JETEM

Journal of Economic Theory and Econometrics

The Journal of the Korean Econometric Society

March 2008

Volume 19. Issue 1. pp.1-16

# The Generalized Method of Moments in the Presence of Nonstationary Variables

**Jae-Young Kim***

**Abstract** This paper studies the generalized method of moments (GMM) in the presence of nonstationary time series with a unit root. We investigate asymptotic properties of the GMM estimator in such a situation. It is shown that the GMM estimator is a consistent estimator regardless of whether or not the model under study contains nonstationarity. On the other hand, the asymptotic distribution of the GMM estimator is nonstandard in the presence of nonstationarity. Such nonstandard limiting behavior of the GMM estimator in the presence of unit roots causes difficulty in statistical inference. However, under a reasonable condition implied in an equilibrium relation the asymptotic distribution of the GMM estimator is described by a mixed normal distribution. The mixed normality of the estimator itself is, in general, not of direct use for statistical inference. However, the mixed normality of the estimator sometimes enables us to derive results that are useful for statistical inference. We show that, under some reasonable conditions, the asymptotic behavior of the J-statistic for testing the validity of moment restrictions is characterized by a chi-square distribution.

**Keywords** GMM, nonstationarity, unit root, asymptotic properties, J-statistic

**JEL Classification** C11, C14, C2, C3, C5

* The author thanks Myoung-Jae Lee and two anonymous referees for helpful comments on an earlier version of the paper. Correspondence: School of Economics, Seoul National University, Seoul 151-732, Korea. +82-2-880-6390, e-mail: jykim017@snu.ac.kr

# 1 Introduction

The generalized method of moments (GMM) estimator defined and analyzed in Hansen (1982) has been applied for a wide variety of economic problems that imply a set of economic theoretical restrictions. Hansen (1982) and subsequent works in GMM literature assume sufficient stationarity in the variables of the model. Little is known, however, about the GMM estimator when there exist some nonstationary variables in the model. Kitamura and Phillips (1997) study GMM estimation for a linear regression model with some nonstationary variables. This is an important deficiency in the literature since many economic time series are regarded as nonstationary processes characterized by the unit root hypothesis, as documented by many researchers since Nelson and Plosser (1982). This paper studies the GMM when the model contains some nonstationary time series with a unit root.

In the presence of a unit root, the behavior of the GMM estimator is not standard, in general, as in the case of most of the other classical estimators. In this paper we study large sample properties of the GMM estimator, when some variables in the model contain a unit root. We will show that the GMM estimator is a consistent estimator regardless of whether or not the model under study contains some nonstationary variables with a unit root. - On the other hand, the asymptotic distribution of the GMM estimator in such a situation is usually represented by a function of nonstandard stochastic processes. Such nonstandard limiting behavior of the GMM estimator in the presence of unit roots causes difficulty in statistical inference. However, under a reasonable condition the asymptotic distribution of the GMM estimator is described by a mixed normal distribution. Although the mixed normal distribution is not a standard distribution, we can construct a chi-square test statistic for testing the validity of theoretical restrictions implied in an econometric model.

The discussion of this paper goes as follows. Section 2 studies consistency of the GMM estimator in the presence of nonstationary variables. In section 3, we investigate the asymptotic distribution of the GMM estimator. Section 4 develops a procedure for testing the validity of moment restrictions of the model by applying the result of section 3. Section 5 concludes the paper.

## 2 Consistency of the GMM Estimator

Let $x_t$ be a $p \times 1$ vector of stochastic processes defined on a probability space $(\Omega, \mathcal{F}, P)$. A finite segment of a particular realization of the process, i.e., $\{x_t(\omega_0) : 1 \leq t \leq T\}$ for some $\omega_0 \in \Omega$ is the observed data.

The GMM procedure is based on the moment condition

$$E[f(x_t, \beta_0)] = 0, \tag{2.1}$$

where the function $f(\cdot, \cdot)$ represents an $r \times 1$ vector of moment restrictions, and $\beta$ is a $q \times 1$ vector of parameters. In economics, a set of theoretical restrictions are usually given by functional relations among variables:

$$0 = F(x_{1t}, \beta_0)$$

where $x_{1t}$ is a $p_1 \times 1$ vector of variables. An econometric model for these theoretical relations is given by

$$u_t = F(x_{1t}, \beta_0) \tag{2.2}$$

where $u_t$ is a disturbance. To estimate the model (2.2) we sometimes need a set of instrumental variables $z_t$ that is obtained from $x_{2t}$, a $p_2 \times 1$ vector of variables:

$$z_t = G(x_{2t}, \beta_0), \tag{2.3}$$

which, with the disturbance $u_t$, satisfies a set of orthogonality conditions

$$E[u_t \otimes z_t] = 0. \tag{2.4}$$

Notice that $u_t$ may be a vector, and each element of $z_t$ in this case is applied to each element of $u_t$. Now, writing

$$f(x_t, \beta_0) = F(x_{1t}, \beta_0) \otimes G(x_{2t}, \beta_0), \tag{2.5}$$

where $x_t = (x'_{1t}, x'_{2t})'$, we have the moment condition (2.1).

One of the key assumptions in the GMM literature is that the variable $x_t$ is stationary and ergodic. In recent years, however, many economic and financial time series are found to be nonstationary processes, especially $I(1)$ processes as defined by Engle and Granger (1987). Little is known about the GMM estimator when there exist some nonstationary variables in the model. In this paper we investigate large sample properties of the GMM estimator and discuss some inference issue when some of variables in $x_t$ are nonstationary.

Thus, let $x_t = (x_{1t}, \cdots, x_{pt})'$ be such that

$$x_{it} = \rho_i x_{it-1} + \epsilon_{it}.$$

where $\rho_i \leq 1$ is a constant. If $\rho_i = 1$ then $x_{it}$ is an I(1) process. While allowing that some of variables in $x_t$ are $I(1)$, we assume that the following is satisfied for the disturbance $u_t$, instrument $z_t$, and the driving force $\epsilon_{it}$:

**Assumption 2.1** *Let $w_t = u_t \otimes z_t$. Assume that $(w_t, \epsilon_t)$ is jointly stationary and ergodic.*

A sufficient condition for the Assumption 2.1 is

**Assumption 2.1$'$** *The process $(u_t, z_t, \epsilon_t)$ is jointly stationary and ergodic.*

Usually the equation (2.2) is specified as an approximation to an equilibrium relation among components of $x_t$. If the equilibrium concept is to have any relevance for the specification of econometric models, it should be the case that the *equilibrium error $u_t$* usually has a small value and does not persist for a long time. Therefore, the assumption of $u_t$ being stationary is reasonable. On the other hand, the assumption of the instrumental variable $z_t$ being stationary seems difficult to make when some components of $x_t$ are $I(1)$. One candidate for the stationary instrument $z_t$ is obtained from

$$z_t = \alpha' x_{t-1} \tag{2.6}$$

where $\alpha$ is a cointegration vector. However, we could obtain an instrument from a nonlinear relation $G$ satisfying (2.3) and (2.4) that are better than $\alpha' x_{t-1}$ in (2.6).

**Assumption 2.2** *The parameter space $S$ is a convex subset of $\mathbb{R}^q$ that contains $\beta_0$ in its interior, and the metric space $(S, \sigma)$ is separable, where $\sigma$ denotes a metric on $S$.*

**Assumption 2.3** *$f(\cdot, \beta)$ is Borel measurable for each $\beta \in S$, and $f(x, \cdot)$ is continuous on $S$ for each $x \in \mathbb{R}^q$.*

The GMM estimation combines cross equation information by a sequence of random weighting matrices $\{a_T, T \geq 1\}$ that is of $(s \times r)$ where $q \leq s \leq r$.

**Assumption 2.4** *The sequence of random weighting matrices $\{a_T, T \geq 1\}$ converge to a constant matrix $a_0$ in probability.*

We will denote

$$f_t(\omega, \beta) = f[x_t(\omega), \beta)]$$

and introduce some notations:

$$g_T(\omega, \beta) = \frac{1}{T} \sum_{t=1}^{T} f_t(\omega, \beta),$$

$$h_T(\omega, \beta) = a_T(\omega) g_T(\omega, \beta).$$

**Definition 2.1** *The GMM estimator $\{b_T(\omega) : T \geq 1\}$, for $\omega \in \Omega$ and for a suitably chosen $a_T(\omega)$, is a sequence of random vectors such that*

$$b_T(\omega) = \arg\min_{\beta} h_T(\omega, \beta)' h_T(\omega, \beta).$$

Under Assumptions 2.2-2.3 the GMM estimator is well defined for any given $a_T$. Hansen (1982) finds an optimality condition for the choice of $a_T$. In the case of the variable $x_t$ being stationary the optimal choice of $a_T$ is such that $a_T' a_T = \hat{S}^{-1}$ where $\hat{S}$ is a consistent estimator of the long-run variance-covariance of the series $\{w_t\}$.

To get consistency of the GMM estimator, we introduce the following condition.

**Assumption 2.5** *Let $U(\beta_0, \eta)$ be an open neighborhood of $\beta_0$ (an open ball in $S$ centered at $\beta_0$) with radius $\eta$. For any $\eta > 0$ there exists $k(\eta) > 0$ such that*

$$\limsup_{T \nearrow \infty} P[\sup_{S \setminus U(\beta_0, \eta)} [h_T(\beta_0)' h_T(\beta_0) - h_T(\beta)' h_T(\beta)] \leq -k(\eta)] = 1,$$

*where* $\lim P[h_T(\beta_0)' h_T(\beta_0) < \infty] = 1$.

The above assumption is essentially an identification condition for $\beta_0$. Usually this condition is more easily satisfied when $x_t$ contains some $I(1)$ processes than the case of pure stationarity. This is because $h_T(\beta_0) \to 0$, while $h_T(\beta) \nearrow \infty$ for $\beta \neq \beta_0$ as $T \nearrow \infty$, which is shown by the following: By a law of large numbers, it can be shown that

$$h_T(\beta_0) \equiv a_T \frac{1}{T} \sum_{t=1}^{T} w_t \xrightarrow{P} a_0 E[f(\beta_0)] = 0$$

for $\{w_t\}$ satisfying Assumption 1, where "P" denotes "in P-probability." To see that $h_T(\beta) \nearrow \infty$ for $\beta \neq \beta_0$ as $T \nearrow \infty$, consider a linear model, without loss of generality:

$$x_{1t} = \beta_{10} + \beta_{20} x_{2t} + \beta_{30} x_{3t} + w_t, \tag{2.7}$$

where $x_{2t}$ and $x_{3t}$ are scalr processes such that

$$\begin{aligned} x_{2t} &= \rho x_{2t-1} + e_{3t}, \quad |\rho| < 1, \\ x_{3t} &= x_{3t-1} + e_{3t} \end{aligned}$$

where $e_{3t}$ is an iid random variable. Thus, $x_{2t}$ is a stationary process, and $x_{3t}$ is an $I(1)$ process. Suppose that (2.7) can be written by

$$f(x_t, \beta_0) = w_t,$$

where

$$f(x_t, \beta) = x_{1t} - \beta_1 - \beta_2 x_{2t} - \beta_3 x_{3t}.$$

For $\beta \neq \beta_0$,

$$\begin{aligned} T^{-1} \sum_{t=1}^{T} f(x_t, \beta) &= T^{-1} \sum_{t=1}^{T} f(x_t, \beta_0) + \{T^{-1} \sum_{t=1}^{T} f(x_t, \beta) - T^{-1} \sum_{t=1}^{T} f(x_t, \beta_0)\} \\ &= T^{-1} \sum_{t=1}^{T} f(x_t, \beta_0) + (\beta_1 - \beta_{10}) + (\beta_2 - \beta_{20}) T^{-1} \sum_{t=1}^{T} x_{2t} + (\beta_3 - \beta_{30}) T^{-1} \sum_{t=1}^{T} x_{3t}. \end{aligned} \tag{2.8}$$

But, by a functional central limit theorem,

$$T^{-3/2} \sum_{t=1}^{T} x_{3t} \Longrightarrow \int_0^1 W(r) dr, \tag{2.9}$$

where $W(\cdot)$ is a standard Wiener process, and '$\Longrightarrow$' denotes weak convergence of one process to another, so that

$$P\left[\ |T^{-1}\sum_{t=1}^{T} x_{3t}| > M\ \right] = P\left[\ |T^{1/2}T^{-3/2}\sum_{t=1}^{T} x_{3t}| > M\ \right] \longrightarrow 1$$

as $T \nearrow \infty$, where $M$ is a big real number. Hence from (2.8) $T^{-1}\sum_{t=1}^{T} f(x_t, \beta)$ diverges as $T \nearrow \infty$ for $\beta = (\beta_{10}, \beta_{20}, \beta_3)$ with $\beta_3 \neq \beta_{30}$. Notice that from (2.9)

$$\lim_{T \nearrow \infty} P[|T^{-1}\sum_{t=1}^{T} x_{3t}| = 0] = 0, \tag{2.10}$$

in other words, $|T^{-1}\sum_{t=1}^{T} x_{3t}| > 0$ in probability.

(2.9) and (2.10) implies that

$$(\beta_3 - \beta_{30})T^{-1}\sum_{t=1}^{T} x_{3t} = O_p(T^{1/2}) \tag{2.11}$$

for $\beta_3 \neq \beta_{30}$. By a similar argument, we have for $i = 1, 2$

$$(\beta_i - \beta_{i0})T^{-1}\sum_{t=1}^{T} x_{it} = O_p(1) \tag{2.12}$$

for $\beta_i \neq \beta_{i0}$. It follows, then, that for $\beta \in S \setminus U(\beta_0, \eta)$

$$h_T(\beta_0)'h_T(\beta_0) - h_T(\beta)'h_T(\beta) = O_p(T) \tag{2.13}$$

as the dominant term of the left hand side of (2.13) is of $O_p(T)$ from (2.11) and (2.12). Thus, we have the condition in Assumption 2.5 satisfied for the model (2.7).

Assumptions 2.1-2.5 are sufficient for the existence and consistency of the GMM estimator.

**Theorem 2.1** *Under Assumptions 2.1-2.5, a GMM estimator $b_T$ exists and converges to $\beta_0$ in probability.*

**Remark 1:** (1) When all components of $x_t$ are $I(1)$ variables, the convergence of the GMM estimator is usually achieved only in "in probability" sense as the limiting behavior of some parts of the model is characterized in the sense of "weak convergence"

as in (2.9).
(2) The assumption 2.1 and 2.5 are introduced for a model with some $I(1)$ processes. On the other hand, the other assumptions (2.2-2.4) are not particular for a model with $I(1)$ variables.
(3) A key condition for the consistency of the GMM estimator is the identification condition (Assumption 2.5). As is explained above Assumption 2.5 is more easily satisfied in the case of $I(1)$ variables than in the case of pure stationarity. Thus, consistency of the estimator is in a sense more easily established in the former case, following the approach of Wooldridge and White (1985).

# 3 The Asymptotic Distribution of the GMM Estimator

In this section we investigate the asymptotic distribution of the GMM estimator when some of variables in the model (2.1) are $I(1)$ processes. As in the previous section, we begin our discussion by describing assumptions needed to obtain the result.

**Assumption 3.1** *$\partial f(\cdot, \beta)/\partial\beta$ is Borel measurable for each $\beta \in S$, and $\partial f(x, \cdot)/\partial\beta$ is continuous on $S$ for each $x \in \mathbb{R}^q$.*

We need some conditions for the first "moment" of $\partial f(x_t, \beta)/\partial\beta$. When $x_t$ contains some $I(1)$ variables, the usual arithmetic sample mean $T^{-1}\sum_{t=1}^{T}\partial f(x_t, \beta)/\partial\beta$ does not have a finite limit since each component of $\sum_{t=1}^{T}\partial f(x_t, \beta)/\partial\beta$ does *not* have a homogeneous order of magnitude. To see this, consider the example (2.7). By a law of large numbers and by a functional central limit theorem, respectively, we have

$$T^{-1}\sum_{t=1}^{T}\frac{\partial f(x_t, \beta)}{\partial\beta_2} = T^{-1}\sum_{t=1}^{T}x_{2t} \rightarrow E[x_{2t}], \tag{3.1}$$

$$T^{-3/2}\sum_{t=1}^{T}\frac{\partial f(x_t, \beta)}{\partial\beta_3} = T^{-3/2}\sum_{t=1}^{T}x_{3t} \Longrightarrow \int_0^1 W(r)dr, \tag{3.2}$$

which implies that

$$\sum_{t=1}^{T}\frac{\partial f(x_t, \beta)}{\partial\beta_2} = O_p(T),$$

but

$$\sum_{n=1}^{T} \frac{\partial f(x_t, \beta)}{\partial \beta_3} = O_p(T^{3/2}).$$

To avoid the above difficulty in defining a sample moment of $\partial f(x_t, \beta)/\partial \beta$, we introduce a diagonal matrix $\Upsilon_T$ that properly normalizes $\sum_{t=1}^{T} \partial f(x_t, \beta_0)/\partial \beta$: Thus, let

$$\Upsilon_T = diag(O_{p,1}, ..., O_{p,q})$$

where $O_{p,i} = \max_j\{O_{p,ij}\}$ where $O_{p,ij}$ is the order of magnitude of the $ij^{th}$ element of $\{\sum_{t=1}^{T} \partial f(x_t, \beta_0)/\partial \beta\}$ for $i = 1, ..., r$ and $j = 1, ..., q$. In the following analysis it becomes apparent that $O_{p,i}$ is the order of magnitude of the $i^{th}$ component of $b_T$. Thus, $\Upsilon_T$ is a diagonal matrix whose $i^{th}$ diagonal element is related to the order of magnitude of the $i^{th}$ element of a random process for which $\Upsilon_T$ is applied. For in Theorem 3.1 the order of magnitude of $b_{iT}$ is the $i^{th}$ diagonal element of $\Upsilon_T$ divided by $\sqrt{T}$. Using $\Upsilon_T$, we can specify the following condition:

**Assumption 3.2** *Let $d_T(\beta) = \{a_T \sum_{t=1}^{T} \partial f(x_t, \beta)/\partial \beta\}\Upsilon_T^{-1}$. Assume that $d_T(\beta_0)$ converges weakly to $d_0$, where $d_0$ is a stochastic process such that $P[|d_{0,ij}| < \infty] = 1$ for all $i = 1, ..., s$, $j = 1, ..., q$, and $P[\det(d_0' d_0) \neq 0] = 1$.*

If all the components of $x_t$ are stationary and ergodic, the normalizing matrix $\Upsilon_T$ reduces to a degenerate matrix, $\Upsilon_T = TI_q$, in which case

$$d_T(\beta_0) = a_T \sum_{t=1}^{T} \frac{\partial f(x_t, \beta_0)}{\partial \beta} \Upsilon_T^{-1} = a_T T^{-1} \sum_{t=1}^{T} \frac{\partial f(x_t, \beta_0)}{\partial \beta} \longrightarrow a_0 E[\frac{\partial f(x_t, \beta_0)}{\partial \beta}].$$

We may establish convergence of $d_T(\beta)$ in Assumption 3.2 under some conditions on $\epsilon_t$. For example, we can assume that the process $\epsilon_t$ is such that (a) $E[\epsilon_{it}] = 0$ for all $t$; (b) $\sup_t E|\epsilon_{it}|^{\beta} < \infty$ for some $\beta > 2$; (c) A long-run variance of $\epsilon_t$ exists; (d) $\{\epsilon_t\}_1^{\infty}$ is a strong mixing with mixing coefficients $\alpha_m$ that satisfy $\sum_1^{\infty} \alpha_m^{1-2/\beta} < \infty$.

We need a continuity condition for $d_T(\beta)$ with respect to $\beta$:

**Assumption 3.3** *Let $\lambda_T(\omega, \beta, \delta) = sup\{\|d_T(\omega, \beta) - d_T(\omega, \alpha)\| : \alpha \in S, \sigma(\beta, \alpha) < \delta\}$ for*

*$T \geq 1$, $\omega \in \Omega$, $\delta > 0$, and $\|\cdot\|$ denotes the norm of a matrix.*[2] *Assume that*

$$\lim_{T\nearrow\infty}\lim_{\delta\downarrow 0} P[\lambda_T(\omega,\beta_0,\delta) < \varepsilon] = 1$$

*for each $\varepsilon > 0$.*

From the above conditions we have the following result:

**Lemma 3.1** *Suppose that Assumptions 2.1-2.5 and 3.1-3.3 are satisfied. Then*

$$d_T(b_T) \equiv a_T T \frac{\partial g_T(b_T)}{\partial \beta} \Upsilon_T^{-1} \Longrightarrow d_0.$$

Recall that

$$w_t = f(x_t, \beta_0) \text{ for } t = 1, 2, ...$$

Let

$$v_j = E[w_{t+j}|w_t, w_{t-1}, ...] - E[w_{t+j}|w_{t-1}, w_{t-2}, ...] \text{ for } j \geq 0.$$

Notice that the basic properties of $v_j$ does not depend on the index $t$ since $\{w_t\}$ is a stationary process. Also, notice that $\{v_j\}$ is a martingale difference sequence. The processes $w_t$ and $v_j$ satisfy the following conditions:

**Assumption 3.4** *$E[w_t w_t']$ exists and is finite, $E[w_{t+j}|w_t, w_{t-1}, ...]$ converges in mean square to zero, and $\sum_{j=0}^{\infty} E[v_j' v_j]^{1/2}$ is finite.*

The Assumption 3.4 provides sufficient conditions for applying a central limit theorem for the stationary and ergodic process $w_t$ by Gordin (1969). Now denote

$$R_w(j) = E[w_{j+1} w_1']$$

and

$$S_w = \sum_{j=1}^{\infty} R_w(j).$$

The $(r \times r)$ matrix $S_w$ is the long-run variance-covariance matrix of the process $\{w_t\}$. The following result is a central limit theorem due to Gordin (1969):

[2] The norm of an $n \times n$ matrix $A$ is defined by $\|A\| = \sup\{|Ax| : x \in \mathbb{R}^n, |x| \leq 1\}$ where $|Ax|$ is the usual Euclidean norm on $\mathbb{R}^n$.

**Lemma 3.2** *Under Assumptions 2.1 and 3.4,*

$$T^{-1/2}\sum_{t=1}^{T} w_t \Longrightarrow Z_w,$$

*where* $Z_w \sim N(0, S_w)$.

We can also establish joint convergence of $(d_T(b_T), T^{-1/2}\sum_1^T w_t)$:

$$(d_T(b_T), T^{-1/2}\sum_1^T w_t) \Longrightarrow (d_0, Z_w)$$

under the same conditions for Lemmas 3.1 and 3.2. Now we obtain the following result:

**Theorem 3.1** *Suppose that Assumptions 2.1-2.5 and 3.1-3.4 are satisfied. Let* $D_T = \Upsilon_T/\sqrt{T}$. *Then*

$$D_T(b_T - \beta_0) \Longrightarrow \mathcal{B}_2^{-1}\mathcal{B}_1$$

*where* $\mathcal{B}_1 = d_0' a_0 Z_w$ *and* $\mathcal{B}_2 = d_0' d_0$.

**Remark 2:** (1) The limiting random process $\mathcal{B}_2^{-1}\mathcal{B}_1 = (d_0' d_0)^{-1} d_0' a_0 Z_w$ in Theorem 3.1 is a mixed normal process. When all the components of $x_t$ are stationary variables, the limiting process becomes a normal random variable.
(2) The assumptions in this section can be easily checked for a linear model. For a nonlinear model the assumptions may not be easy to check depending on the form of the nonlinear specifications. However, once the order of magnitude of each component of the sum $\{\sum_{t=1}^{T} \partial f(x_t, \beta_0)/\partial\beta\}$ is known, we can figure out what the behavior of the GMM estimator is. For a given function $f$ the exact forms of the limiting processes, $\mathcal{B}_1$ and $\mathcal{B}_2$, can be derived by a functional central limit theorem.

# 4 Testing Over-identifying Restrictions

If the number of orthogonality conditions $r$ exceeds the number of parameters to be estimated $q$, tests of the restrictions implied by the econometric model are available.

When all the elements of $x_t$ are stationary, the test procedure is discussed in Hansen (1982). The procedure is viewed as an extension of Sargan (1958)'s test for the validity of instruments and is often called a test for "over-identifying" restrictions. In this section, we discuss a procedure for testing over-identifying restrictions when some elements of $x_t$ are nonstationary processes.

In the last section we found that, if there exist some unit roots in a model, the standard GMM estimator has a mixed normal asymptotic distribution. It is not a standard distribution, and its distribution function is unknown. Fortunately, however, we can construct a test statistic for testing over-identifying restrictions which has a chi-square asymptotic distribution. Our test is based on the following statistic:

$$\Xi = (T)^{-1} \sum_{t=1}^{T} f(x_t, b_T)' \hat{S}_T^{-1} \sum_{t=1}^{T} f(x_t, b_T)$$

where $\hat{S}_T$ is a consistent estimator of $S_w$.

The asymptotic distribution of the test statistic $\Xi$ is given by the following theorem:

**Theorem 4.1** *Under Assumptions 2.1-2.5 and 3.1-3.4,*

$$\Xi \xrightarrow{d} \chi^2(r-q)$$

*where* $\chi^2(r-q)$ *is a chi-square random variable with* $(r-q)$ *degrees of freedom.*

The above theorem extends the test of over-identifying restrictions of Hansen (1982) to the case of nonstationary variables.

# 5 Conclusions

We have considered the GMM when the model contains some nonstationary time series with a unit root. We have shown that the GMM estimator is a consistent estimator regardless of the existence of the unit roots. However, different from the case of stationarity the rate of convergence of the estimator depends on the location of the unit roots. On the other hand, the asymptotic distribution of the GMM estimator is nonnormal and depends on the functional form of the theoretical restrictions implied by an econometric model. However, under a reasonable condition that implies that the economic relations usually have mild and non-persistent fluctuations around the state of equilibrium, the asymptotic distribution of the GMM estimator is described by a mixted normal distribution. We have shown that in this case it is possible to test overidentifying restrictions based on the usual chi-square test in the GMM framework.

## References

Chan, N. H. and C. Z. Wei, 1988, Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes, *Annals of Statistics*, 16, 367-401.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration And Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55, 251-276.

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

Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054.

Herrndorf, N., 1984, A Functional Central Limit Theorem For Weakly Dependent Sequences of Random Variables, *The Annals of Probability*, 12, 141-153.

Jennrich, R., 1969, Asymptotic Properties of Nonlinear Least Square Estimators, *Annals of Mathematical Statistics*, 40, 633-643.

Kitamura, Y., and P.C.B. Phillips, 1997, Fully Modified IV, GIVE and GMM Estimation with Possibly Non-stationary Regressors and Instruments, *Journal of Econometrics*, 80, 85-123.

Nelson, C.R., and C.I. Plosser, 1982, Trends and Random Walks in Macroeconomic Time Series, *Journal of Monetary Economics*, 10, 139-192.

Sargan, J.D., 1958, The Estimation of Economic Relationships Using Instrumental Variables, *Econometrica*. 26. 393-415.

Wooldridge, J. and H. White, 1985, Consistency of Optimization Estimators, Discussion Paper, University of California, San Diego.

# Appendix: Mathematical Proofs

## Proof of Theorem 2.1

Given Assumptions 2.3-2.4, the existence of the GMM estimator is a direct consequence of lemma 2 of Jennrich (1969). Also, given Assumptions 2.2-2.5, the consistency of the GMM estimator is a direct consequence of theorem 2.3 of Wooldridge and White (1985).

## Proof of Lemma 3.1

Under Assumptions 2.1-2.5, $b_T$ is well defined and $b_T \to \beta_0$ as is shown in Theorem 2.1. By adding Assumption 3.1 to these assumptions, $d_T(b_T)$ is well defined. Now, by Assumption 3.3 $\|d_T(b_T) - d_T(\beta_0)\| \to 0$ as $b_T \to \beta_0$. But by Assumption 3.2 $d_T(\beta_0) \Rightarrow d_0$.

## Proof of Lemma 3.2

By Theorem 1 in Hannan (1973), the conclusion follows.

## Proof of Theorem 3.1

Write $Dg_T = \partial g_T / \partial \beta$. Let

$$Dg_T(\beta^1, ..., \beta^r) = \begin{pmatrix} \frac{\partial g_T^1}{\partial \beta}(\beta^1) \\ \vdots \\ \frac{\partial g_T^r}{\partial \beta}(\beta^r) \end{pmatrix}.$$

By Taylor's theorem and Assumptions 2.2 and 3.1, for sufficient large $T$ we can write

$$g_T(b_T) = g_T(\beta_0) + Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)(b_T - \beta_0),$$

where $\bar{b}_T^i$ is on the line segment between $\beta_0$ and $b_T$ for $i = 1, ..., r$. Premultiplying by $a_T^* = d_T(b_T)' a_T$, we get

$$a_T^* g_T(b_T) = a_T^* g_T(\beta_0) + a_T^* Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)(b_T - \beta_0). \tag{1}$$

From the first order condition for maximization, $a_T^* g_T(b_T) = 0$, so that for sufficiently large $T$, the equation (1) can be written as

$$(b_T - \beta_0) = -[a_T^* Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)]^{-1} a_T^* g_T(\beta_0). \tag{2}$$

Rewriting the above equation,

$$\begin{aligned} D_T(b_T - \beta_0) &= -[a_T^* T Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1}]^{-1} a_T^* \sqrt{T} g_T(\beta_0) \\ &= -[d_T' a_T T Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1}]^{-1} d_T' a_T \sqrt{T} g_T(\beta_0) \end{aligned} \tag{3}$$

where $D_T = \Upsilon_T / \sqrt{T}$. Since $b_T \xrightarrow{p} \beta_0$ under Assumptions 2.1-2.5, it follows that $\bar{b}_T^i \xrightarrow{p} \beta_0$ for $i = 1, ..., r$. Thus, Lemma 3.1 implies that $a_T T Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1} \Rightarrow d_0$. Notice that $\sqrt{T} g_T(\beta_0) \xrightarrow{d} Z_w$ by Lemma 3.2. Also, the joint convergence of $(d_T(b_T), T^{-1/2} \sum_1^T w_t)$ is established under the same conditions as in Lemmas 3.1 and 3.2. Then, by a functional central limit theorem and the continuous mapping theorem as in Herrndorf (1984) and Chan and Wei (1988),

$$D_T(b_T^* - \beta_0) \Longrightarrow (d_0' d_0)^{-1} d_0' a_0 Z_w.$$

## Proof of Theorem 4.1

By Taylor's theorem and Assumptions 2.2 and 3.1, for sufficient large T we can write

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{T} f(x_t, b_T) &= T^{-1/2} \sum_{t=1}^{T} f(x_t, \beta_0) \\ &\quad + (T Dg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1}) \frac{\Upsilon_T}{\sqrt{T}} (b_T - \beta_0), \end{aligned} \tag{4}$$

where $\bar{b}_T^i$ for $i = 1, ..., r$, and $TDg_T(\bar{b}_T^1, ..., \bar{b}_T^r)$ are as defined in the proof of Theorem 3.1. Also from (3) in the proof of Theorem 3.1, we have the following results

$$D_T(b_T - \beta_0) = -[d_T(b_T)'(a_T TDg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1})]^{-1} d_T(b_T)' a_T \sqrt{T} g_T(\beta_0).$$

Hence, (4) can be rewritten as

$$\begin{aligned}
(5) \qquad & T^{-1/2} \sum_{t=1}^{T} f(x_t, b_T) \\
&= \{I - (TDg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1})[d_T(b_T)'(a_T TDg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1})]^{-1} \\
&\quad \times\ d_T(b_T)' a_T\} T^{-1/2} \sum_{t=1}^{T} f(x_t, \beta_0).
\end{aligned}$$

Let

$$A_T = I - (TDg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1})[d_T(b_T)'(a_T TDg_T(\bar{b}_T^1, ..., \bar{b}_T^r)\Upsilon_T^{-1})]^{-1} d_T(b_T)' a_T.$$

We can check that $A_T$ is an idempotent matrix with rank $r - q$. Also, $\hat{S}_T^{-1} A_T \Rightarrow S_w^{-1} A_0$ where $S_w^{-1} A_0$ is a symmetric matrix with $\hat{S}_T^{-1} = a_T' a_T$ for optimality of the GMM estimator. Then,

$$\begin{aligned}
\Xi &= \left[\ T^{-1/2} \textstyle\sum_{t=1}^{T} f(x_t, b_T)\ \right]' \hat{S}_T^{-1} \left[\ T^{-1/2} \textstyle\sum_{t=1}^{T} f(x_t, b_T)\ \right] \\
&= \left[S_w^{-1/2}\ T^{-1/2} \textstyle\sum_{t=1}^{T} f(x_t, \beta_0)\ \right]' M_T \left[S_w^{-1/2}\ T^{-1/2} \textstyle\sum_{t=1}^{T} f(x_t, \beta_0)\ \right]
\end{aligned}$$

where $M_T = (\hat{S}_T^{-1/2} A_T S_w^{1/2})'(\hat{S}_T^{-1/2} A_T S_w^{1/2})$. Notice that

$$M_T \Rightarrow (S_w^{-1/2} A_0 S_w^{1/2})'(S_w^{-1/2} A_0 S_w^{1/2}) = S_w^{-1/2} A_0 S_w^{1/2},$$

which is a symmetric and idempotent matrix with rank $r - q$. Then, with Lemma 3.2 the conclusion follows.