$ belongs to $C \cap L^2(h)$. $G \in C(g)$ implies

$$(G_1, h(\tau))_{H^0} = 0 \quad \forall \tau. \quad (\text{A.2})$$

Next we compute the norm of G and show that it is minimal for $G_1 = 0$.

$$\|G\|_{H^0}^2 = \|G_0\|_{H^0}^2 + \|G_1\|_{H^0}^2 + 2(G_0, G_1)_{H^0}.$$

The condition (A.2) implies that

$$\begin{aligned} \sum_l \omega_l (G_1, h(\tau_l))_{H^0} &= 0 \\ (G_0, G_1)_{H^0} &= 0. \end{aligned}$$

Hence G_0 has minimal norm. The result follows.

Proof of Proposition 4.1. The GMM estimator is efficient if the inverse of its asymptotic variance equals the Information matrix that is the (i, j) element of Σ satisfies:

$$\left(E^\theta (\nabla_{\theta_i} h(\theta)), E^\theta (\nabla_{\theta_j} h(\theta)) \right)_K = E^\theta \left[\left(\frac{\partial \ln f_\theta}{\partial \theta_i} (X_t | X_{t-1}^{L-1}) \right) \left(\frac{\partial \ln f_\theta}{\partial \theta_j} (X_t | X_{t-1}^{L-1}) \right)' \right]$$

at $\theta = \theta_0$. Note that $\|E^{\theta_0} (\nabla_{\theta_i} h)\|_K^2 = \|-E^{\theta_0} (\overline{\nabla_{\theta_i} h})\|_K^2$. Using the notation of Section 3, $(\cdot, \cdot)_{H^0}$ denotes the inner product defined by the covariance operator K . Define

$$C(g) = \{G \mid g(\tau) = (G, h(\tau))_{H^0} \text{ for all } \tau \in I\}.$$

By Proposition 3.1 (i), the GMM estimator will be efficient if the following two conditions are satisfied:

- (a) $G_0 = \nabla_{\theta_j} \ln f_\theta (x_t | x_{t-1}^{L-1})$ belongs to the set $C(g)$ with $g = -E^{\theta_0} (\overline{\nabla_{\theta_j} h})$.
- (b) G_0 belongs to $L^2(h)$.

We successively consider the case where $\{h(\tau, X_t^L)\}$ is a martingale difference sequence and the general case where $\{h(\tau, X_t^L)\}$ is autocorrelated.

1. $\{h(\tau, X_t^L)\}$ is a m.d.s. As $\{h(\tau, X_t^L)\}$ are uncorrelated, the covariance operator K has for kernel

$$k(\tau_1, \tau_2) = E^{\theta_0} \left[h(\tau_1, X_t^L) \overline{h(\tau_2, X_t^L)} \right] \equiv (h, h)_{H^0}.$$

Using the notation of Proposition 3.1, we define

$$C(g) = \left\{ G | g(\tau) = E^\theta \left(G \overline{h(\tau)} \right), \forall \tau \in I \right\}$$

To show (a), we need to prove that

$$-E^{\theta_0} \left(\overline{\nabla_{\theta_j} h} \right) = \int \nabla_{\theta_j} \ln f_\theta \left(x_t | x_{t-1}^{L-1} \right) \overline{h(\tau, x_t^L)} f_\theta \left(x_t^L \right) dx_t^L.$$

Taking the complex conjugate on each side, this is equivalent to show

$$-E^{\theta_0} \left(\nabla_{\theta_j} h \right) = \int \nabla_{\theta_j} \ln f_\theta \left(x_t | x_{t-1}^{L-1} \right) h(\tau, x_t^L) f_\theta \left(x_t^L \right) dx_t^L. \quad (\text{A.3})$$

Note that $E^\theta \left(h(\tau, X_t^L, \theta) \right) = 0$ for all θ . Differentiating with respect to θ on each sides yields

$$\begin{aligned} -E^\theta \left(\nabla_{\theta_j} h(\tau, X_t^L, \theta) \right) &= \int \nabla_{\theta_j} \ln f_\theta \left(x_t^L \right) h(\tau, x_t^L) dx_t^L \\ &= \int \nabla_{\theta_j} f_\theta \left(x_t | x_{t-1}^{L-1} \right) f_\theta \left(x_{t-1}^{L-1} \right) h(\tau, x_t^L) dx_t^L \quad (\text{A.4}) \\ &\quad + \int f_\theta \left(x_t | x_{t-1}^{L-1} \right) \nabla_{\theta_j} f_\theta \left(x_{t-1}^{L-1} \right) h(\tau, x_t^L) dx_t^L \quad (\text{A.5}) \\ &= \int \nabla_{\theta_j} \ln f_{\theta_j} \left(x_t | x_{t-1}^{L-1} \right) f_\theta \left(x_t^L \right) h(\tau, x_t^L) dx_t^L + 0 \end{aligned}$$

because $\{h(\tau, X_t^L)\}$ is a m.d.s. and

$$\int \nabla_{\theta_j} f_\theta \left(x_{t-1}^{L-1} \right) \left\{ \int h(\tau, x_t^L) f_\theta \left(x_t | x_{t-1}^{L-1} \right) dx_t \right\} dx_{t-1}^{L-1} = 0.$$

Hence (A.3) holds.

2. $\{h(\tau, X_t^L)\}$ is autocorrelated. The covariance operator has for kernel

$$\begin{aligned} k(\tau_1, \tau_2) &= \sum_{j=-\infty}^{\infty} E^{\theta_0} h(\tau_1, x_t^L) h(\tau_2, x_{t-j}^L) \\ &= \sum_{j=-\infty}^{\infty} E^{\theta_0} h(\tau_1, S^j Y) h(\tau_2, Y) \\ &\equiv (h, h)_{H^0} \end{aligned}$$

where $Y_t = X_t^L$ and S is a measure preserving shift operator as defined in Stout (1974, page 171). To show that G_0 belongs to $C(g)$ with $g = -E^{\theta_0}(\overline{\nabla_{\theta_j} h})$, we need to establish⁴

$$\begin{aligned} -E^{\theta}(\nabla_{\theta_j} h(\tau, x_t^L, \theta)) &= \sum_{j=-\infty}^{\infty} E^{\theta} \left[\nabla_{\theta_j} \ln f_{\theta}(X_t | X_{t-1}^{L-1}) h(\tau, X_{t-j}^L) \right] \\ &= \sum_{t \in \mathbb{Z}} E^{\theta} \left[\nabla_{\theta_j} \ln f_{\theta}(X_{L-t} | X_{L-t-1}^{L-1}) h(\tau, X_L^L) \right] \\ &\equiv \sum_{t \in \mathbb{Z}} \gamma(t). \end{aligned}$$

As before, we use the fact that $E^{\theta}(h(\tau, x_t^L, \theta)) = 0$ for all θ . Again differentiating with respect to θ gives $-E^{\theta}(\nabla_{\theta_j} h(\tau, x_t^L, \theta)) = (\text{A.4}) + (\text{A.5})$. As $\{h(\tau, X_t^L)\}$ is no longer a m.d.s., the term (A.5) does not vanish. We now turn our attention on the terms $\gamma(t)$. For all $t < 0$, $\gamma(t) = 0$.

$$\begin{aligned} \gamma(0) &= \int \nabla_{\theta_j} \ln f_{\theta}(x_p | x_{L-1}^{L-1}) h(\tau, x_L^L) f_{\theta}(x_L^L) dx_L^L \\ &= (\text{A.4}). \end{aligned}$$

To conclude the proof, it remains to show that

$$\sum_{t=1}^{\infty} \gamma(t) = (\text{A.5}).$$

We have

$$\sum_{t=1}^{\infty} \gamma(t) = \lim_{T \rightarrow \infty} E^{\theta_0} \left[\sum_{t=1}^T \nabla_{\theta_j} \ln f_{\theta}(X_{L-t} | X_{L-t-1}^{L-1}) h(\tau, X_L^L) \right]$$

and

$$\begin{aligned} &E^{\theta_0} \left[\sum_{t=1}^T \nabla_{\theta_j} \ln f_{\theta}(X_{L-t} | X_{L-t-1}^{L-1}) h(\tau, X_L^L) \right] \tag{A.6} \\ &= E^{\theta_0} \left[\nabla_{\theta_j} \ln f_{\theta}(X_{L-1}, \dots, X_{L-T} | X_{L-T-1}^{L-1}) h(\tau, X_L^L) \right] \\ &= \int \nabla_{\theta_j} \ln f_{\theta}(x_{L-1}^{T-1} | x_{L-T-1}^{L-1}) h(\tau, x_L^L) f_{\theta}(x_L^{T+L}) dx_L^{T+L} \\ &= \int \nabla_{\theta_j} f_{\theta}(x_{L-1}^{T-1} | x_{L-T-1}^{L-1}) h(\tau, x_L^L) f_{\theta}(x_L | x_{L-1}^{L-1}) f_{\theta}(x_{L-T-1}^{L-1}) dx_L^{T+L}. \end{aligned}$$

⁴This equality appears in Newey and McFadden (1994) in the independent case. It has been proved for the dynamic case by Hall and Inoue (2003) independently to us.

Using

$$\begin{aligned}
& \int \nabla_{\theta_j} f_{\theta} \left(x_{L-1}^{T-1} | x_{L-T-1}^{L-1} \right) dx_{-1} dx_{-2} \dots dx_{L-T} \\
&= \nabla_{\theta_j} \int f_{\theta} \left(x_{L-1}^{T-1} | x_{L-T-1}^{L-1} \right) dx_{-1} dx_{-2} \dots dx_{L-T} \\
&= \nabla_{\theta_j} f_{\theta} \left(x_{L-1}^{L-1} | x_{L-T-1}^{L-1} \right),
\end{aligned}$$

we obtain

$$(A.6) = \int \nabla_{\theta_j} f_{\theta} \left(x_{L-1}^{L-1} | x_{L-T-1}^{L-1} \right) h \left(\tau, x_L^L \right) f_{\theta} \left(x_L | x_{L-1}^{L-1} \right) f_{\theta} \left(x_{L-T-1}^{L-1} \right) dx_L^L dx_{L-T-1}^{L-1}.$$

Hence letting T go to infinity and applying Assumption 4.1 (ii) yield

$$\begin{aligned}
\sum_{t=1}^{\infty} \gamma(t) &= \int \nabla_{\theta} f_{\theta} \left(x_{L-1}^{L-1} \right) h \left(\tau, x_L^L \right) f_{\theta} \left(x_L | x_{L-1}^{L-1} \right) f_{\theta} \left(x_{L-T-1}^{L-1} \right) dx_L^L dx_{L-T-1}^{L-1} \\
&= (A.5).
\end{aligned}$$

References

- [1] Anatolyev, S. (2003) “Redundancy of Lagged Regressors in a Conditionally Heteroskedastic Time Series Regression”, Problem 03.1.2, *Econometric Theory*, 225.
- [2] Broze, L., C. Francq, and J-M. Zakoian (2001) “Non-redundancy of high order moment conditions for efficient GMM estimation of weak AR processes”, *Economics Letters*, 317-322.
- [3] Carrasco, M. and J. P. Florens (2000) “Generalization of GMM to a continuum of moment conditions”, *Econometric Theory*, 16, 797-834.
- [4] Carrasco, M. and J. P. Florens (2001) “Efficient GMM Estimation Using the Empirical Characteristic Function”, mimeo, University of Rochester.
- [5] Carrasco, M. and J. P. Florens (2002) “Simulation Based Method of Moments and Efficiency”, *Journal of Business & Economic Statistics*, Vol. 20, No. 4, 482-492.
- [6] Carrasco, M., M. Chernov, J. P. Florens, and E. Ghysels (2002) “Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions”, working paper, University of Rochester.
- [7] Chamberlain, G. (1987) “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions”, *Journal of Econometrics*, 34, 305-334.
- [8] Dunford, N. and J. Schwartz (1988) *Linear Operators, Part I: General Theory*, Wiley, New York.
- [9] Feuerverger, A. (1990) “An efficiency result for the empirical characteristic function in stationary time-series models”, *Canadian Journal of Statistics*, 18, 155-161.
- [10] Gallant, A. R. and J. R. Long (1997) “Estimating Stochastic Differential Equations Efficiently by Minimum Chi-squared”, *Biometrika*, 84, 125-141.
- [11] Gallant, R. and D. Nychka (1987) “Semi-Nonparametric Maximum Likelihood Estimation”, *Econometrica*, 55, 363-390.

- [12] Gallant, A. R., and G. Tauchen (1996) “Which Moments to Match?”, *Econometric Theory*, 12, 657-681.
- [13] Gourieroux, C., A. Monfort (1996) *Simulation Based Econometric Methods*, CORE Lectures, Oxford University Press, New York.
- [14] Gourieroux, C., A. Monfort, and E. Renault (1993) “Indirect Inference”, *Journal of Applied Econometrics*, 8, S85-S118.
- [15] Hall, A. and A. Inoue (2003) “Information in Empirical Likelihood and Generalized Method of Moments Estimation”, mimeo, North Carolina State University.
- [16] Hansen, L. (1982) “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica*, 50, 1029-1054.
- [17] Hansen, L. (1985) “A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators”, *Journal of Econometrics*, 30, 203-238.
- [18] Hansen, L., J. Heaton, and M. Ogaki (1988) “Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions”, *Journal of the American Statistical Association*, Vol. 83, 863-871.
- [19] Kim, Y., H. Qian, and P. Schmidt (1999) “Efficient GMM and MD estimation of autoregressive models”, *Economics Letters*, 265-270.
- [20] Newey, W. and D. McFadden (1994) “Large Sample Estimation and Hypothesis Testing”, in *Handbook of Econometrics*, Vol. IV, eds R.F. Engle and D. McFadden, Elsevier.
- [21] Parzen, E. (1970) “Statistical Inference on time series by RKHS methods”, *12th Biennial Seminar Canadian Mathematical Congress Proc.*, R. Pyke, ed., Canadian Mathematical Society, Montreal.
- [22] Rosenblatt, M. (1971) *Markov Processes, Structure and Asymptotic Behavior*, Springer Verlag, Berlin.
- [23] Saitoh, S. (1997) *Integral transforms, reproducing kernels and their applications*, Longman.

- [24] Sansone, G. *Orthogonal Functions*, Dover Publications, New York.
- [25] Singleton, K. (2001) “Estimation of Affine Pricing Models Using the Empirical Characteristic Function”, *Journal of Econometrics*, 102, 111-141.
- [26] Tauchen, G. (1997) “New Minimum Chi-Square Methods in Empirical Finance”, in *Advances in Econometrics, Seventh World Congress*, eds. D. Kreps and K. Wallis, Cambridge University Press, Cambridge.
- [27] Stout, W. (1974) *Almost Sure Convergence*, Academic Press, New York.
- [28] West, K. (2001) “On Optimal Instrumental Variables Estimation of Stationary Time Series Models”, *International Economic Review*, Vol. 42, 1043-1050.