The strengths and weaknesses of L2 approximable regressors

Mynbaev, Kairat

Federal University of Ceará, Fortaleza, CE Brazil

2001

Online at https://mpra.ub.uni-muenchen.de/9056/
MPRA Paper No. 9056, posted 10 Jun 2008 06:39 UTC
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

\begin{equation}
    y_n = X_n \beta + u_n \tag{1.1}
\end{equation}

where $X_n$ is a nonstochastic $n \times K$ matrix, $\beta$ is a $K \times 1$ parameter vector and $u_n$ a stochastic error vector with mean zero. Let $x_{n1}^1, ..., x_{nK}^K$ be the columns of $X_n$. The asymptotics of the OLS estimator

\begin{equation}
    \hat{\beta}_n = \left( X_n^T X_n \right)^{-1} X_n^T y_n \tag{1.2}
\end{equation}

is expressed in terms of some characteristics of sequences $\{x_{n1}^1\}, ..., \{x_{nK}^K\}$ (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

\begin{equation}
    z_{ni} = \sqrt{n} \int_{(i-1)/n}^{i/n} F(x) dx \ , \ i = 1, ..., n, \tag{1.3}
\end{equation}

(see Milbrodt (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
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

$$f_n x_n = F(t/n) + o(1).$$

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.

$$X_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nx]} e_i, \quad 0 \leq x \leq 1,$$

where $[nx]$ is the integer part of $nx$ and $e_i$ 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

$$\sum_{i=1}^{n} z_{ni} u_{ni}$$

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

$$\left( \frac{X_n}{\|x_n\|}, \ldots, \frac{X_n^K}{\|x_n^K\|} \right) = X_n \left( \text{diag} [\|x_n\|, \ldots, \|x_n^K\|] \right)^{-1}$$
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 $\{z_t; t \geq 0\}$ 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

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

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

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

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.
\[ D_n z = \sqrt{n} \sum_{i=1}^{n} z_i 1(i_i) \]

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 \( z_n \in \mathbb{R}^n \) for any natural \( n \). We say that \( \{z_n\} \) is \( L_2 \)-approximable if there exists a function \( F \in L_2 \) such that

\[
(2.1) \quad \lim_{n \to \infty} \|z_n - d_n F\| = 0
\]

(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

\[
(2.2) \quad \|D_n z\| = \left( \sum_{i} |z_i|^n n \int_{i} dx \right)^{1/2} = \|z\|, \quad z \in \mathbb{R}^n,
\]

and by the Cauchy-Schwarz inequality

\[
(2.3) \quad \|d_n F\| \leq \left( \sum_{i} n \int_{i} F^2 dx \right)^{1/2} = \|F\|, \quad n \geq 1.
\]

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

\[ P_n F = n \sum_{i} \int_{i} F(x) dx 1(i_i) \]

Therefore (2.2) and (2.3) imply

\[
(2.4) \quad \|P_n F\| \leq \|F\|, \quad F \in L_2.
\]

It is well known that
\begin{equation}
\lim_{n \to \infty} \left\| F - P_n F \right\| = 0, \quad F \in L_2
\end{equation}

(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

\begin{equation}
\lim_{n \to \infty} \left\| D_n z_n - F \right\| = 0.
\end{equation}

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

\begin{equation*}
\left\| D_n z_n - F \right\| \leq \left\| D_n (z_n - d_n F) \right\| + \left\| P_n F - F \right\| = \left\| z_n - d_n F \right\| + \left\| P_n F - F \right\| \to 0.
\end{equation*}

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

\begin{equation*}
\left\| z_n - d_n F \right\| = \left\| D_n z_n - P_n F \right\| \leq \left\| D_n z_n - F \right\| + \left\| F - P_n F \right\| \to 0.
\end{equation*}

**Property 2.** If \( \{z_n\} \) is \( L_2 \)-approximable, then

\begin{equation}
\lim_{n \to \infty} \max \left| z_n \right| = 0.
\end{equation}

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

\begin{equation*}
\max_i \left| d_n F \right| \leq \max_i \left( \int d_n F^2 \right)^{1/2} \to 0, \quad n \to \infty.
\end{equation*}

This relation and (2.1) yield

\begin{equation*}
\max_i \left| z_n \right| \leq \left\| z_n - d_n F \right\| + \max_i \left| d_n F \right| \to 0.
\end{equation*}

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

\begin{equation*}
\lim_{n \to \infty} \left( z_n^i \right) z_n^i = \int_0^1 F_i(x) F_2(x) dx.
\end{equation*}

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

\begin{equation}
\lim_{n \to \infty} \left\| z_n^i \right\| = \lim_{n \to \infty} \left\| D_n z_n^i \right\| = \left\| F_i \right\|, \quad i = 1,2.
\end{equation}

In Mynbaev (2000) it has been proved that for \( L_2 \)-generated sequences
\[
\lim_{n \to \infty} (d_n F_1) d_n F_2 = \int_0^1 F_1(x) F_2(x) dx.
\]

Using these equations and (2.1), we get

\[
\begin{align*}
&\left| z_n^2 - \int_0^1 F_1 F_2 dx \right| \leq \left| z_n^2 - d_n F_1 \right|_n \leq \left| (d_n F_1) \left( z_n^2 - d_n F_2 \right) \right| \\
&\left| d_n F_1 \right| d_n F_2 - \int_0^1 F_1 F_2 dx \leq \left| z_n^2 - d_n F_1 \right|_n \left\| z_n^2 - d_n F_2 \right\| \\
&\left| d_n F_1 \right| d_n F_2 - \int_0^1 F_1 F_2 dx \to 0.
\end{align*}
\]

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 \( \{ \psi_j \, : \, j \geq 0 \} \) of real numbers define operators \( \Psi_n : R^n \to R^n \) and \( \Phi_n : R^n \to l_2 \) by

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

Let

\[
\alpha = \sum_j |\psi_j|, \quad \beta = \sum_j |\psi_j|, \quad \gamma = \sum_j \psi_j.
\]

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

\[
(2.9) \quad \left\| \Psi_n z \right\| \leq \alpha \|z\|, \quad \left\| \Phi_n z \right\| \leq \alpha \|z\|, \quad z \in R^n, \quad n \geq 1,
\]

and that \( \beta < \infty \) implies \( \alpha < \infty \) and convergence of \( \gamma \).

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

\[
\lim_{n \to \infty} \left\| (\Psi_n - \gamma) z_n \right\| = 0, \quad \lim_{n \to \infty} \left\| \Phi_n z_n \right\| = 0.
\]
Proof: Let \( \{ z_n \} \) be \( L_2 \)-approximated by \( F \). In Mynbaev (2000) it has been proved that

\[
\lim_{n \to \infty} \| (\Psi_n - \gamma)d_n F \| = \lim_{n \to \infty} \| \Phi_n d_n F \| = 0.
\]

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

\[
\| (\Psi_n - \gamma)z_n \| \leq \| (\Psi_n - \gamma)(z_n - d_n F) \| + \| (\Psi_n - \gamma)d_n F \|
\]

\[
\leq (\alpha + |\gamma|) \| z_n - d_n F \| + \| (\Psi_n - \gamma)d_n F \| \to 0
\]

and

\[
\| \Phi_n z_n \| \leq \| \Phi_n (z_n - d_n F) \| + \| \Phi_n d_n F \| \to 0.
\]

Denote \( M_n F = D_n \Psi_n d_n F \). In Mynbaev (2000) it has been proved that

\[
\| M_n F - \gamma F \| \to 0.
\]

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 \( MF = \sum m_k c_k \exp(ikx) \) for a given sequence of numbers \( \{ m_k \} \).

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

\[
\| D_n z_n - F \| = \text{esssup}_{x \in (0,1)} \| D_n z_n \| (x) - F(x) \| \to 0.
\]

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 \( p_1, p_2, \ldots, p_n \) such that \( p_t \in i_t \) for any \( t = 1, \ldots, n \). Put \( z_{nt} = n^{-1/2} F(p_t) \), \( t = 1, \ldots, n \). Then \( \{ z_n \} \) is \( L_2 \)-approximable by \( F \).

**Proof:** Statement a) follows from the inequality

\[
\| D_n z_n - F \| \leq \| D_n z_n - F \|_{\infty}.
\]

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

\[
\| F(p_t) - F(x) \| \leq \varepsilon, \quad x \in i_t, \quad n \geq n_0.
\]
Hence,
\[
\left\| D_n z_n - F \right\|_\infty = \left\| \sum_{i=1}^n F(p_i)1(i) - F \right\| = \max_{x, i} \left| F(p_i) - F(x) \right| \leq \varepsilon, \quad n \geq n_0.
\]

It remains to apply part a).

**Proposition 1.** Consider a polynomial trend
\[
p_n = (1^{k-1}, 2^{k-1}, \ldots, n^{k-1})
\]
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
\[
\sum_{i=1}^n t^i = (1 + o(1)) \frac{n^{l+1}}{l+1}, \quad l = 1, 2, \ldots
\]

Therefore
\[
\| p_n \| = (1 + o(1)) \left( \frac{n^{2k-1}}{l(2k-1)} \right)^{1/2}
\]

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

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

Hence,
\[
D_n z_n = (1 + o(1)) \sqrt{2k-1} \sum_{i=1}^n \left( \frac{t}{n} \right)^{k-1} 1(i),
\]

wherefore
\[
\left\| D_n z_n - F \right\|_\infty = \sqrt{2k-1} \max_{1 \leq i \leq n} \max_{x, k} \left( \left( \frac{t}{n} \right)^{k-1} (1 + o(1)) - x^{k-1} \right)
\]

Since the last expression tends to zero, the statement follows from Property 5.
Consider a geometric progression

\[ g_n = (a^n, a^{n-1}, \ldots, a^1) \quad a \in \mathbb{R}. \]

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

\[
z_n = (1 + o(1))\sqrt{1 - a^2}(a^0, a^{n-1})
\]

so that

\[
D_n z_n = (1 + o(1))\sqrt{n(1 - a^2)} \sum_{i=1}^{n} a^{i-1} (i_t)
\]

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

\[
(2.10) \quad \int_\varepsilon^1 \left| D_n z_n \right|^2 dx \leq \sum_{t=t_\varepsilon}^{n} \int \left| D_n z_n \right|^2 dx = (1 + o(1))n(1 - a^2) \sum_{r=t_\varepsilon}^{n} a^{2(r-1)} \frac{1}{n}
\]

\[
\leq (1 + o(1))(1 - a^2) \sum_{r=\lceil n\varepsilon \rceil - 1}^{\infty} a^{2(r-1)} \leq ca^{2\varepsilon} \to 0.
\]

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

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

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

\[
(2.11) \quad \|F\| = 1.
\]

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

\[
(2.12) \quad D_n z_n = \sum_{i=1}^{n} (-1)^{i-1} 1(i_i).
\]

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

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

Therefore, denoting \(S_n = \bigcup_{i=[na]+1}^{[nb]+1} i_i\), we can write

\[
(2.13) \quad \|F\|_{S_n} \leq \int_{[na]/n}^{[nb]/n} Fdx + \int_{[na]+1/n}^{[nb]} Fdx + \int_{b}^{(2.12)} Fdx.
\]

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

\[
(2.14) \quad \int_{S_n} Fdx \leq \int (F - D_n z_n)dx + \int D_n z_n dx
\]

\[
\leq \|F - D_n z_n\| + 1/n \rightarrow 0
\]

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

\[
(2.15) \quad \int_{a}^{b} Fdx = 0 \text{ for any } (a, b) \subset (0,1)
\]

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)

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

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

**Proof.** Denote

\[I_k(n) = \int_1^n \ln^k x \, dx, \quad k \geq 0.\]

Obviously,

\[I_k(n) = x \ln^k x \bigg|_1^n - k \int_1^n \ln^{k-1} x \, dx = n \ln^k n - k I_{k-1}, \quad k \geq 1,\]

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

By recurrent substitution we see that there exist numbers $C_k, \ldots, C_0$ 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 $k \geq 1$

\[(2.16) \quad I_k(n) = (1 + o(1)) n \ln^k n.\]

This implies

\[(2.17) \quad I_k(n+1) = (1 + o(1)) (n+1) \ln^k (n+1)\]

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

\[= (1 + o(1)) n \ln^k n.\]
Note that

\[(2.18) \quad I_{2k}(n) \leq \sum_{t=2}^{n} \ln^{2k} t \leq \|\alpha_n\|^2 \leq I_{2k}(n+1).\]

(2.16) – (2.18) imply

\[\|\alpha_n\| = (1+o(1))\sqrt{n \ln^k n}.\]

So

\[z_n = \frac{1+o(1)}{\sqrt{n \ln^k n}} (\ln^k 1, \ldots, \ln^k n)\]

and

\[D_n z_n = \frac{1+o(1)}{\ln^k n} \sum_{i=1}^{n} |i| \ln^k t.\]

Since \(\ln t / \ln n^k \leq 1, 1 \leq t \leq n\), the difference between \(D_n z_n\) and \(f_n\) defined by

\[f_n = \frac{1}{\ln^k n} \sum_{i=1}^{n} |i| \ln^k t\]

tends to zero uniformly on \([0,1]\).

Fix \(\varepsilon \in (0,1)\). If \([\varepsilon n]+1 \leq t \leq n\), then \(\varepsilon \leq t/n \leq 1\) and there exists \(c_{i}(\varepsilon) > 0\) such that

\[|\ln(t/n)| \leq e_i \text{ for } [\varepsilon n]+1 \leq t \leq n.\]

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

\[(2.19) \quad \left|\left(\frac{\ln t}{\ln n}\right)^k - 1\right| = \left|\left(\frac{\ln n + \ln(t/n)}{\ln n}\right)^k - 1\right| \leq \varepsilon, \quad n \geq n_1(\varepsilon).\]

If \(1 \leq t \leq [\varepsilon n]\), then

\[(2.20) \quad \left|\left(\frac{\ln t}{\ln n}\right)^k - 1\right| \leq \left|\left(\frac{\ln t}{\ln n}\right)^k\right| + 1 \leq 2.\]

Obviously,

\[f_n - F = S_1 + S_2\]
where

\[ S_1 = \sum_{i=1}^{[\epsilon n]} \left( \frac{\ln t}{\ln n} \right)^k I(i), \quad S_2 = \sum_{i=[\epsilon n]+1}^{n} \left( \frac{\ln t}{\ln n} \right)^k I(i). \]

From \( \bigcup_{i=1}^{[\epsilon n]} I_i = (0, \frac{[\epsilon n]}{n}) \subset (0, \epsilon) \) and (2.20) it follows that (\( \text{mes} \) denotes the Lebesgue measure)

\[ \|S_1\| \leq 2\left( \text{mes} \left( \bigcup_{i=1}^{[\epsilon n]} I_i \right) \right)^{1/2} \leq 2\epsilon^{1/2}. \]

(2.19) implies

\[ \|S_2\| \leq \epsilon \left( \text{mes} \left( \bigcup_{i=[\epsilon n]+1}^{n} I_i \right) \right)^{1/2} \leq \epsilon. \]

Thus,

\[ \|f_n - F\| \leq \|S_1\| + \|S_2\| \leq 2\epsilon^{1/2} + \epsilon, \quad n \geq n_1(\epsilon), \]

which proves the statement.

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

(2.21)

\[ u_n = \sum_{j=0}^{[\epsilon n]} e_{nt-j} \psi_j, \quad n = 1,2,\ldots \]

where the \( \psi_j \) are the same as in Property 4. For a sequence \( \{Z_n: n > K\} \) of \( n \times K \) nonstochastic matrices with columns \( z_1^n, \ldots, z^K_n \) define random vectors

\[ Z_n^*u_n = \left( \sum_{i=1}^{n} z_i^K u_{ni} \right)^K. \]

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

\[ V = \int_0^1 F^*F dx = \left( \int_0^1 F_k(x)F_j(x) dx \right)_{k,j=1}^K. \]

**Theorem 1.** Suppose that
A) \( E(e_n^2 | G_{n,t-1}) = \sigma^2 \) for some \( \sigma > 0 \) and all \( t, n \),
B) \( e_n^4 \) are uniformly integrable,
C) the sequence \( \{e_n^k\} \) 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) \( \lim_{n \to \infty} Z_n^\prime Z_n = V \).

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 \( u_n \) with deterministic weights \( z_n \). 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

\[
\sup_{t,n} E e_n^4 < \infty.
\]

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

\[
\| x_n \| \sqrt{n} f_n D_n z_n - F \| \xrightarrow{a.s.} 0.
\]

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) \( Y_n = \text{diag} \|x_n^1\|, ..., \|x_n^K\| \), \( Z_n = X_n Y_n^{-1} \)

From (1.1) and (1.2) it is easy to obtain
Application of Theorem 1 immediately leads to the following result.

**Theorem 2.** Let \( e_n, z_n^k \), and \( \psi_j \) satisfy assumptions of Theorem 1. Then

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

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

\[
\psi_0 = 1, \; \psi_j = 0, \; j \geq 1,
\]

we are taken back to Anderson’s (1971) result. He has imposed conditions (2.7) and (2.23) with \( \text{det} \; V \neq 0 \) (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 \( \text{det} V \neq 0 \) 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 of 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 \( Y_n = \text{diag}[y_{1n}, \ldots, y_{nk}] \) with positive elements on the main diagonal and put \( Z_n = X_n \bar{Y}_n^{-1} \). We say that \( \{\bar{Y}_n\} \) is a conventional-scheme-compliant (CSC) normalizer if

\[
(3.5) \quad \text{there exists } \lim_{n \to \infty} Z_n Z_n^T = \bar{V}, \quad \det \bar{V} \neq 0,
\]

and

\[
(3.6) \quad \begin{cases} 
\mathcal{Z}_n u_n \xrightarrow{d} N(0, (\sigma^2) V) \text{ for any } e_n \text{ and } \psi_j \\
\text{satisfying conditions A), B), E) of Theorem 1}
\end{cases}
\]

The columns of \( Z_n \) are not required to be \( L_2 \)-approximable in this definition.

**Proposition 4.** If \( \{\bar{Y}_n\} \) is a CSC normalizer and \( \{\Delta_n\} \) is a sequence of \( K \times K \) diagonal matrices with positive elements such that

\[
(3.7) \quad \text{there exists } \lim_{n \to \infty} \Delta_n = \Delta, \quad \det \Delta \neq 0,
\]

then \( \Delta_n \bar{Y}_n \) is also a CSC normalizer.

**Proof.** From (3.5) and (3.7)

\[
\lim(X_n^* (\Delta_n \bar{Y}_n)^{-1} X_n (\Delta_n \bar{Y}_n)^{-1} = \lim(X_n^* \bar{Y}_n^{-1} \Delta_n^{-1} X_n \bar{Y}_n^{-1} \Delta_n^{-1} = \\
\lim \Delta_n^{-1} Z_n Z_n^T \Delta_n^{-1} = \Delta^{-1} \bar{V} \Delta^{-1}, \quad \det \Delta^{-1} \bar{V} \Delta^{-1} \neq 0
\]

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

\[
(X_n^* (\Delta_n \bar{Y}_n)^{-1} u_n = (Z_n^T \Delta_n u_n)^{-1} \xrightarrow{d} N(0, (\sigma^2) \Delta^{-1} \bar{V} \Delta^{-1})
\]

for any \( e_n \) and \( \psi_j \) satisfying conditions A), B), E) of Theorem 1. Hence, \( \Delta_n \bar{Y}_n \) is a CSC normalizer.

**Proposition 5.** If \( \bar{Y}_n \) 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 \( Y_n = \Delta_n \bar{Y}_n \).

**Proof.** Denote \( y_n^k = \|x_n^k\| \), \( \bar{y}_n^k \), \( \bar{v}_k \) the diagonal elements of \( Y_n \), \( \bar{Y}_n \), and \( \bar{V} \), respectively. The main diagonal of the limit relation in (3.5) gives
\[(\mathbf{z}_n^k) \mathbf{z}_n^k = \| \mathbf{z}_n^k \|^2 / (\mathbf{y}_n^k)^2 \rightarrow \mathbf{v}_{ik}, \]

that is \( y_n^k / \mathbf{y}_n^k \rightarrow (\mathbf{v}_{ik})^{1/2} \). In matrix notation this means that \( Y_n \mathbf{v}_n^{-1} \rightarrow \Delta \) where

\[
\Delta = \text{diag}[(\mathbf{v}_{i1})^{1/2}, \ldots, (\mathbf{v}_{ik})^{1/2}], \quad \det \Delta \neq 0
\]

Denoting \( \Delta_n^s = Y_n \mathbf{v}_n^{-1} \), we see that (3.7) is true, \( Y_n = \Delta_n^s \mathbf{v}_n^{-1} \), 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

\[
\varphi_n = \frac{R' \hat{\mathbf{b}}_n - r}{\sqrt{s^2 R'(X_n'X_n)^{-1} R}}
\]

used to test \( H_0: R'\hat{\mathbf{b}} = r \) against the alternative \( H_a: R'\hat{\mathbf{b}} \neq r \). Here the vector \( R = (R_1, \ldots, R_K)' \) and the real number \( r \) are given and \( s^2 \) is the estimator of \( \sigma^2 \),

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

(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

\[
\rho_n = Y_n^{-1} R, \quad h_n = (Z_n'Z_n)^{-1/2}, \quad f_n = h_n \rho_n / \| h_n \rho_n \|
\]

and using the null hypothesis, we have

\[
\varphi_n = \frac{\left( Y_n^{-1} R \right) Y_n \left( \hat{\mathbf{b}}_n - \mathbf{b} \right)}{\sqrt{s^2 (Y_n^{-1} R)(Z_n'Z_n)^{-1} Y_n^{-1} R}} = \\
\frac{\rho_n (Z_n'Z_n)^{-1} Z_n' e_n}{\sqrt{s^2 \rho_n (Z_n'Z_n)^{-1} \rho_n}} = \frac{\left( h_n \rho_n \right) h_n Z_n' e_n}{\sqrt{s^2 \left( h_n \rho_n \right) h_n \rho_n}} = \frac{1}{s} f_n Z_n' e_n.
\]

(3.8)

By Theorem 1 \( Z_n' e_n \) converges in distribution. Assuming that
\( \lim h_n = V^{-1/2} \)

and

\( \lim f_n = f, \)

we can pass to the limit in (3.8) (using also \( \text{plims} = \sigma \) 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, \( \tau_n \) can be \( \sqrt{n} \)). In place of (3.8) we can write

\[ \Phi_n = \frac{R(Z_n \tilde{Z}_n)^{-1} \tilde{Z}_n e_n}{\sqrt{s^2 R'(Z_n \tilde{Z}_n)^{-1} R}}. \]

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 \( x_n^1, \ldots, x_n^k \) be polynomial trends of degrees 0, \ldots, \( K-1 \), respectively. Instead of normalizing \( X_n \) by the canonical normalizer

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

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

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

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 \( \psi_j \) satisfy A), B), E), then by Theorem 2 (3.3) is true with

\[
V = \begin{pmatrix}
1 & 1/2 & \ldots & 1/K \\
1/2 & 1/3 & \ldots & 1/(K+1) \\
\vdots & \vdots & \ddots & \vdots \\
1/K & 1/(K+1) & \ldots & 1/(2K+1)
\end{pmatrix}
\]

\( V \) 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).

**Example 2.** If \( K > 1 \), and

\[
x_n^k = (\ln^1 1, \ldots, \ln^k n), \quad k = 1, \ldots, K,
\]

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 \( F_k \equiv 1, \quad k = 1, \ldots, K \). 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

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

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

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

where both factors at the right converge.
Theorem 3. Under the conditions of Theorem 1 one has

\[(3.12) \quad D_n (\hat{y}_n - X_n \hat{\beta}) \overset{d}{\rightarrow} \xi_1 F_1 + \ldots + \xi_K F_K\]

where \(\xi \sim N(0, (\sigma^2 V))\) (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 \(\xi\) 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
