The strengths and weaknesses of L2 approximable regressors

Mynbaev, Kairat

Federal University of Ceará, Fortaleza, CE Brazil

2001

The Strengths and Weaknesses of L_2-Approximable Regressors

Kairat T. Mynbaev
Ivan Castelar
Economics Department – CAEN
Federal University of Ceará
Fortaleza, CE 60020-181
Brazil

1. Introduction

Authors of some home pages on the Internet warn visitors that "the page is under construction". We want to give an instant photo of a theory that is currently under development. The most part of the paper is about modeling (or approximating) nonstochastic regressors. One of our long-range objectives is to show that within our framework it is possible to study an autoregressive model with nonstochastic exogenous regressors. Since no such results are available at the moment, no mention will be made of models with stochastic regressors.

Consider a linear model

(1.1) $$y_n = X_n\beta + u_n$$

where X_n is a nonstochastic $n\times K$ matrix, β is a $K\times 1$ parameter vector and u_n a stochastic error vector with mean zero. Let $x_n^1,...,x_n^K$ be the columns of X_n. The asymptotics of the OLS estimator

(1.2) $$\hat{\beta}_n = \left(X_n^{'}X_n\right)^{-1}X_n^{'}y_n$$

is expressed in terms of some characteristics of sequences $\{x_n^1\},...,\{x_n^K\}$ (multiplied by some normalizing factor). Since it is hard to grasp the behavior of and manage these sequences, it is a good idea to represent them (or their normalized descendants) as images of some functions of a continuous argument. In statistical context this idea has been pursued in Moussatat (1976) and Millar (1982). Milbrodt (1992) applied it to AR(p) processes with a nonparametric trend. Precisely, L_2-generated sequences are defined as follows. Let F be a square-integrable function on (0,1). For any natural n, let z_n denote a vector with coordinates

(1.3) $$z_{nt} = \sqrt{n}\int_{(t-1)/n}^{t/n} F(x)dx \ , t = 1, ..., n,$$

(see Mibrodt (1992)). The sequence $\{z_n\}$ is called L_2-generated. With volatility of economic data, it is hard to accept such sequences as (normalized) regressors in econometrics. Therefore Mynbaev (1997) has suggested to work with sequences $\{z_n\}$ satisfying

(1.4) $$\sum_{t=1}^{n}(z_{nt}-\sqrt{n}\int_{(t-1)/n}^{t/n}F(x)dx)^2\to 0.$$

We call such a sequence L_2-approximable by F. A similar condition has been imposed by Vogelsang (1998): there exists a sequence $\{f_n\}$ of positive numbers and a function F such that

(1.5) $$f_n x_{nt}=F(t/n)+o(1).$$

As to the comparison of (1.4) and (1.5), see our comments in the end of Section 2.

All statements of asymptotic theory are based on central limit theorems (CLT's), laws of large numbers and sometimes functional central limit theorems (FCLT's). There are no universally applicable stochastic limit theorems. Each researcher has to derive his or her own results, depending on the goal and the means used. With regularly behaved regressors, such results are easily obtained from the FCLT for partial sums of random walk

(1.6) $$X_n(x)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}e_t,\quad 0\le x\le 1,$$

where $[nx]$ is the integer part of nx and e_t can be martingale differences or their moving averages (see, e.g., Bai, Lumsdaine and Stock (1998), Canjels and Watson (1997), Vogelsang (1998)). The results are expressed in terms of functionals of standard Brownian motion. This is inconvenient when one needs to know the correlation between the functionals which must be calculated independently.

In Section 2 we review the known properties of L_2-generated sequences and show that L_2-approximable ones inherit all of them. Our approach does not appeal to standard Brownian motion and allows for less smooth approximating functions. We deal with weighted sums of the form

(1.7) $$\sum_{t=1}^{n}z_{nt}u_{nt}$$

with so irregular z_n that application of the FCLT for (1.6) is not possible. This is why it takes so long to arrive to stochastic limit results. The functional-theoretical part of the job has been done in Mynbaev (2000). Among other facts, we prove that normalized polynomial and logarithmic trends are L_2-approximable.

In Section 3 we justify the choice of the normalizer. In order to do so, we derive the asymptotics of the OLS estimator from the CLT obtained in Section 2. Apart from the relaxed restrictions on the errors, that asymptotics is not new. We use it to show that normalization of X_n by the Euclidean norms of columns

$$\left(\frac{x_n^1}{\|x_n^1\|},\ldots,\frac{x_n^K}{\|x_n^K\|}\right)=X_n\left(\mathrm{diag}\left[\|x_n^1\|,\ldots,\|x_n^k\|\right]\right)^{-1}$$

is in some sense unique. We call this normalization canonical. Even though it is common knowledge in the profession, some recent authoritative sources, such as Hamilton (1994), do not mention it (or, better to say, Hamilton does not try to find a general explanation for a variety of normalizers he uses). The only rational explanation that comes to our mind is that its uniqueness has been unknown.

Because of the uniqueness, it makes sense to use it in all asymptotic statements to normalize nonstochastic regressors. We show that replacement of the classical $\sqrt{T}$-normalizer by the canonical one is not as trivial as it might seem. Section 3 is concluded by a generalization of Mynbaev's (1997) result on the asymptotics of the fitted value for model (1.1). Unlike the asymptotics of the OLS estimator, this result has no precedents and shows in full the strength of L_2-approximability.

2. L_2-approximability and a Central Limit Theorem

Let L_2 denote the space of square-integrable functions F on (0,1) provided with the norm

$$\|F\| = \left(\int_0^1 |F(x)|^2 dx \right)^{1/2}.$$

Its discrete analogue l_2 consists of sequences $\{z_t: t \ge 0\}$ having a finite norm

$$\|z\| = \left(\sum_t |z_t|^2 \right)^{1/2}.$$

R^n is the Euclidean space.

For any natural n denote

$$i_t = \left(\frac{t-1}{n}, \frac{t}{n} \right), \quad t = 1,\dots,n.$$

The discretization operator d_n maps a function $F \in L_2$ to a column-vector d_nF with coordinates

$$(d_n F)_t = \sqrt{n} \int_{i_t} F(x)dx, \quad t = 1,\dots,n.$$

The sequence $\{d_nF\}$ was called L_2-generated in the Introduction (see (1.3)). The interpolation operator D_n takes a vector $z \in R^n$ to a simple function

$$D_n z = \sqrt{n}\sum_{t=1}^{n} z_t 1(i_t)$$

Where $1(A)$ denotes the indicator of a set A:

$$1(A) = \begin{cases} 1 & \text{on } A \\ 0 & \text{outside } A \end{cases}$$

L_2-generated sequences possess some useful properties which allow one to obtain asymptotic results for linear regression models by requiring that normalized nonstochastic regressors be L_2-generated. However, in econometrics this requirement would be too restrictive. The range of applicability of L_2-generated sequences is extended by using the following definition.

Definition. Let $\{z_n\}$ be a sequence of vectors such that $z_n \in R^n$ for any natural n. We say that $\{z_n\}$ is L_2-approximable if there exists a function $F \in L_2$ such that

(2.1) $$\lim_{h\to\infty} \|z_n - d_n F\| = 0$$

(this is a compact way of writing (1.4)).

Note that F, as a member of L_2, is defined almost everywhere (a.e.), may be discontinuous and unbounded. Below we list some properties of L_2-generated and L_2-approximable sequences. First note that

(2.2) $$\|D_n z\| = \left(\sum_t |z_t|^n n\int_{i_t} dx\right)^{1/2} = \|z\|, \quad z \in R^n,$$

and by the Cauchy-Schwarz inequality

(2.3) $$\|d_n F\| \le \left(\sum_t n\int_{i_t} F^2 dx n^{-1}\right)^{1/2} = \|F\|, \quad n \ge 1.$$

Further, it is easy to check that the product $D_n\, d_n$ coincides with the Haar projector P_n where

$$P_n F = n\sum_t \int_{i_t} F(x)dx 1(i_t).$$

Therefore (2.2) and (2.3) imply

(2.4) $$\|P_n F\| \le \|F\|, \quad F \in L_2.$$

It is well known that

(2.5) $$\lim_{n\to\infty}\|F-P_nF\|=0, \quad F\in L_2$$

(see, e.g., Millar (1982)).

Property 1. $\{z_n\}$ is L_2-approximable if and only if there exists $F\in L_2$ such that

(2.6) $$\lim_{n\to\infty}\|D_nz_n-F\|=0.$$

Proof. (2.1), (2.2), and (2.5) imply

$$\|D_nz_n-F\|\le\|D_n(z_n-d_nF)\|+\|P_nF-F\|=\|z_n-d_nF\|+\|P_nF-F\|\to 0.$$

Conversely, from (2.6), (2.2), and (2.5) we get

$$\|z_n-d_nF\|=\|D_nz_n-P_nF\|\le\|D_nz_n-F\|+\|F-P_nF\|\to 0.$$

Property 2. If $\{z_n\}$ is L_2-approximable, then

(2.7) $$\lim_{n\to\infty}\max_{1\le t\le n}|z_{nt}|=0.$$

Proof: By the Cauchy-Schwarz inequality and absolute continuity of the Lebesgue integral

$$\max_t|(d_nF)_t|\le\max_t\left(\int_{i_t}F^2dx\right)^{1/2}\to 0, \quad n\to\infty.$$

This relation and (2.1) yield

$$\max_t|z_{nt}|\le\|z_n-d_nF\|+\max_t|(d_nF)_t|\to 0.$$

Property 3. If z_n^i is L_2-approximated by F_i, i = 1,2, then

$$\lim_{n\to\infty}(z_n^1)'z_n^2=\int_0^1F_1(x)F_2(x)dx.$$

Proof. By (2.2), (2.6), and the continuity of the norm

(2.8) $$\lim_{n\to\infty}\|z_n^i\|=\lim_{n\to\infty}\|D_nz_n^i\|=\|F_i\|, \quad i=1,2.$$

In Mynbaev (2000) it has been proved that for L_2-generated sequences

$$\lim_{n\to\infty}(d_nF_1)'d_nF_2=\int_0^1 F_1(x)F_2(x)dx.$$

Using these equations and (2.1), we get

$$\left|(z_n^1)'z_n^2-\int_0^1 F_1F_2dx\right|\le\left|(z_n^1-d_nF_1)'z_n^2\right|+\left|(d_nF_1)'(z_n^2-d_nF_2)\right|$$
$$+\left|(d_nF_1)'d_nF_2-\int_0^1 F_1F_2dx\right|\le\|z_n^1-d_nF_1\|\|z_n^2\|+\|d_nF_1\|\|z_n^2-d_nF_2\|$$
$$+\left|(d_nF_1)'d_nF_2-\int_0^1 F_1F_2dx\right|\to 0.$$

If the normalized regressors are L_2-approximable, then, using Properties 2 and 3 and stochastic limit results from Davidson (1994), it is possible to replace independent errors by martingale differences (m.d.'s) in Anderson's (1971) asymptotics of the OLS estimator. These days a more general error structure, such as mixing or moving averages of m.d.'s, is common in the econometrics literature (see the references in Davidson (1994) regarding mixing and in Vogelsang (1998) concerning moving averages and the so-called local-to-unity asymptotics). To extend the Anderson theorem to errors which are moving averages of m.d.'s we need the following property.

For a given sequence $\{\psi_j: j\ge 0\}$ of real numbers define operators $\Psi_n: R^n\to R^n$ and $\Phi_n: R^n\to l_2$ by

$$\Psi_n z=\left(\sum_{j=t}^n z_j\psi_{j-t}\right)_{t=1}^n,\quad \Phi_n z=\left(\sum_{j=1}^n z_j\psi_{j+t}\right)_{t=0}^\infty.$$

Let

$$\alpha=\sum_j|\psi_j|,\quad \beta=\sum_j j|\psi_j|,\quad \gamma=\sum_j\psi_j.$$

It is easy to prove that if $\alpha<\infty$, then

(2.9) $$\|\Psi_n z\|\le\alpha\|z\|,\quad \|\Phi_n z\|\le\alpha\|z\|,\quad z\in R^n,\quad n\ge 1,$$

and that $\beta<\infty$ implies $\alpha<\infty$ and convergence of γ.

Property 4. If $\{z_n\}$ is L_2-approximable and $\beta<\infty$, then

$$\lim_{n\to\infty}\|(\Psi_n-\gamma)z_n\|=0,\quad \lim_{n\to\infty}\|\Phi_n z_n\|=0.$$

Proof: Let $\{z_n\}$ be L_2-approximated by F. In Mynbaev (2000) it has been proved that

$$\lim_{n\to\infty}\|(\Psi_n-\gamma)d_nF\|=\lim_{n\to\infty}\|\Phi_nd_nF\|=0.$$

Hence, taking also into account (2.1) and (2.9)

$$\|(\Psi_n-\gamma)z_n\|\le\|(\Psi_n-\gamma)(z_n-d_nF)\|+\|(\Psi_n-\gamma)d_nF\|$$

$$\le(\alpha+|\gamma|)\|z_n-d_nF\|+\|(\Psi_n-\gamma)d_nF\|\to 0$$

and

$$\|\Phi_nz_n\|\le\|\Phi_n(z_n-d_nF)\|+\|\Phi_nd_nF\|\to 0.$$

Denote $M_nF=D_n\Psi_nd_nF$. In Mynbaev (2000) it has been proved that

$$\|M_nF-\gamma F\|\to 0.$$

This property is not applied in econometrics but it is interesting because the operator M_n is similar to the operator M in the Fourier analysis where for a function F on the unit circle decomposed as $F=\sum c_k\exp(ikx)$ one can put $MF=\sum m_kc_k\exp(ikx)$ for a given sequence of numbers $\{m_k\}$.

Property 5. a) Suppose that for a given $\{z_n\}$ there exists F from the space L_∞ of essentially bounded on (0,1) functions such that

$$\|D_nz_n-F\|=\operatorname*{ess\,sup}_{x\in(0,1)}|(D_nz_n)(x)-F(x)|\to 0.$$

Then $\{z_n\}$ is L_2-approximable by F.

b) Let F be continuous on [0,1] and suppose that for each n there are points $p_1,p_2,...,p_n$ such that $p_t\in i_t$ for any t = 1,...,n. Put $z_{nt}=n^{-1/2}F(p_t),\quad t=1,...,n$. Then $\{z_n\}$ is L_2-approximable by F.

Proof: Statement a) follows from the inequality

$$\|D_nz_n-F\|\le\|D_nz_n-F\|_\infty.$$

b) By uniform continuity of F on [0,1] for any $\varepsilon>0$ there exists n_0 such that

$$|F(p_t)-F(x)|\le\varepsilon,\quad x\in i_t,\quad n\ge n_0.$$

Hence,

$$\|D_n z_n - F\|_\infty = \left\|\sum_{t=1}^{n} F(p_t)1(i_t) - F\right\|_\infty = \max_t \max_{x\in i_t}|F(p_t) - F(x)| \le \varepsilon, \quad n \ge n_0.$$

It remains to apply part a).

Proposition 1. Consider a polynomial trend

$$p_n = (1^{k-1}, 2^{k-1}, \dots, n^{k-1})$$

where k is natural. Let $z_n = p_n/\|p_n\|$ be the normalized sequence. Then it is L_2-approximable by $F(x) = \sqrt{2k-1}x^{k-1}$.

Proof. In Hamilton (1994), p. 456, it is shown that

$$\sum_{t=1}^{n} t^l = (1+o(1))\frac{n^{l+1}}{l+1}, \quad l = 1,2,\dots$$

Therefore

$$\|p_n\| = (1+o(1))\left(n^{2k-1}/(2k-1)\right)^{1/2}$$

and

$$z_n = (1+o(1))\frac{p_n}{\left(n^{2k-1}/(2k-1)\right)^{1/2}}$$

$$= (1+o(1))\left(\frac{2k-1}{n}\right)^{1/2}\left(\left(\frac{1}{n}\right)^{k-1}, \left(\frac{2}{n}\right)^{k-1}, \dots, \left(\frac{n}{n}\right)^{k-1}\right).$$

Hence,

$$D_n z_n = (1+o(1))\sqrt{2k-1}\sum_{t=1}^{n}\left(\frac{t}{n}\right)^{k-1} 1(i_t),$$

wherefore

$$\|D_n z_n - F\|_\infty = \sqrt{2k-1}\max_{1\le t\le n}\max_{x\in i_t}\left|\left(\frac{t}{n}\right)^{k-1}(1+o(1)) - x^{k-1}\right|.$$

Since the last expression tends to zero, the statement follows from Property 5.

Consider a geometric progression

$$g_n = \left(a^0, a^1, \ldots, a^{n-1}\right), \quad a \in R.$$

When a = 1, g_n is a (constant) polynomial trend. All other cases are covered in the next proposition.

Proposition 2. If $a \neq 1$, then $z_n = g_n / \|g_n\|$ is not L_2-approximable.

Proof. Consider $|a| < 1$. From

$$\|g_n\| = \left(\sum_{t=0}^{n-1} a^{2t}\right)^{1/2} = \left(\frac{1-a^{2n}}{1-a^2}\right)^{1/2} = \frac{1+o(1)}{\sqrt{1-a^2}}$$

it follows that

$$z_n = (1+o(1))\sqrt{1-a^2}\left(a^0, \ldots, a^{n-1}\right),$$

so that

$$D_n z_n = (1+o(1))\sqrt{n\left(1-a^2\right)}\sum_{t=1}^{n} a^{t-1} 1\left(i_t\right).$$

For a fixed $\varepsilon \in (0,1)$ denote $t_\varepsilon = [n\varepsilon] + 1$ where $[n\varepsilon]$ is the integer part of $n\varepsilon$. Since $\varepsilon \in i_{t_\varepsilon}$, we have

$$\text{(2.10)} \qquad \int_\varepsilon^1 |D_n z_n|^2 dx \le \sum_{t=t_\varepsilon}^{n} \int_{i_t} |D_n z_n|^2 dx = (1+o(1)) n\left(1-a^2\right)\sum_{t=t_\varepsilon}^{n} a^{2(t-1)}\frac{1}{n}$$

$$\le (1+o(1))\left(1-a^2\right)\sum_{t=[n\varepsilon]+1}^{\infty} a^{2(t-1)} \le c a^{2[n\varepsilon]} \to 0.$$

Suppose, $\{z_n\}$ is L_2-approximable. (2.6) and (2.10) give

$$\left(\int_\varepsilon^1 F^2 dx\right)^{1/2} \le \left(\int_0^1 |F - D_n z_n|^2 dx\right)^{1/2} + \left(\int_\varepsilon^1 |D_n z_n|^2 dx\right)^{1/2} \to 0.$$

Since $\varepsilon > 0$ can be arbitrarily small, this means that F = 0 a.e. On the other hand, (2.8) (applied to z_n and F) and normalization of z_n give

$$\text{(2.11)} \qquad \|F\| = 1.$$

The contradiction finishes the proof in the case $|a| < 1$.

The case $|a| > 1$ is treated similarly. The difference is that

$$\|g_n\| = (1+o(1))\frac{|a|^n}{\sqrt{a^2-1}}, \quad z_n = (1+o(1))\frac{\sqrt{a^2-1}}{|a|^n}(a^0,...,a^n)$$

and $F = 0$ a.e. on intervals $(0, 1-\varepsilon)$.

Let $a = -1$. Then

$$\|g_n\| = \sqrt{n}\,, \quad z_n = n^{-1/2}((-1)^0, (-1)^1, ..., (-1)^{n-1})$$

and

(2.12) $$D_n z_n = \sum_{t=1}^{n} (-1)^{t-1} 1(i_t).$$

Suppose that $\{z_n\}$ is L_2-approximable by F and consider any interval $(a, b) \subset (0,1)$. One has

$$\frac{[na]}{n} \le a < \frac{[na]+1}{n}, \quad \frac{[nb]}{n} \le b < \frac{[nb]+1}{n}.$$

Therefore, denoting $S_n = \bigcup_{t=[na]+1}^{[nb]+1} i_t$, we can write

(2.13) $$\left|\int_a^b F dx\right| \le \left|\int_{S_n} F dx\right| + \left|\int_{[na]/n}^{a} F dx\right| + \left|\int_b^{([nb]+1)/n} F dx\right|.$$

The last two terms at the right tend to zero by absolute continuity of the Lebesgue integral. We bound the first one as follows

(2.14) $$\left|\int_{S_n} F dx\right| \le \left|\int_{S_n} (F - D_n z_n) dx\right| + \left|\int_{S_n} D_n z_n dx\right|$$

$$\le \| F - D_n z_n \| + 1/n \to 0$$

where we have used the Cauchy-Schwarz inequality and (2.12). Thus,

(2.15) $$\int_a^b F dx = 0 \text{ for any } (a,b) \subset (0,1)$$

and $F = 0$ a.e. This conclusion contradicts (2.11).

Note that exponential trends

$$(e^b,\ldots,e^{nb}),\quad b\in R,$$

are geometric progressions and are not L_2-approximable, unless $b = 0$. Next we consider logarithmic trends (k is natural)

$$\lambda_n = (\ln^k 1,\ldots,\ln^k n).$$

<u>Proposition 3.</u> The sequence $z_n = \lambda_n/\|\lambda_n\|$ is L_2-approximable by $F(x) \equiv 1$ (for any k).

<u>Proof.</u> Denote

$$I_k(n) = \int_1^n \ln^k x dx,\ \ k \ge 0.$$

Obviously,

$$I_k(n) = x\ln^k x\Big|_1^n - k\int_1^n \ln^{k-1} xdx = n\ln^k n - kI_{k-1},\quad k \ge 1,$$

$$I_0(n) = \int_1^n dx = n-1.$$

By recurrent substitution we see that there exist numbers C_k, ..., C_0 which do not depend on n and such that

$$I_k(n) = n\ln^k n + C_k n\ln^{k-1} n + \ldots + C_1 n + C_0.$$

Hence, for any $k \ge 1$

(2.16) $$I_k(n) = (1+o(1))n\ln^k n.$$

This implies

(2.17) $$I_k(n+1) = (1+o(1))(n+1)\ln^k(n+1)$$

$$= (1+o(1))(n\ln^k n)(1+\frac{1}{n})\left(\frac{\ln n + \ln(1+1/n)}{\ln n}\right)^k =$$

$$= (1+o(1))n\ln^k n.$$

Note that

$$(2.18)\qquad I_{2k}(n) \le \sum_{t=2}^{n} \ln^{2k} t \le \|\lambda_n\|^2 \le I_{2k}(n+1).$$

(2.16) – (2.18) imply

$$\|\lambda_n\| = (1+o(1))\sqrt{n}\ln^k n.$$

So

$$z_n = \frac{1+o(1)}{\sqrt{n}\ln^k n}(\ln^k 1,...,\ln^k n)$$

and

$$D_n z_n = \frac{1+o(1)}{\ln^k n}\sum_{t=1}^{n} 1(i_k)\ln^k t.$$

Since $|\ln t/\ln n|^k \le 1,\ 1 \le t \le n,$ the difference between $D_n z_n$ and f_n defined by

$$f_n = \frac{1}{\ln^k n}\sum_{t=1}^{n} 1(i_t)\ln^k t$$

tends to zero uniformly on [0, 1].

Fix $\varepsilon \in (0,1)$. If $[\varepsilon n]+1 \le t \le n,$ then $\varepsilon \le t/n \le 1$ and there exists $c_1(\varepsilon) > 0$ such that

$$|\ln(t/n)| \le e_1 \text{ for } [\varepsilon n]+1 \le t \le n.$$

Hence, there exists $n_1(\varepsilon)$ such that for these t

$$(2.19)\qquad \left|\left(\frac{\ln t}{\ln n}\right)^k - 1\right| = \left|\left(\frac{\ln n + \ln(t/n)}{\ln n}\right)^k - 1\right| \le \varepsilon, \qquad n \ge n_1(\varepsilon)..$$

If $1 \le t \le [\varepsilon n],$ then

$$(2.20)\qquad \left|\left(\frac{\ln t}{\ln n}\right)^k - 1\right| \le \left|\left(\frac{\ln t}{\ln n}\right)^k\right| + 1 \le 2.$$

Obviously,

$$f_n - F = S_1 + S_2$$

where

$$S_1 = \sum_{t=1}^{[\varepsilon n]} \left(\left(\frac{\ln t}{\ln n} \right)^k - 1 \right) 1(i_t), \ \ S_2 = \sum_{t=[\varepsilon n]+1}^{n} \left(\left(\frac{\ln t}{\ln n} \right)^k - 1 \right) 1(i_t).$$

From $\bigcup_{t=1}^{[\varepsilon n]} i_t = (0, \frac{[\varepsilon n]}{n}) \subset (0,\varepsilon)$ and (2.20) it follows that (mes denotes the Lebesgue measure)

$$\|S_1\| \le 2\left(\text{mes}\left(\bigcup_{t=1}^{[\varepsilon n]} i_t \right) \right)^{1/2} \le 2\varepsilon^{1/2}.$$

(2.19) implies

$$\|S_2\| \le \varepsilon\left(\text{mes}\left(\bigcup_{t=[\varepsilon n]+1}^{n} i_t \right) \right)^{1/2} \le \varepsilon.$$

Thus,

$$\|f_n - F\| \le \|S_1\| + \|S_2\| \le 2\varepsilon^{1/2} + \varepsilon, \ n \ge n_1(\varepsilon),$$

which proves the statement.

Let $\{\{e_{nt}, G_{nt}\}: -\infty < t \le n;\ n = 1,2,\ldots\}$ be an m.d. array (see Davidson (1994) for all probability notions and facts; as a first approximation, it is sufficient to think of e_{nn}, $e_{n,n\text{-}1}$, …, $e_{n,n\text{-}j}$,… as independent identically distributed). Denote u_n the moving averages of e_{nt}:

$$u_n = \left(\sum_{j=0}^{\infty} e_{n,t-j}\psi_j \right)_{t=1}^{n}, \quad n = 1,2,... \tag{2.21}$$

where the ψ_j are the same as in Property 4. For a sequence $\{Z_n: n > K\}$ of $n\times K$ nonstochastic matrices with columns $z_n^1,...,z_n^K$ define random vectors

$$Z_n' u_n = \left(\sum_{t=1}^{n} z_{nt}^k u_{nt} \right)_{k=1}^{K}.$$

For a row-vector $F = (F_1,\ldots,F_K)$ with $F_k \in L_2$, put

$$V = \int_0^1 F'F dx = \left(\int_0^1 F_k(x) F_l(x) dx \right)_{k,l=1}^{K}.$$

<u>Theorem 1.</u> Suppose that

A) $E(e_{nt}^2 \mid G_{n,t-1}) = \sigma^2$ for some $\sigma > 0$ and all t, n,

B) e_{nt}^2 are uniformly integrable,

C) the sequence $\{z_n^k\}$ is L_2-approximable by $F_k \in L_2$, $k = 1, ..., K$,

D) V is positive definite (that is, $F_1, ..., F_K$ are linearly independent),

E) $\beta < \infty$ and $\gamma \neq 0$.

Then

(2.22) $$Z_n' u_n \xrightarrow{d} N(0, (\sigma\gamma)^2 V),$$

(2.23) $$\lim_{n\to\infty} Z_n' Z_n = V.$$

For L_2-generated $\{z_n^k\}$ this result has been proved in Mynbaev (2000). To obtain the proof for the case under consideration, it suffices to use Properties 3 and 4 instead of Lemmas 1 and 6, respectively, in the proof given in Mynbaev (2000).

Some comments are in order. CLT's have many formats, depending on the intended application. Our CLT is about convergence in distribution of weighted sums (1.7) of random variables u_{nt} with deterministic weights z_{nt}. There are few papers devoted specifically to this type. The results in Srinivasan and Zhou (1995) and Yoshihara (1997a, 1997b) are aimed at censored regression models and hard to compare with Theorem 1. Many econometrics papers explicitly or implicitly contain CLT's as intermediate steps. As we can judge by the most recent sources (Bai, Lumsdaine and Stock (1998), Canjels and Watson (1997), Vogelsang (1998)), conditions A), D), and E) are standard requirements. Instead of B) these authors assume a stronger condition

$$\sup_{t,n} Ee_{nt}^4 < \infty \quad .$$

Regarding C), the only alternative we have met in the literature is Volgelsang's (1998) condition (1.5). Since it involves point values of F, we think that F should be continuous even though Vogelsang does not mention continuity. For continuous F (1.5) is equivalent to

$$\left\| \| x_n \| \sqrt{n} f_n D_n z_n - F \right\|_\infty \to 0.$$

This condition cannot be directly compared to the condition from Property 5a) sufficient for L_2-approximability because of the unspecified sequence $\{f_n\}$. But if $f_n = n^{-1/2}/\|x_n\|$, then it implies L_2-approximability.

3. Normalization of Nonstochastic Regressors

Here we consider model (1.1) with u_n defined in (2.21). Denote

(3.1) $$Y_n = \mathrm{diag}\left[\|x_n^1\|, ..., \|x_n^K\|\right], \quad Z_n = X_n Y_n^{-1}$$

From (1.1) and (1.2) it is easy to obtain

(3.2) $$Y_n(\hat{\beta}_n - \beta) = (Z_n^{'}Z_n)^{-1} Z_n^{'} u_n.$$

Application of Theorem 1 immediately leads to the following result.

Theorem 2. Let e_{nt}, z_n^k, and ψ_j satisfy assumptions of Theorem 1. Then

(3.3) $$Y_n(\hat{\beta}_n - \beta) \xrightarrow{d} \xi \in N(0, (\sigma\gamma)^2 V^{-1}).$$

In principle, Theorem 2 is not new. The model considered is so simple that it is difficult to indicate an immediate predecessor. All comments about the conditions A) through E) apply here. In particular, we believe that conditions B) and C) are more general than those which allow one to derive a CLT from the FCLT for (1.6). The statement, besides being conditional on the literature we have access to, also depends on the sequence $\{\psi_j\}$. In the trivial case

(3.4) $$\psi_0 = 1,\ \ \psi_j = 0, \quad j \ge 1,$$

we are taken back to Anderson's (1971) result. He has imposed conditions (2.7) and (2.23) with det $V \ne 0$ (his assumption of independent errors is easily relaxed to m.d.'s). These conditions are weaker than the pair C) + D) by Properties 2 and 3. Theorem 2 covers polynomial and logarithmic trends as we show in Examples 1 and 2 below (it is well known that geometric progressions and exponential trends stand out: convergence takes place but the limiting distribution in general is not normal).

The main reason we state Theorem 2 is to discuss one point that seems to have been missed in the econometrics literature: the choice of the normalizer. We need a couple of definitions for the discussion.

Our derivation of (3.3) follows the conventional scheme that can be described as follows. 1) Using some diagonal matrix, such as Y_n, the OLS estimator is transformed to (3.2). 2) Condition (2.23) along with $\det V \ne 0$ is imposed. 3) A CLT is applied to prove convergence of $Z_n^{'}u_n$ in distribution. Convergence of the product at the right of (3.2) then follows from Cramér's theorem.

We call Y_n defined in (3.2) a canonical normalizer. It was used, for example, in Grenander and Rosenblatt (1957) and Anderson (1971). Traditionally another normalizer, $\sqrt{n}$, is widely used in econometrics. Polynomial trends give rise to other powers of n (see, e.g., Hamilton (1994)). Thus, there is uncertainty as to the choice or uniqueness of the normalizer. We shall show that, as for as a model with nonstochastic regressors is concerned, the normalizer Y_n is in some sense unique. The fact that the normalizer must depend on the model is common knowledge, but interaction with our colleagues convinced us that its uniqueness for a given model is not.

Consider a sequence of diagonal matrices $\overline{Y}_n = \mathrm{diag}[\overline{y}_{n1},...,\overline{y}_{nk}]$ with positive elements on the main diagonal and put $\overline{Z}_n = X_n \overline{Y}_n^{-1}$. We say that $\{\overline{Y}_n\}$ is a conventional-scheme-compliant (CSC) normalizer if

(3.5) there exists $\lim_{n\to\infty} \overline{Z}_n^{'} \overline{Z}_n = \overline{V}$, $\det \overline{V} \neq 0$,

and

(3.6) $$\begin{cases} \overline{Z}_n^{'} u_n \xrightarrow{d} N(0, (\sigma\gamma)^2 V) \text{ for any } e_{nt} \text{ and } \psi_j \\ \text{satisfying conditions A), B), E) of Theorem 1} \end{cases}$$

The columns of $\overline{Z}_n$ are not required to be L_2-approximable in this definition.

<u>Proposition 4</u>. If $\{\overline{Y}_n\}$ is a CSC normalizer and $\{\Delta_n\}$ is a sequence of $K\times K$ diagonal matrices with positive elements such that

(3.7) there exists $\lim_{n\to\infty} \Delta_n = \Delta$, $\det \Delta \neq 0$,

then $\Delta_n \overline{Y}_n$ is also a CSC normalizer.

<u>Proof</u>. From (3.5) and (3.7)

$$\lim (X_n (\Delta_n \overline{Y}_n)^{-1})' X_n (\Delta_n \overline{Y}_n)^{-1} = \lim (X_n \overline{Y}_n^{-1} \Delta_n^{-1})' X_n \overline{Y}_n^{-1} \Delta_n^{-1} =$$

$$= \lim \Delta_n^{-1} \overline{Z}_n^{'} \overline{Z}_n \Delta_n^{-1} = \Delta^{-1} \overline{V} \Delta^{-1}, \quad \det \Delta^{-1} \overline{V} \Delta^{-1} \neq 0$$

By the Cramér theorem (3.6) and (3.7) imply

$$(X_n (\Delta_n \overline{Y}_n)^{-1})^{'} u_n = (\overline{Z}_n \Delta_n^{-1})^{'} u_n \xrightarrow{d} N(0, (\sigma\gamma)^2 \Delta^{-1} \overline{V} \Delta^{-1})$$

for any e_{nt} and ψ_j satisfying conditions A), B), E) of Theorem 1. Hence, $\Delta_n \overline{Y}_n$ is a CSC normalizer.

Proposition 4 means that it makes sense to talk about uniqueness of the canonical normalizer up to a factor satisfying (3.7). All such a factor does is change the variance of the limit distribution in (2.22) and (3.3).

<u>Proposition 5.</u> If $\overline{Y}_n$ is some CSC normalizer, then the canonical normalizer is also, and there exists a sequence $\{\Delta_n\}$ of diagonal matrices satisfying (3.7) such that $Y_n = \Delta_n \overline{Y}_n$.

<u>Proof.</u> Denote $y_n^k = \|x_n^k\|$, $\overline{y}_n^k$, $\overline{v}_{kk}$ the diagonal elements of Y_n, $\overline{Y}_n$, and $\overline{V}$, respectively. The main diagonal of the limit relation in (3.5) gives

$$(\bar{z}_n^{\,k})'\bar{z}_n^{\,k} = \|x_n^k\|^2/(\bar{y}_n^{\,k})^2 \to \bar{v}_{kk},$$

that is $y_n^k/\bar{y}_n^{\,k} \to (\bar{v}_{kk})^{1/2}$. In matrix notation this means that $Y_n\bar{Y}_n^{-1} \to \Delta$ where

$$\Delta = \mathrm{diag}[(\bar{v}_{11})^{1/2},\ldots,(\bar{v}_{kk})^{1/2}],\ \ \det\Delta \neq 0$$

Denoting $\Delta_n = Y_n\bar{Y}_n^{-1}$, we see that (3.7) is true, $Y_n = \Delta_n\bar{Y}_n^{-1}$, so by Proposition 4 $\{Y_n\}$ is a CSC normalizer.

Summarizing, the canonical normalizer is more flexible (it adjusts to the regressor) and is unique up to a factor (with a nondegenerate limit) which preserves convergence in distribution to a normal variable. If for a model with nonstochastic regressors there exists some CSC normalizer, then Y_n can be used as well. It would be mathematically correct and didactically justified to rewrite all classical statements of the asymptotic theory using Y_n. This is a formidable task we do not undertake. We consider just one statistic to show that not everything is as straightforward as it might seem at the first glance.

Consider the statistic

$$\varphi_n = \frac{R'\hat{\beta}_n - r}{\sqrt{s^2 R'(X_n'X_n)^{-1}R}}$$

used to test H$_0$: $R'\beta = r$ against the alternative H$_a$: $R'\beta \neq r$. Here the vector $R = (R_1,\ldots, R_K)'$ and the real number r are given and s^2 is the estimator of σ^2,

$$s^2 = \frac{e_n'(I - X_n(X_n'X_n)^{-1}X_n')e_n}{n-K}$$

(for simplicity we assume (3.4) and maintain all other hypotheses of Theorem 2). Following the assumed normalization, Y_n should be introduced everywhere. Denoting

$$\rho_n = Y_n^{-1}R,\ \ h_n = \left(Z_n'Z_n\right)^{-1/2},\ \ f_n = h_n\rho_n/\|h_n\rho_n\|$$

and using the null hypothesis, we have

$$\varphi_n = \frac{\left(Y_n^{-1}R\right)'Y_n\left(\hat{\beta}_n - \beta\right)}{\sqrt{s^2\left(Y_n^{-1}R\right)'\left(Z_n'Z_n\right)^{-1}Y_n^{-1}R}} =$$

(3.8)

$$= \frac{\rho_n\left(Z_n'Z_n\right)^{-1}Z_n'e_n}{\sqrt{s^2\rho_n'\left(Z_n'Z_n\right)^{-1}\rho_n}} = \frac{(h_n\rho_n)'h_nZ_n'e_n}{\sqrt{s^2(h_n\rho_n)'h_n\rho_n}} = \frac{1}{s}f_n'h_nZ_n'e_n.$$

By Theorem 1 $Z_n'e_n$ converges in distribution. Assuming that

(3.9) $\lim h_n = V^{-1/2}$

and

(3.10) there exists $\lim f_n = f$,

we can pass to the limit in (3.8) (using also $\text{plim} s = \sigma$ which is proved as usually).

Conditions (3.9) and (3.10) have been chosen as the most plausible, in view of (2.23) and the normalization $\|f_n\| = 1$. Observe that (3.9) does not follow from (2.23). The reason is that the square root of a matrix is not a continuous function of its argument (see Kato (1966)). It would be wrong to require existence of a nondegenerate $\lim \rho_n$ instead of (3.10), because usually $\lim Y_n = \infty$ (excluding such pathologies as geometric progressions).

The transformation in (3.8) and conditions (3.9) and (3.10) are the best we could think of (any suggestions are welcome). To compare, consider the case of a scalar identity Y_n,

$$Y_n = \tau_n I_k$$

where $\tau_n \in (0, \infty)$ (in particular, τ_n can be $\sqrt{n}$). In place of (3.8) we can write

$$\varphi_n = \frac{R'(Z_n' Z_n)^{-1} Z_n' e_n}{\sqrt{s^2 R'(Z_n' Z_n)^{-1} R}}.$$

Using Theorem 2, we can pass to the limit without imposing conditions of type (3.9), (3.10). Thus, the fact that in general Y_n is not a scalar identity matrix forces us to impose new conditions in order to be able to find the limit statistic. Analysis of some other statements of the classical asymptotic theory in the light of the canonical normalizer will appear in Mynbaev and Lemos (to be published).

Example 1. Let $x_n^1, ..., x_n^k$ be polynomial trends of degrees 0,...,K–1, respectively. Instead of normalizing X_n by the canonical normalizer

$$\text{diag}\left[n^{1/2}, \left(\frac{n^3}{3} \right)^{1/2}, ..., \left(\frac{n^{2k-1}}{2k-1} \right)^{1/2} \right]$$

(see the proof of Proposition 1), we can use a simpler matrix

$$Y_n = \text{diag}[n^{1/2}, n^{3/2}, ..., n^{(2k-1)/2}].$$

This corresponds to L_2-approximation of $x_n^k / \left(\|x_n^k\| \sqrt{2k-1} \right)$ by $F_k(x) = x^{k-1}$ in which case

$$\int_0^1 F_k F_l dx = \int_0^1 x^{k+l-2} = 1/(k+l-1).$$

Hence, if e_{nt} and ψ_j satisfy A), B), E), then by Theorem 2 (3.3) is true with

$$V = \begin{pmatrix} 1 & 1/2 & ... & 1/K \\ 1/2 & 1/3 & ... & 1/(K+1) \\ ... & ... & ... & ... \\ 1/K & 1/(K+1) & ... & 1/(2K+1) \end{pmatrix}$$

V is known under the name of a Hilbert matrix.

This application is not new (see, e.g., parts (a) and (d) of Lemma 1 and references in Sims, Stock, and Watson (1990) or Section 16.1 in Hamilton (1994)). The main reason we state this and the next example is to show that Proposition 5 can be used both in a positive sense (if some CSC-compliant normalizer exists, then the canonical normalizer can be used as well, as in Example 1) and in a negative sense (if the canonical normalizer is not CSC-compliant, then there is no CSC-compliant normalizer, as in Example 2).

<u>Example 2.</u> If $K > 1$, and

$$x_n^k = (\ln^k 1, ..., \ln^k n), \qquad k = 1, ..., K,$$

then there is no CSC-compliant normalizer. If $K = 1$, then Theorem 2 is applicable.

Indeed, if $K > 1$ and there were one, then we could use the canonical normalizer. The normalized columns would be L_2-approximable by $F_k \equiv 1, \quad k = 1, ..., K$. But these functions are linearly dependent (all of the elements of V are equal to 1).

L_2-approximability allows one to obtain new, unprecedented asymptotic results. One example is the asymptotics of the fitted value

$$\hat{y}_n = X_n \hat{\beta}_n = X_n (X_n' X_n)^{-1} X_n' y_n$$

obtained in Mynbaev (1997). Similar to (3.2), one has

(3.11) $$\hat{y}_n - X_n \beta = Z_n [(Z_n' Z_n)^{-1} Z_n' u_n].$$

The term in the brackets at the right converges by Theorem 2 but the factor Z_n in front of it does not, because of (2.7). However, requiring L_2-approximability of the columns of Z_n, we can premultiply (3.11) by D_n to get

$$D_n (\hat{y}_n - X_n \beta) = [D_n Z_n][(Z_n' Z_n)^{-1} Z_n' u_n]$$

where both factors at the right converge.

Theorem 3. Under the conditions of Theorem 1 one has

$$D_n(\hat{y}_n - X_n\beta) \xrightarrow{d} \xi_1 F_1 + \ldots + \xi_K F_K \tag{3.12}$$

where $\xi \in N(0,(\sigma\gamma)^2 V)$ (see (3.3)).

The linear combination at the right of (3.12) is a random element of L_2. It is the random vector of coefficients ξ that is normally distributed, not the linear combination itself. When regressing on trends, results such as (3.12) can be used to perform interval estimation and hypothesis testing for quantities measured by the area under the fitted curve.

References

Anderson, T. W. (1971) The Statistical Analysis of Times Series. Wiley and Sons. New York.

Bai, J., R. L. Lumsdaine and J. H. Stock (1998) Testing for and Dating Common Breaks in Multivariate Time Series. Review of Economic Studies, 65, 395-432.

Canjels, E. and M. W. Watson (1997) Estimating Deterministic Trends in the Presence of Serially Correlated Errors. Review of Economics and Statistics. 79, 184-200.

Davidson, J. (1994) Stochastic Limit Theory. Oxford: Oxford University Press.

Grenander, U. and M. Rosenblatt (1957) Statistical Analysis of Stationary Time Series. New York. Chelsea Publishing Company.

Hamilton, J. D. (1994) Time Series Analysis. Princeton University Press. Princeton, New Jersey.

Kato, T. (1966) Perturbation Theory for Linear Operators. New York: Springer – Verlag.

Milbrodt, H. (1992) Testing Stationarity in the Mean of Autoregressive Processes with a Nonparametric Regression Trend. Ann. Statist. 20, N.º 3, 1426-1440.

Millar, P. W. (1982) Optimal Estimation of a General Regression Function. Ann. Statist. 10, 717-740.

Moussatat, M. W. (1976) On the Asymptotic Theory of Statistical Experiments and Some of Its Applications. Ph.D. dissertation, Univ. California, Berkeley.

Mynbaev, K. T. (1997) Linear Models with Regressors Generated by Square-Integrable Functions. Programa e Resumos, 7 Escola de Séries Temporais e Econometria. 6 a 8 de agosto, 80-82.

Mynbaev, K. T. (2000) Limits of Weighted Sums of Random Variables. Discussion Text. Economics Dept. Federal University of Ceará, Fortaleza, Brazil. 20pp.

Mynbaev, K. T. and A. Lemos (to be published). Econometria. Um Guia Conciso da Teoria Clássica. Editora da Fundação Getúlio Vargas, Rio de Janeiro, RJ.

Sims, C. A., J. H. Stock, M. W. Watson (1990) Inference in Linear Time Series Models with Some Unit Roots. Econometrica, 58, N.º 1, 113-144.

Srinivasan, C. and M. Zhou (1995) A Central Limit Theorem for Weighted and Integrated Martingales. Scand. J. Statist. 22, 493-504.

Vogelsang, T. J. (1998) Trend Function Hypothesis Testing in the Presence of Serial Correlation. Econometrica 66, N.º 1, 123-148.

Yoshihara, Ken-ichi (1997a) Central Limit Theorems for Weighted D[0, 1]-Valued Mixing Sequences. I. Functional Central Limit Theorems for Weighted Sums. Proceedings of the Second World Congress of Nonlinear Analysts, Part 6, (Athens, 1966). Nonlinear Analysis, 30, No. 6, 3569-3573.

Yoshihara, Ken-ichi (1997b) Central Limit Theorems for Weighted D[0, 1]-Valued Mixing Sequences. II. Functional Central Limit Theorems for Integrated Variables. Yokohama Math. J. 44, N.º 2, 157-167.