Asymptotic properties of OLS estimates in autoregressions with bounded or slowly growing deterministic trends

Mynbaev, Kairat

Kazakhstan Institute of Management, Economics, and Strategic Research (KIMEP)

2003

ASYMPTOTIC PROPERTIES OF OLS ESTIMATES IN AUTOREGRESSIONS WITH BOUNDED OR SLOWLY GROWING DETERMINISTIC TRENDS

Kairat T. Mynbaev
Department of Economics
KIMEP
Abai Ave. 4, 480100 Almaty, Kazakhstan
kairat@kimep.kz

Key Words: autoregression; deterministic trend; OLS estimator asymptotics.

ABSTRACT

We propose a general method of modeling deterministic trends for autoregressions. The method relies on the notion of L_2-approximable regressors previously developed by the author. Some facts from the theory of functions play an important role in the proof. In its present form, the method encompasses slowly growing regressors, such as logarithmic trends, and leaves open the case of polynomial trends.

1. INTRODUCTION

Consider the following autoregressive model:

$$y_i = \beta_1 t_i + \beta_2 y_{i-1} + e_i, \quad i = 1, \dots, n, \tag{1.1}$$

where the parameters β_1 and β_2 are to be estimated by Ordinary Least Squares (OLS). The regressor $t = (t_1, \dots, t_n)'$ is assumed to be nonstochastic (in applications it is often a time trend); the coefficient β_2 satisfies the stability condition $|\beta_2| < 1$; the errors e_i are martingale differences satisfying certain second- and fourth-order conditional moment restrictions (in particular, the errors can be normal independent identically distributed (i.i.d.) with mean zero and variance σ^2). Denote $\beta = (\beta_1, \beta_2)'$ and let $\hat{\beta}$ be the OLS estimator of β based on a sample of size n. The logarithmic trend

$$t_i = \ln i, \quad i = 1, \dots, n, \tag{1.2}$$

and polynomial trend

$$t_i = i^k, \quad i = 1, \ldots, n, \tag{1.3}$$

are examples of growing trends (here k is some natural number). The most recent papers about models with growing trends include Ng and Vogelsang (2002), Sibbertsen (2001), and Rahbek *et al.* (1999). Bounded trends are also interesting for modeling seasonal variations (see Nabeya (2000) and Tam and Reinsel (1998)). Leonenko and Šilac-Benšić (1997) treat the continuous case and the stress is on the singular errors.

The abundance of papers about models with particular types of trends testifies to the continuing interest in deterministic trends and calls for a general method that would be applicable to all types. One such method in a setup different from ours has been developed by Andrews and McDermott (1995). We pursue an approach based on the notion of L_2-approximable regressors introduced in Mynbaev (2001) (a narrower notion of L_2-generated regressors has been suggested in Moussatat (1976)). Specifically, our purpose is to find the asymptotic distribution of $\hat{\beta}$, as $n \to \infty$, when the normalized exogenous regressor is L_2-approximable. Mynbaev and Castelar (2001) have shown that the last condition holds true for (1.2) and (1.3). In the same paper it is proved that normalization of the geometric progression $x_n = (a^0, a^1, \ldots, a^{n-1})$, where $a \neq 1$ is real, and the exponential trend $x_n = (e^a, \ldots, e^{na})$, where $a \neq 0$ is real, does not lead to L_2-approximable sequences. This is because both the geometric progression and exponential trend are too concentrated at one end of their domain, while L_2-approximability implies some "smearing" over the domain. It is well known that regressing on the geometric progression or exponential trend leads to bad asymptotic properties for the OLS estimator.

When there are no autoregressive terms, the solution to this problem does not require the L_2-approximability assumption, is relatively simple and given by Anderson (1971), Theorem 2.6.1. For the case $\beta_2 \neq 0$ and $|\beta_2| < 1$, the most advanced result, including stochastic t, is contained in Anderson and Kunitomo (1992). However, that result does not cover growing regressors like (1.2) and (1.3). Sims, Stock, and Watson (1990), in order to find the asymptotics of $\hat{\beta}$ in the case of a simple linear trend, found the asymptotics for a transformed regression. This method is not feasible because the transformation involves unknown parameters. The exposition of their approach can also be found in Hamilton (1994) (see Chapter 16). The feasibility problem does not exist in our case since we just normalize the exogenous variables.

To explain the nature of difficulties arising in case (1.3), we need to review the way the OLS asymptotics is usually derived. Let us write the linear model in the form

$$y = X\beta + e \tag{1.4}$$

where X is a $n \times k$ matrix of linearly independent regressors, β is a $k \times 1$ parameter vector to be estimated, and e is an error vector. The OLS estimator for (1.4) is

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'e\ .$$

By transferring β to the left side and premultiplying the resulting equation by a nondegenerate diagonal matrix M we obtain

$$M(\hat{\beta} - \beta) = [(XM^{-1})'XM^{-1}]^{-1}(XM^{-1})'e = (H'H)^{-1}H'e \tag{1.5}$$

where $H = XM^{-1}$. The *conventional scheme* of deriving the asymptotics of $\hat{\beta}$ consists in choosing the matrix M in such a way that the matrix $Q = H'H$ converges in probability to a nondegenerate matrix Q_∞ and the factor $w = H'e$ converges in distribution to a normal vector w_∞. Then it immediately follows that $M(\hat{\beta}-\beta)$ converges in distribution to a normal vector. The matrix M is called a normalizer. Usually, Q_∞ is the variance of w_∞.

An obvious problem is that of choosing M. When $\beta_1 = 0$ and $\beta_2 \neq 0, |\beta_2| < 1$, the standard choice is $M = \sqrt{n}$. When $\beta_2 = 0$ and $\beta_1 \neq 0$, Anderson (1971) suggested to put $M = (\sum_{i=1}^n t_i^2)^{1/2}$. These two facts helped us to come up with the normalizer in Theorem 2.1 below.

Another problem is that when the exogenous regressor grows quickly (like a polynomial trend), the vector $H'e$ converges in distribution to a degenerate normal vector, whose second coordinate is proportional to the first one. For this reason the limit of $H'H$ is degenerate in case (1.3). In this case we have proved convergence of w and Q but not $M(\hat{\beta}-\beta)$. The idea of the method is explained in the paragraph preceding Lemma 2.1. The proof is pretty involved. It relies on properties of L_2-approximable sequences established in Mynbaev (2001) as well as on a martingale Weak Law of Large Numbers (WLLN) by Chow (1971) and Davidson (1994), mixingale WLLN due to Andrews (1988) and Davidson (1994), the McLeish (1974) Central Limit Theorem (CLT), and Burkholder's (1973) theorem on transforms of martingales. All these results, for the reader's convenience, are gathered in the Appendix. The main result is stated as Theorem 2.1 in Section 2.

The author hopes to consider elsewhere the model with q deterministic exogenous regressors and p lags of the dependent variable

$$y_i = \sum_{j=1}^{q} \beta_j^1 t_{ji} + \sum_{j=1}^{p} \beta_j^2 y_{i-j} + e_i.$$

This is why the intercept term is not included in (1.1): the intercept would be just another L_2-approximable regressor, and its inclusion, within the framework suggested, would not be any easier than considering more trends. The exogenous regressors will be required to satisfy the L_2-approximability condition (see assumption A2) below).

The L_2-approximability notion was applied in Mynbaev (2001) to find a limiting distribution of quadratic forms of random variables, in Mynbaev (1997) to find the asymptotics of the fitted value for a linear regression with nonstochastic regressors, and in Mynbaev (2003) to prove a CLT applicable to an SUR-type system of linear regressions without autoregressive terms. In response to referee's question, I am pretty confident that this notion can be applied to nonstationary models (with unit roots). One way this would be possible to do is by proving an invariance principle parallel to the central limit theorem contained in Mynbaev (2001).

2. MAIN RESULT

If (Ω, μ) is a probability space with measure μ, then $L_p(\Omega, \mu)$ denotes the set of measurable functions $F : \Omega \to R$ provided with the norm $\|f\|_p = (\int_\Omega |f(x)|^p d\mu(x))^{1/p}, \quad 1 \le p < \infty$. When $\Omega = (0, 1)$ and μ is the Lebesgue measure, we write L_2 instead of $L_2((0, 1), \mu)$ and $\|f\|$ instead of $\|f\|_2$. The space ℓ_2, a discrete analog of L_2, consists of sequences $\{z_j : j \in J\}$

with a finite norm $\|z\| = (\sum_{j\in J} |z_j|^2)^{1/2}$; the set of indices J depends on the context. R^n is the Euclidean space provided with this norm. plim (dlim) means a limit in probability (in distribution, respectively). $N(m, V)$ denotes the set of normal vectors with mean m and a matrix variance V.

The discretization operator $d_n : L_2 \to R^n$ is defined as follows. For a function $f \in L_2$, the vector $d_n f \in R^n$ has components

$$(d_n f)_j = \sqrt{n} \int_{i_j} f(x)dx, \quad j = 1, \ldots, n,$$

where the intervals $i_j = ((j-1)/n, j/n)$ form a partition of $(0, 1)$. The sequence $\{d_n f : \; n = 1, 2, \ldots\}$ is called L_2-generated by f. The notion of L_2-generated sequences was introduced by Moussatat (1976). A sequence $\{u_n : \; n = 1, 2, \ldots\}$, where $u_n \in R^n$ for each n, is called L_2-approximable, if there exists a function $f \in L_2$ such that $\|u_n - d_n f\| \to 0, \quad n \to \infty$. Besides, in this case $\{u_n\}$ is called L_2-approximated by f. L_2-approximable sequences have been introduced and studied by Mynbaev (2001). In statistics often sequences of vectors with an increasing number of coordinates are used. Conditions on such sequences imposed in terms of limits of different expressions involving them look awkward and are difficult to check. The idea behind L_2-approximability is to approximate sequences with functions of a continuous argument and then derive (instead of imposing) the required properties of sequences from properties of functions. This is facilitated by the fact that the theory of L_2 spaces and operators in them is well developed. A comparison of properties of L_2-approximable sequences contained in Mynbaev (2001) with those imposed directly in, say, Anderson (1971) shows that not very much is lost in terms of generality.

Before we state the main result we need to do a little housekeeping. We assume that in (1.4)

$$y = (y_1, \ldots, y_n)', \quad X = (x_1, x_2), \quad x_1 = (t_1, \ldots, t_n)',$$

$$x_2 = (y_0, \ldots, y_{n-1})', \quad e = (e_{n1}, \ldots, e_{nn})',$$

where $\{e_{ni}, F_{ni}\}_{i=1}^n$ is a martingale difference (m.d.) sequence for each n, that is, F_{ni} are σ-fields such that $F_{n1} \subset \ldots \subset F_{nn}$ and $E(e_{ni}|F_{n,i-1}) = 0$.

Now we state and discuss the main assumptions.

A1) $\beta_1\beta_2 \neq 0, \quad |\beta_2| < 1.$

The cases $\beta_1 = 0$ and $\beta_2 = 0$ are excluded as known (see Anderson (1971) and Hamilton (1994)).

A2) $\|x_1\| > 0$ for all large n and the sequence $u_n = x_1/\|x_1\| = t/\|t\|$ is L_2-approximable.

Mynbaev and Castelar (2001) have shown that if $u_n = t/\|t\|$, where t is defined by (1.2) or (1.3), then u_n is L_2-approximable. See Theorem 3.1 and Lemma 2.1 about implications of L_2-approximability.

A3) The initial condition y_0 is a square-integrable random variable.

As usual, the influence of y_0 is asymptotically negligible.

A4) $\{e_{ni}, F_{ni}\}$ is a p-integrable m.d. sequence such that $\sup_{n,i} \|e_{ni}\|_p < \infty$ for some $p > 4$ and

$$E(e_{ni}^2|F_{n,i-1}) = \sigma^2 \quad \forall n, i,$$

where $0 < \sigma^2 < \infty$, and with some $\sigma_1^2 > \sigma^4$ and $c > 0$

$$E(e_{ni}^4|F_{n,i-1}) = \sigma_1^2, \quad E(|e_{ni}^2 - \sigma^2||F_{n,i-1}) \geq c \quad \forall n, i.$$

For example, if $\{e_i\}$ is i.i.d. normal, then

$$E(e_i^4|F_{i-1}) = Ee_i^4 = 3\sigma^4 > \sigma^4, \quad E(|e_i^2 - \sigma^2||F_{i-1}) = \|e_i^2 - \sigma^2\|_1 = c > 0.$$

A5) The limit

$$\lambda = \lim_{n\to\infty} \frac{\sqrt{n}}{\|t\|} = \lim_{n\to\infty} \frac{\sqrt{n}}{\|x_1\|} \in [0, \infty]$$

exists.

The limit λ measures the relative magnitude of the error term and the regressor t. When t is a polynomial with $k > 0$, one has $\lambda = 0$. If t is a logarithmic trend, then $0 < \lambda < \infty$. Since $\lambda = \infty$ is admitted, in the formulas that follow we put $1/\infty = 0, \quad \infty/\infty = 1$. For L_2-approximable normalized regressors we find a general answer, which covers (1.2) but not (1.3). If the regressor grows quickly relative to the error, then $\lambda = 0$, which, in turn, renders degenerate the matrix Q_∞ from (2.4). In the latter case we suggest a conjecture for profession's discussion.

To state the main result, we need to define the elements of the conventional scheme. Let

$$m_1 = \|x_1\|,\ m_2 = \|x_1\| + \sqrt{n},\ M = \mathrm{diag}[m_1, m_2].$$

With this M, the matrix $H = XM^{-1}$ from (1.5) has the vectors

$$h_1 = x_1/\|x_1\|,\ h_2 = x_2/(\|x_1\| + \sqrt{n}) \tag{2.1}$$

as its columns: $H = (h_1, h_2)$. Therefore in (1.5)

$$Q = H'H = \begin{pmatrix} h_1' \\ h_2' \end{pmatrix} (h_1 h_2) = \begin{pmatrix} h_1'h_1 & h_1'h_2 \\ h_2'h_1 & h_2'h_2 \end{pmatrix} = \begin{pmatrix} 1 & h_1'h_2 \\ h_2'h_1 & \|h_2\|^2 \end{pmatrix} \tag{2.2}$$

and

$$w = H'e = \begin{pmatrix} h_1'e \\ h_2'e \end{pmatrix}. \tag{2.3}$$

Denote

$$\gamma = \frac{\beta_1}{(1+\lambda)(1-\beta_2)},\quad Q_\infty = \begin{pmatrix} 1 & \gamma \\ \gamma & \gamma^2 + (\frac{\sigma\lambda}{1+\lambda})^2 \frac{1}{1-\beta_2^2} \end{pmatrix}. \tag{2.4}$$

Obviously, $\det Q_\infty = 0$ if and only if $\lambda = 0$.

Theorem 2.1. Under assumptions A1) through A5), one has

$$w_\infty \equiv \mathrm{dlim}\ w \in N(0, \sigma^2 Q_\infty),\quad \mathrm{plim}\ Q = Q_\infty. \tag{2.5}$$

Hence, if $\lambda > 0$, then $\mathrm{dlim}\ M(\hat\beta - \beta) \in N(0, \sigma^2 Q_\infty^{-1})$.

From the point of view of this theorem, the case $\lambda = 0$ presents a problem. There are reasons to believe that the following is true.

Conjecture. If one chooses $M = |\det Q|^{1/2}\mathrm{diag}\,[m_1, m_2]$ in case $\lambda = 0$, then $M(\hat\beta - \beta)$ will converge in distribution to a vector w_∞ such that $w_{\infty 1} = \gamma w_{\infty 2}$.

By the Cramér-Wold theorem, to prove convergence of w in distribution to an element of $N(0, \sigma^2 Q_\infty)$, it is sufficient to prove, for any vector $a \in R^2$, convergence of $a'w$ to an element of $N(0, \sigma^2 a' Q_\infty a)$. The last problem will be reduced to another one, using the fact that the influence of the initial condition y_0 is negligible. Replacing e_i by e_{ni} in (1.1), by induction we obtain the solution

$$y_i = \sum_{k=1}^{i} \beta_2^{i-k}(\beta_1 x_{1k} + e_{nk}) + \beta_2^i y_0, \quad 1 \le i \le n. \tag{2.6}$$

Using (2.1), (2.3), and (2.6), rearrange $a'w$ as follows

$$a'w = a_1 h_1' e + a_2 h_2' e = \sum_{i=2}^{n}(a_1 h_{1i} + \frac{a_2}{m_2} y_{i-1}) e_{ni} + (a_1 h_{11} + \frac{a_2}{m_2} y_0) e_{n1} =$$

$$= \sum_{i=2}^{n}\Big\{a_1 h_{1i} + \frac{a_2}{m_2}\Big[\sum_{k=1}^{i-1} \beta_2^{i-1-k}(\beta_1 x_{1k} + e_{nk}) + \beta_2^{i-1} y_0\Big]\Big\} e_{ni} + (a_1 h_{11} + \frac{a_2}{m_2} y_0) e_{n1} =$$

$$= \sum_{i=2}^{n} Y_{ni} + Z_n,$$

where we put

$$Y_{ni} = \Big[a_1 h_{1i} + \frac{a_2}{m_2} \sum_{k=1}^{i-1} \beta_2^{i-1-k}(\beta_1 x_{1k} + e_{nk})\Big] e_{ni}, \quad Z_n = \frac{a_2}{m_2} \sum_{i=1}^{n} \beta_2^{i-1} y_0 e_{ni} + a_1 h_{11} e_{n1}.$$

Using conditions A1) through A4) and the fact that $m_2 \to \infty, n \to \infty$, we have by Hölder's inequality

$$\|Z_n\|_1 \le \frac{|a_2|}{m_2} \sum_{i=1}^{n} |\beta_2|^{i-1} \|y_0\|_2 \|e_{ni}\|_2 + |a_1 h_{11}| \|e_{n1}\|_1 \le \frac{c}{m_2} \sum_{i=0}^{\infty} |\beta_2|^i + c|h_{11}| \to 0.$$

Here $h_{11} \to 0$ by Theorem 3.1b). Hence, plim $Z_n = 0$ and

$$\text{dlim } a'w = \text{dlim } \sum_{i=2}^{n} Y_{ni}. \tag{2.7}$$

Next we derive the main representation of Y_{ni}. Decompose it as

$$Y_{ni} = \Big(a_1 h_{1i} + \frac{a_2}{m_2} \|x_1\| \beta_1 \sum_{k=1}^{i-1} \beta_2^{i-1-k} \frac{x_{1k}}{\|x_1\|}\Big) e_{ni} + \frac{a_2}{m_2} \sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk} e_{ni} = A_{ni} + B_{ni}, \ \ i \ge 2, \tag{2.8}$$

where we put

$$A_{ni} = \Big(a_1 h_{1i} + \frac{\beta_1 a_2}{1+\lambda_n} \sum_{k=1}^{i-1} \beta_2^{i-1-k} h_{1k}\Big) e_{ni}; \quad \lambda_n = \frac{\sqrt{n}}{\|x_1\|}; \quad B_{ni} = \frac{a_2}{m_2}\Big(\sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk}\Big) e_{ni}. \tag{2.9}$$

In definition (3.3) put $\psi_j = 0, \ j \ge 1; \ \psi_j = \beta_2^{-j}, \ j \le 0$. Then (see also (3.1))

$$\Psi_n z = \Big(\sum_{k=1}^{i} \beta_2^{i-k} z_k\Big)_{i=1}^{n}; \ \ \sum_{k=1}^{i-1} \beta_2^{i-1-k} z_k = (L_n \Psi_n z)_i, \quad i \ge 2. \tag{2.10}$$

With the notation

$$\mu_n = \frac{\beta_1 a_2}{1+\lambda_n},\ \nu_n = \frac{a_2}{m_2},\ u_n = h_1,\ g_n = a_1 u_n + \mu_n L_n \Psi_n u_n \tag{2.11}$$

we see that

$$a_1 h_{1i} + \frac{\beta_1 a_2}{1+\lambda_n} \sum_{k=1}^{i-1} \beta_2^{i-1-k} h_{1k} = (a_1 u_n + \mu_n L_n \Psi_n u_n)_i = g_{ni}. \tag{2.12}$$

h_1 is denoted by u_n because of its special role in the proof. Thus, we have representation (2.8) of Y_{ni} in terms of variables

$$A_{ni} = g_{ni} e_{ni},\quad B_{ni} = \nu_n (L_n \Psi_n e)_i e_{ni},\quad i \geq 2.$$

Besides, if we denote $\mu = \frac{\beta_1 a_2}{1+\lambda}$, then from A5) we get

$$\lim \lambda_n = \lambda,\ \lim \mu_n = \mu,\ \lim n\nu_n^2 = \Big(\frac{a_2 \lambda}{1+\lambda}\Big)^2 \tag{2.13}$$

for all $0 \leq \lambda \leq \infty$.

Now we are in a position to outline the idea of the proof of convergence of $\sum Y_{ni}$. According to the McLeish CLT (Theorem 3.4), we need to consider $\sum EY_{ni}^2$. (2.8) and (2.9) imply $Y_{ni}^2 = A_{ni}^2 + 2A_{ni}B_{ni} + B_{ni}^2$ where

$$A_{ni}^2 = g_{ni}^2 e_{ni}^2, \tag{2.14$'$}$$

$$B_{ni}^2 = \nu_n^2 \Big(\sum_{k=1}^{i-1} \beta_2^{2(i-1-k)} e_{nk}^2 + 2 \sum_{1 \leq k < l \leq i-1} \beta_2^{2i-2-k-l} e_{nk} e_{nl} \Big) e_{ni}^2, \tag{2.14$''$}$$

$$A_{ni} B_{ni} = g_{ni} \nu_n \Big(\sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk} \Big) e_{ni}^2. \tag{2.14$'''$}$$

The sum $\sum A_{ni}^2$ is responsible mainly for the contribution of the exogenous regressor; $\sum B_{ni}^2$ accounts for the contribution of the autoregressive term, and $\sum A_{ni} B_{ni}$ controls interaction between the two. Each of these three sums needs separate treatment. Before doing that we gather in one lemma various implications of Theorem 3.1.

Lemma 2.1. Under assumptions A1), A2), and A5) the following is true.

a) For any $a \in R^2$ and $\lambda \in [0, \infty]$ (see (2.4) for the notation of γ)

$$\lim_{n\to\infty} \|a_1 u_n + \mu_n L_n \Psi_n u_n\| = |a_1 + a_2 \gamma|.$$

b) The constants $c_{ni} = g_{ni}^2$, $2 \le i \le n$, (see (2.11)) satisfy conditions (b) and (c) of Theorem 3.2 and

$$\max_i c_{ni} \to 0. \tag{2.15}$$

c) The constants $c_{ni} = \nu_n^2$ satisfy conditions (b) and (c) of Theorem 3.2.

d) $\lim_{n\to\infty} u_n'(\mu_n L_n \Psi_n u_n) = \gamma$.

Proof.

a) Theorem 3.1 (part b)), identity (3.4), assumptions A1) and A2), and the choice of ψ_j imply

$$\|L_n\Psi_n u_n - \Psi_n L_n u_n\| \le \max_{1\le j\le n} |u_{nj}| \Big[\sum_j \psi_j^2 + \Big(\sum_j |\psi_j|\Big)^2\Big]^{1/2} \to 0.$$

Hence, by Theorem 3.1, parts a), c), and d), we have

$$\Big\|\sum \psi_j u_n - L_n\Psi_n u_n\Big\| \le \Big\|\Big(\sum \psi_j - \Psi_n\Big)u_n\Big\| + \|\Psi_n(u_n - L_n u_n)\| +$$

$$+ \|\Psi_n L_n u_n - L_n \Psi_n u_n\| \to 0.$$

Now using normalization $\|u_n\| = 1$, the identity $\sum_j \psi_j = 1/(1-\beta_2)$, (3.2), Theorem 3.1a) and (2.13), we obtain the desired result:

$$\Big|\|a_1 u_n + \mu_n L_n\Psi_n u_n\| - |a_1 + a_2\gamma|\Big| = \Big|\|a_1 u_n + \mu_n L_n \Psi_n u_n\| - \Big\|\Big(a_1 + \mu \sum \psi_j\Big)u_n\Big\|\Big| \le$$

$$\le \Big\|\mu_n L_n \Psi_n u_n - \mu \sum \psi_j u_n\Big\| \le |\mu_n - \mu| \|L_n\Psi_n u_n\| + |\mu| \Big\|L_n\Psi_n u_n - \sum \psi_j u_n\Big\| \le$$

$$\le |\mu_n - \mu|\alpha_\psi \|u_n\| + |\mu| \Big\|\sum \psi_j u_n - L_n\Psi_n u_n\Big\| \to 0.$$

b) From (3.2), Theorem 3.1a), normalization of u_n and (2.13), we see that condition (b) of Theorem 3.2 is satisfied:

$$\limsup_{n\to\infty} \sum_{i=2}^n c_{ni} \le \limsup_{n\to\infty} \|a_1 u_n + \mu_n L_n \Psi_n u\|^2 \le \tag{2.16}$$

$$\le \limsup_{n\to\infty} [|a_1| \|u_n\| + |\mu_n| \alpha_\psi \|u_n\|]^2 < \infty.$$

Further, (2.15) follows from (2.13), assumption A1), and Theorem 3.1b):

$$\max_i c_{ni} = \max_i \Big(a_1 u_{ni} + \mu_n \sum_{k=1}^{i-1} \beta_2^{i-1-k} u_{nk}\Big)_i^2 \le c(\max_i |u_{ni}|)^2 \to 0.$$

This bound and (2.16) imply condition (c) of Theorem 3.2:

$$\lim_{n\to\infty}\sum_{i=2}^{n}c_{ni}^2\le\lim_{n\to\infty}\max_{2\le j\le n}c_{nj}\sum_{i=2}^{n}c_{ni}=0.$$

c) Since $c_{ni}=\nu_n^2\le c/n$, we do not need to use Theorem 3.1:

$$\limsup_{n\to\infty}\sum_{i=2}^{n}c_{ni}\le c\sup_{n}\sum_{i=2}^{n}1/n<\infty,$$

$$\limsup_{n\to\infty}\sum_{i=2}^{n}c_{ni}^2\le c^2\lim_{n\to\infty}\sum_{i=2}^{n}1/n^2=0.$$

d) Choosing $a_1=-\gamma,\ a_2=1$ in property a) above, we get by normalization of u_n

$$|u_n'(\mu_nL_n\Psi_nu_n)-\gamma|=|u_n'(\mu_nL_n\Psi_nu_n-\gamma u_n)|\le\|u_n\|\|\mu_nL_n\Psi_nu_n-\gamma u_n\|\to 0.$$

The proof is complete.

In order to apply the McLeish CLT, we need to normalize Y_{ni} by Σ_n, which is defined by

$$\Sigma_n=\Big(\sum_{i=2}^{n}EY_{ni}^2\Big)^{1/2},$$

and study the asymptotical behavior of Σ_n. From now on we assume that all conditions A1)-A5) hold.

Lemma 2.2. With notation (2.11) one has

$$EY_{ni}^2=\sigma^2g_{ni}^2+\nu_n^2\sigma^4\frac{1-\beta_2^{2(i-1)}}{1-\beta_2^2}, \tag{2.17}$$

$$\Sigma_n^2=\sigma^2(\|g_n\|^2-g_{n1}^2)+\frac{\nu_n^2\sigma^4}{1-\beta_2^2}\Big(n-\frac{1-\beta_2^{2n}}{1-\beta_2^2}\Big), \tag{2.18}$$

$$\lim_{n\to\infty}\Sigma_n^2=\sigma^2a'Q_\infty a=\sigma^2\Big[(a_1+\gamma a_2)^2+\Big(\frac{a_2\sigma\lambda}{1+\lambda}\Big)^2\frac{1}{1-\beta_2^2}\Big]. \tag{2.19}$$

Proof. Assumption A4) and identities $(2.14'),(2.14'')$, and $(2.14''')$ imply by the Law of Iterated Expectations (LIE)

$$EA_{ni}^2=\sigma^2g_{ni}^2, \tag{2.20$'$}$$

$$EB_{ni}^2=\sigma^4\nu_n^2\sum_{k=1}^{i-1}\beta_2^{2(i-1-k)}=\sigma^4\nu_n^2\frac{1-\beta_2^{2(i-1)}}{1-\beta_2^2}, \tag{2.20$''$}$$

$$EA_{ni}B_{ni} = 0. \tag{2.20'''}$$

These equations immediately yield (2.17). Hence, (2.18) follows:

$$\Sigma_n^2 = \sigma^2 \sum_{i=2}^n g_{ni}^2 + \frac{\sigma^4 \nu_n^2}{1-\beta_2^2} \sum_{i=2}^n (1-\beta_2^{2(i-1)}) = \sigma^2(\|g_n\|^2 - g_{n1}^2) + \frac{\sigma^4\nu_n^2}{1-\beta_2^2}\Big(n - \frac{1-\beta_2^{2n}}{1-\beta_2^2}\Big).$$

Since $g_{n1} \to 0$ by (2.15), (2.19) follows from the last equation, Lemma 2.1a), and the last equation in (2.13). The proof is finished.

From Y_{ni} we pass to normalized variables $X_{ni} \equiv Y_{ni}/\Sigma_n$. The objective of the next three lemmas is to show that

$$q_n(X) \equiv \sum_{i=2}^n X_{ni}^2 - \sum_{i=2}^n EX_{ni}^2 = \Sigma_n^{-2} \sum_{i=2}^n (Y_{ni}^2 - EY_{ni}^2) \xrightarrow{p} 0.$$

By (2.20''')

$$Y_{ni}^2 - EY_{ni}^2 = (A_{ni}^2 - EA_{ni}^2) + (B_{ni}^2 - EB_{ni}^2) + 2A_{ni}B_{ni}.$$

Lemma 2.3. plim $\sum_{i=2}^n (A_{ni}^2 - EA_{ni}^2) = 0$.

Proof. The constants c_{ni} from Lemma 2.1b) satisfy conditions of Theorem 3.2. From assumption A4), (2.14'), and (2.20'), it follows that $A_{ni}^2 - EA_{ni}^2$ is a martingale difference:

$$E(A_{ni}^2 - EA_{ni}^2|F_{n,i-1}) = c_{ni}[E(e_{ni}^2|F_{n,i-1}) - \sigma^2] = 0.$$

By assumption A4) the functions $e_{ni}^2 - \sigma^2$ are uniformly integrable. By Theorem 3.2 $\sum_{i=2}^n (A_{ni}^2 - EA_{ni}^2)$ converges to zero in L_1 and, hence, in probability.

Lemma 2.4. plim $\sum_{i=2}^n (B_{ni}^2 - EB_{ni}^2) = 0$.

Proof. Denote $I_{ni} = B_{ni}^2 - EB_{ni}^2$. This time we use the mixingale WLLN because $\{I_{ni}, F_{ni}\}$ is not a m.d. sequence.

Put

$$I_{ni} = 0, \quad F_{ni} = \{\emptyset, \Omega\}, \quad i \le 1; \quad c_{ni} = \nu_n^2 \quad \forall i.$$

We shall show that $\{I_{ni}, F_{ni}\}$ satisfies conditions 1) through 3) of the definition of a L_1-mixingale from the Appendix.

1) Obviously, F_{ni} form an increasing sequence of σ-subfields of F.

2) Now we show that the family $\{I_{ni}/c_{ni}\}$ is uniformly integrable. Note that since EB_{ni}^2/c_{ni} are uniformly bounded (see (2.20″)), it suffices to prove that the variables $J_{ni} \equiv B_{ni}^2/c_{ni}$ are uniformly integrable. The estimate (see (2.14″) and assumption A4))

$$EJ_{ni} = \sigma^2 \Big\| \sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk} \Big\|_2^2 = \sigma^4 \sum_{k=1}^{i-1} \beta_2^{2(i-1-k)} \le c \tag{2.21}$$

proves uniform L_1-boundedness. By assumption A4) and Hölder's inequality with $r = p/4$ we have

$$\Big\| \prod_{j=1}^{4} e_{nk_j} \Big\|_r \le \prod_{j=1}^{4} \|e_{nk_j}\|_p \le c \tag{2.22}$$

for any $k_j \ge 1$. (2.21) and (2.22) imply (r' is defined from $1/r + 1/r' = 1$ and $1(A)$ is the indicator of a set A)

$$E\Big[\Big|\prod_{j=1}^{4} e_{nk_j}\Big| 1(J_{ni} > N)\Big] \le \Big\| \prod_{j=1}^{4} e_{nk_j} \Big\|_r [E1(J_{ni} > N)]^{1/r'} \le$$

$$\le c_1 N^{-1/r'} (EJ_{ni})^{1/r'} \le c_2 N^{-1/r'}.$$

Hence, uniformly with respect to n and i

$$EJ_{ni} 1(J_{ni} > N) \le \sum_{k=1}^{i-1} \beta_2^{2(i-1-k)} E e_{nk}^2 e_{ni}^2 1(J_{ni} > N) +$$

$$+2 \sum_{1 \le k < l \le i-1} |\beta_2|^{2i-2-k-l} E|e_{nk} e_{nl} e_{ni}^2 1(J_{ni} > N)| \le c_3 N^{-1/r'} \to 0, \ N \to \infty.$$

Thus, the functions I_{ni}/c_{ni} are uniformly integrable.

3) Bounds (3.5) and (3.6) are trivial for $i \le 1$. Let $i \ge 2$. For $m \ge 0$ and all $k \le i-1$ one has $F_{nk} \subset F_{ni} \subset F_{n,i+m}$. From (2.14″) then

$$E(I_{ni}|F_{n,i+m}) = I_{ni} \ \ \forall m \ge 0, \tag{2.23}$$

so (3.6) is trivial. To prove (3.5), consider three cases.

3.1) $m = 0$. (2.23) applies and yields, by the LIE, (2.14″) and (2.20″),

$$\|E(I_{ni}|F_{ni})\|_1 = \|I_{ni}\|_1 = \nu_n^2 \Big\| \Big(\sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk}\Big)^2 e_{ni}^2 - \sigma^4 (1 - \beta_2^{2(i-1)})/(1 - \beta_2^2) \Big\|_1 \le$$

$$\leq \nu_n^2 \Big\{ E\Big[\Big(\sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk}\Big)^2 E(e_{ni}^2|F_{n,i-1})\Big] + c_1 \Big\} = \nu_n^2 \Big(\sigma^2 \Big\| \sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk} \Big\|_2^2 + c_1\Big) \leq c_2 \nu_n^2.$$

Here we have used also assumptions A1) and A4).

3.2) $i - 1 \geq m \geq 1$. Noting that $F_{n,i-m} \subset F_{n,i-1}$ and using assumption A4), (2.14″) and (2.20″), we get

$$E(I_{ni}|F_{n,i-m}) = E[E(I_{ni}|F_{n,i-1})|F_{n,i-m}] =$$

$$= \nu_n^2 \sigma^2 E\Big[\sum_{k=1}^{i-1} \beta_2^{2(i-1-k)} (e_{nk}^2 - \sigma^2) + 2 \sum_{1 \leq k < l \leq i-1} \beta_2^{2i-2-k-l} e_{nk} e_{nl} | F_{n,i-m}\Big] =$$

$$= \nu_n^2 \sigma^2 \Big[\sum_{k=1}^{i-m} \beta_2^{2(i-1-k)} (e_{nk}^2 - \sigma^2) + 2 \sum_{1 \leq k < l \leq i-m} \beta_2^{2i-2-k-l} e_{nk} e_{nl}\Big] =$$

$$= \nu_n^2 \sigma^2 \Big[\Big(\sum_{k=1}^{i-m} \beta_2^{i-1-k} e_{nk}\Big)^2 - \sigma^2 \sum_{k=1}^{i-m} \beta_2^{2(i-1-k)}\Big].$$

Hence, with $\zeta_{m+1} \equiv c\beta_2^{2(m-1)}$ by orthogonality

$$\|E(I_{ni}|F_{n,i-m})\|_1 \leq \nu_n^2 \sigma^2 \Big(\Big\| \sum_{k=1}^{i-m} \beta_2^{i-1-k} e_{nk} \Big\|_2^2 + \sigma^2 \sum_{k=1}^{i-m} \beta_2^{2(i-1-k)}\Big) =$$

$$= 2\nu_n^2 \sigma^4 \sum_{k=1}^{i-m} \beta_2^{2(i-1-k)} \leq \nu_n^2 \zeta_{m+1}.$$

3.3) $m > i - 1$. Then $F_{n,i-m} = \{\emptyset, \Omega\}$ by definition, so by assumption A4), (2.14″), and (2.20″)

$$E(I_{ni}|F_{n,i-m}) = E[E(I_{ni}|F_{n,i-1})] =$$

$$= \nu_n^2 \sigma^2 E\Big[\sum_{k=1}^{i-1} \beta_2^{2(i-1-k)} (e_{nk}^2 - \sigma^2) + 2 \sum_{1 \leq k < l \leq i-1} \beta_2^{2i-2-k-l} e_{nk} e_{nl}\Big] = 0.$$

Summarizing, (3.5) holds with $\zeta_{m+1} = c\beta_2^{2(m-1)},\ 0 \leq m \leq i-1;\ \zeta_m = 0,\ m > i-1$.

By Lemma 2.1c), the scaling coefficients c_{ni} satisfy the requirements of Theorem 3.3, so $\| \sum_{i=2}^n I_{ni} \|_1 \to 0$, which proves the lemma.

Lemma 2.5. plim $\sum_{i=2}^n A_{ni} B_{ni} = 0$.

Proof. $\{A_{ni} B_{ni}; F_{ni}\}$ is a mixingale but its scaling coefficients c_{ni} do not seem to satisfy the conditions of Theorem 3.3. Therefore the approach here is different from that in Lemma 2.4. Denoting

$$r_{ni} = g_{ni} \nu_n \sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk},$$

we can write (see (2.14‴))

$$A_{ni}B_{ni} = r_{ni}e_{ni}^2 = r_{ni}(e_{ni}^2 - \sigma^2) + \sigma^2 r_{ni}. \tag{2.24}$$

By assumption A4), the variables $x_{ni} = (e_{ni}^2 - \sigma^2)/\sqrt{\sigma_1^2 - \sigma^4}$ satisfy Burkholder's condition from Theorem 3.5:

$$E(x_{ni}^2|F_{n,i-1}) = E\Big(\frac{e_{ni}^4 - 2\sigma^2 e_{ni}^2 + \sigma^4}{\sigma_1^2 - \sigma^4}|F_{n,i-1}\Big) = \frac{\sigma_1^2 - \sigma^4}{\sigma_1^2 - \sigma^4} = 1,$$

$$E(|x_{ni}||F_{n,i-1}) \geq c/\sqrt{\sigma_1^2 - \sigma^4}.$$

Therefore

$$E\Big|\sum_{i=2}^n r_{ni}(e_{ni}^2 - \sigma^2)\Big|^2 \leq (\sigma_1^2 - \sigma^4)E\max_{2\leq k\leq n}\Big|\sum_{i=2}^k r_{ni}x_{ni}\Big|^2 \leq c_1\sum_{i=2}^n Er_{ni}^2. \tag{2.25}$$

(2.12) implies

$$r_{ni} \leq c_2 \max_j |u_{nj}||\nu_n| \sum_{k=1}^{i-1} \beta_2^{i-1-k}|e_{nk}|.$$

Taking also into account Theorem 3.1b) and (2.13), we have

$$\sum_{i=2}^n Er_{ni}^2 \leq c_3(\max_j |u_{nj}|\nu_n)^2 \sum_{i=2}^n E\Big(\sum_{k=1}^{i-1}|\beta_2|^{i-1-k}e_{nk}\Big)^2 \leq$$

$$\leq c_4(\max_j |u_{nj}|\nu_n)^2 \sum_{i=2}^n\Big(\sum_{k=1}^{i-1}|\beta_2|^{i-1-k}\|e_{nk}\|_2\Big)^2 \leq c_5(\max_j |u_{nj}|)^2\nu_n^2 n \to 0.$$

This inequality and (2.25) show that

$$\operatorname{plim}\sum_{i=2}^n r_{ni}(e_{ni}^2 - \sigma^2) = 0. \tag{2.26}$$

Next we show that

$$\Big\|\sum_{i=2}^n r_{ni}\Big\|_2 \to 0. \tag{2.27}$$

Using g_n from (2.11) we have

$$\sum_{i=2}^n r_{ni} = \nu_n \sum_{i=2}^n\sum_{k=1}^{i-1}\beta_2^{i-1-k}g_{ni}e_{nk} = \nu_n\sum_{k=1}^{n-1}e_{nk}\sum_{i=k+1}^n \beta_2^{i-1-k}g_{ni}.$$

Let

$$\Phi_n z = \Big(\sum_{i=k}^{n} \beta_2^{i-k} z_i\Big)_{k=1}^{n}. \tag{2.28}$$

Φ_n is obtained from (3.3) by putting $\psi_j = 0,\ j < 0,\quad \psi_j = \beta_2^j,\ j \geq 0$. Then

$$\sum_{i=2}^{n} r_{ni} = \nu_n \sum_{k=1}^{n-1} e_{nk}(\Phi_n g_n)_{k+1} = \nu_n \sum_{k=2}^{n} e_{n,k-1}(\Phi_n g_n)_k.$$

It follows by orthogonality, Lemma 2.1a), and Theorem 3.1a) that

$$\Big\|\sum_{i=2}^{n} r_{ni}\Big\|_2 = \nu_n \sigma \Big[\sum_{k=2}^{n} (\Phi_n g_n)_k^2\Big]^{1/2} \leq c_1|\nu_n|\|g_n\| \leq c_2|\nu_n| \to 0.$$

Now (2.24), (2.26), and (2.27) prove the lemma.

The next lemma supplies the final ingredient for Theorem 3.4.

Lemma 2.6. If $\lim \Sigma_n > 0$, then $\operatorname{plim} \max_i |X_{ni}| = 0$.

Proof. Since $\lim \Sigma_n > 0$, the statement to be proved is equivalent to $\operatorname{plim} \max_i |Y_{ni}| = 0$. Obviously,

$$P(\max_i |Y_{ni}| > 2\epsilon) \leq P(\max_i |A_{ni}| > \epsilon) + P(\max_i |B_{ni}| > \epsilon).$$

With $p > 2$ we have by assumption A4), Lemma 2.1a) and (2.15)

$$P(\max_i |A_{ni}| > \epsilon) \leq \epsilon^{-p} E \max_i |A_{ni}|^p \leq \epsilon^{-p} \sum_{i=2}^{n} E|A_{ni}|^p =$$

$$= \epsilon^{-p} \sum_{i=2}^{n} |g_{ni}|^{p-2+2} E|e_{ni}|^p \leq c_1 \epsilon^{-p} \max_i |g_{ni}|^{p-2} \|g_n\|_2^2 \to 0.$$

Similarly, using the estimate $|\nu_n| \leq c/\sqrt{n}$, we have from (2.9) by Hölder's inequality and assumption A4)

$$P(\max_i |B_{ni}| > \epsilon) \leq \epsilon^{-p} \sum_{i=2}^{n} E|B_{ni}|^p = \epsilon^{-p}|\nu_n|^p \sum_{i=2}^{n} \Big\|\sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk} e_{ni}\Big\|_p^p \leq$$

$$\leq \epsilon^{-p}|\nu_n|^p \sum_{i=2}^{n} \Big(\sum_{k=1}^{i-1} |\beta_2|^{i-1-k} \|e_{nk} e_{ni}\|_p\Big)^p \leq$$

$$\leq c_1 \epsilon^{-p} n^{-p/2} \sum_{i=2}^{n} \Big(\sum_{k=1}^{i-1} |\beta_2|^{i-1-k} \|e_{nk}\|_{2p} \|e_{ni}\|_{2p}\Big)^p \leq c_2 \epsilon^{-p} n^{1-p/2} \to 0.$$

This completes the proof.

In the following two lemmas we consider convergence in probability of elements of Q (see (2.2)).

Lemma 2.7. $\text{plim}\|h_2\|^2 = \gamma^2 + (\frac{\sigma\lambda}{1+\lambda})^2 \frac{1}{1-\beta_2^2}$.

Proof. Let $G_n = \|h_2\|^2$. From (2.1), (2.6) and (2.10), one has

$$G_n = \frac{1}{(\|x_1\| + \sqrt{n})^2} \sum_{i=0}^{n-1} y_i^2, \ y_i = \beta_1(\Psi_n x_1)_i + (\Psi_n e)_i + \beta_2^i y_0. \tag{2.29}$$

Using notation (2.11) with $a_2 = 1$, we can write G_n as

$$G_n = \nu_n^2 \Big\{ y_0^2 + \sum_{i=1}^{n-1} [\beta_1(\Psi_n x_1)_i + (\Psi_n e)_i + \beta_2^i y_0]^2 \Big\} =$$

(multiplying through by ν_n^2 and using the identity $\beta_1\|x_1\|\nu_n = \mu_n$)

$$= \nu_n^2 y_0^2 + \sum_{i=1}^{n-1} [\mu_n(\Psi_n u_n)_i + \nu_n(\Psi_n e)_i + \nu_n\beta_2^i y_0]^2 =$$

(squaring the parentheses)

$$= \nu_n^2 y_0^2 + \sum_i [(\mu_n\Psi_n u_n)_i^2 + (\nu_n\Psi_n e)_i^2 + (\nu_n\beta_2^i y_0)^2 +$$

$$+2(\mu_n\Psi_n u_n)_i(\nu_n\Psi_n e)_i + 2((\mu_n\Psi_n u_n)_i + (\nu_n\Psi_n e)_i)\nu_n\beta_2^i y_0] = \sum_{i=1}^{5} G_{ni},$$

where we have denoted

$$G_{n1} = \sum_{i=0}^{n-1} (\nu_n\beta_2^i y_0)^2, \quad G_{n2} = \sum_{i=1}^{n-1} (\mu_n\Psi_n u_n)_i^2, \quad G_{n3} = \sum_{i=1}^{n-1} (\nu_n\Psi_n e)_i^2,$$

$$G_{n4} = 2\mu_n\nu_n \sum_{i=1}^{n-1} (\Psi_n u_n)_i(\Psi_n e)_i, \quad G_{n5} = 2\nu_n \sum_{i=1}^{n-1} ((\mu_n\Psi_n u_n)_i + (\nu_n\Psi_n e)_i)\beta_2^i y_0.$$

We consider these terms one by one.

1) $\text{plim}\, G_{n1} = 0$ because

$$\|G_{n1}\|_1 = \nu_n^2 \sum_{i=0}^{n-1} \beta_2^{2i} \|y_0^2\|_1 \le c/n \to 0.$$

2) $L_n\Psi_n u_n$ and $\Psi_n u_n$ have the same limits (see the proof of Lemma 2.1). Therefore, choosing $a_1 = 0$ and $a_2 = 1$ in Lemma 2.1, parts a) and b), we get

$$\lim_{n\to\infty} G_{n2} = \lim_{n\to\infty} (\|\mu_n\Psi_n u_n\|_2^2 - g_{nn}^2) = \gamma^2. \tag{2.30}$$

3) G_{n3} is represented as $G_{n6}+G_{n7}$ where

$$G_{n6}=EG_{n3}=\sigma^2\nu_n^2\sum_{i=1}^{n-1}\sum_{k=1}^{i}\beta_2^{2(i-k)},$$

$$G_{n7}=G_{n3}-EG_{n3}=\nu_n^2\sum_{i=1}^{n-1}\Big[\sum_{k=1}^{i}\beta_2^{2(i-k)}(e_{nk}^2-\sigma^2)+2\sum_{1\le k<l\le i}\beta_2^{2i-k-l}e_{nk}e_{nl}\Big].$$

By (2.13)

$$\lim_{n\to\infty}G_{n6}=\frac{\sigma^2}{1-\beta_2^2}\lim_{n\to\infty}\nu_n^2\sum_{i=1}^{n-1}(1-\beta_2^{2i})=\Big(\frac{\sigma\lambda}{1+\lambda}\Big)^2\frac{1}{1-\beta_2^2}. \tag{2.31}$$

Handling G_{n7} is the most difficult. We start with revealing its martingale nature. Changing the summation order and calculating the inner sums gives

$$G_{n7}=\nu_n^2\Big[\sum_{i=1}^{n-1}\sum_{k=1}^{i}\beta_2^{2(i-k)}(e_{nk}^2-\sigma^2)+2\sum_{i=1}^{n-1}\sum_{k=1}^{i-1}\sum_{l=k+1}^{i}\beta_2^{2i-k-l}e_{nk}e_{nl}\Big]=$$

$$=\nu_n^2\Big[\sum_{k=1}^{n-1}(e_{nk}^2-\sigma^2)\sum_{i=k}^{n-1}\beta_2^{2(i-k)}+2\sum_{k=1}^{n-2}e_{nk}\sum_{i=k+1}^{n-1}\sum_{l=k+1}^{i}\beta_2^{2i-k-l}e_{nl}\Big]=$$

$$=\nu_n^2\Big[\sum_{k=1}^{n-1}a_{nk}(e_{nk}^2-\sigma^2)+2\sum_{k=1}^{n-2}e_{nk}\sum_{l=k+1}^{n-1}e_{nl}b_{nkl}\Big]$$

where we denote

$$a_{nk}=\sum_{i=k}^{n-1}\beta_2^{2(i-k)}=\frac{1-\beta_2^{2(n-k)}}{1-\beta_2^2},\quad b_{nkl}=\sum_{i=l}^{n-1}\beta_2^{2i-k-l}=\beta_2^{l-k}a_{nl}.$$

Changing the order of summation once again and denoting

$$r_{n1}=\nu_n^2a_{n1}(e_{n1}^2-\sigma^2),\quad r_{ni}=\nu_n^2\Big[a_{ni}(e_{ni}^2-\sigma^2)+2e_{ni}\sum_{l=1}^{i-1}e_{nl}b_{nli}\Big],\quad 2\le i\le n-1,$$

we obtain

$$G_{n7}=\nu_n^2\Big[\sum_{k=1}^{n-1}a_{nk}(e_{nk}^2-\sigma^2)+2\sum_{l=2}^{n-1}e_{nl}\sum_{k=1}^{l-1}e_{nk}b_{nkl}\Big]=$$

$$=\nu_n^2\Big\{a_{n1}(e_{n1}^2-\sigma^2)+\sum_{i=2}^{n-1}\Big[a_{ni}(e_{ni}^2-\sigma^2)+2e_{ni}\sum_{k=1}^{i-1}e_{nk}b_{nki}\Big]\Big\}=\sum_{i=1}^{n}r_{ni}.$$

Here $\{r_{ni},F_{ni}\}$ is a m.d. sequence.

By Lemma 2.1c), the constants $c_{ni}=\nu_n^2$ satisfy conditions (b) and (c) of Theorem 3.2. To check the other conditions of that theorem, denote $s_{ni}=r_{ni}/c_{ni}$. For $2\le i\le n-1$

$$\|s_{ni}\|_1\le c_1\|e_{ni}^2-\sigma^2\|_1+2\sup_{i,l,n}\|e_{ni}e_{nl}\|_1\sum_{l=1}^{i-1}|b_{nli}|.$$

Here by Hölder's inequality and assumption A4)

$$\sum_{l=1}^{i-1}|b_{nli}|\le c_2\sum_{l=1}^{i-1}|\beta_2|^{i-l}\le c_3,\quad \|e_{ni}e_{nl}\|_1\le c_4, \tag{2.32}$$

so that $\|s_{ni}\|_1\le c_5$. Further, with $q=p/2$ we have

$$\|e_{ni}e_{nl}\|_q\le\|e_{ni}\|_p\|e_{nl}\|_p\le c_6\quad\forall i,l,n.$$

It follows that (q' is defined by $1/q+1/q'=1$)

$$E|e_{ni}e_{nl}1(s_{ni}>N)|\le\|e_{ni}e_{nl}\|_q[E1(s_{ni}>N)]^{1/q'}\le$$

$$\le c_6N^{-1/q'}\|s_{ni}\|_1^{1/q'}\le c_7N^{-1/q'}.$$

Hence, uniformly in i,n (see also (2.32))

$$E|s_{ni}1(s_{ni}>N)|=E\Big|\Big[a_{ni}(e_{ni}^2-\sigma^2)+2\sum_{l=1}^{i-1}e_{ni}e_{nl}b_{nli}\Big]1(s_{ni}>N)\Big|\le$$

$$\le c_8N^{-1/q'}\to 0,\quad N\to\infty.$$

We have proved that the family $\{s_{ni}\}$ is uniformly integrable. Hence, by Theorem 3.2 $\|G_{n7}\|_1\to 0$.

4) Using definitions (2.10) and (2.28), we can write

$$G_{n4}=2\mu_n\nu_n\sum_{i=1}^{n-1}(\Psi_nu_n)_i\sum_{k=1}^{i}\beta_2^{i-k}e_{nk}=2\mu_n\nu_n\sum_{k=1}^{n-1}e_{nk}\sum_{i=k}^{n-1}\beta_2^{i-k}(\Psi_nu_n)_i=$$

$$=2\mu_n\nu_n\sum_{k=1}^{n-1}e_{nk}\Big[\sum_{i=k}^{n}\beta_2^{i-k}(\Psi_nu_n)_i-\beta_2^{n-k}(\Psi_nu_n)_n\Big]=$$

$$=2\mu_n\nu_n\sum_{k=1}^{n-1}e_{nk}[(\Phi_n\Psi_nu_n)_k-\beta_2^{n-k}(\Psi_nu_n)_n].$$

By orthogonality and Theorem 3.1a)

$$\|G_{n4}\|_2=2|\mu_n\nu_n|\Big[\sum_{k=1}^{n-1}|(\Phi_n\Psi_nu_n)_k-\beta_2^{n-k}(\Psi_nu_n)_n|^2\Big]^{1/2}\le$$

$$\le c_1n^{-1/2}\Big[\|\Phi_n\Psi_nu_n\|_2+\|\Psi_nu_n\|\Big(\sum_{k=1}^{n-1}\beta_2^{2(n-k)}\Big)^{1/2}\Big]\le c_2n^{-1/2}\to 0.$$

5) By Theorem 3.1a) and (2.13)

$$\|G_{n5}\|_1 \le 2|\mu_n\nu_n| \sum_{i=1}^{n-1} |\beta_2^i(\Psi_n u_n)_i| \|y_0\|_1 + 2\nu_n^2 \sum_{i=1}^{n-1} |\beta_2^i| \sum_{k=1}^{i} |\beta_2^{i-k}| \|e_{nk} y_0\|_1 \le$$

$$\le c_1 \Big[\nu_n \Big(\sum_{i=1}^{n-1} \beta_2^{2i}\Big)^{1/2} \|\Psi_n u_n\|_2 + \nu_n^2\Big] \le c_2 n^{-1/2} \to 0.$$

Summarizing, of all terms in the decomposition of G_n, only G_{n2} and G_{n6} have nontrivial limits in probability. (2.30) and (2.31) give the desired result.

Lemma 2.8. plim $h_1'h_2 = \gamma$.

Proof. (2.1), (2.6) and (2.10) lead to

$$h_1'h_2 = \frac{1}{\|x_1\|(\|x_1\| + \sqrt{n})} \sum_{i=1}^{n} x_{1i} y_{i-1} =$$

$$= \frac{1}{\|x_1\|(\|x_1\| + \sqrt{n})} \Big[\sum_{i=1}^{n} x_{1i} \Big(\beta_1(\Psi_n x_1)_{i-1} + (\Psi_n e)_{i-1} + \beta_2^{i-1} y_0\Big)\Big] = G_1 + G_2 + G_3,$$

where

$$G_1 = \mu_n \sum_{i=1}^{n} u_{ni} (L_n \Psi_n u_n)_i = \mu_n u_n' L_n \Psi_n u_n,$$

$$G_2 = \nu_n \sum_{i=1}^{n} u_{ni} \sum_{k=1}^{i-1} \beta_2^{i-1-k} e_{nk} = \nu_n \sum_{k=1}^{n-1} e_{nk} \sum_{i=k+1}^{n} \beta_2^{i-1-k} u_{ni} =$$

$$= \nu_n \sum_{k=2}^{n} e_{n,k-1} \sum_{i=k}^{n} \beta_2^{i-k} u_{ni} = \nu_n \sum_{k=2}^{n} e_{n,k-1} (\Phi_n u_n)_k,$$

$$G_3 = \nu_n \sum_{i=1}^{n} u_{ni} \beta_2^{i-1} y_0.$$

Here we have used definitions (2.11) with $a_2 = 1$ and (2.28).

By virtue of Lemma 2.1d), $\lim_{n\to\infty} G_1 = \gamma$. By orthogonality and Theorem 3.1a)

$$\|G_2\|_2 \le \nu_n \|\Phi_n u_n\| \le c\nu_n \to 0.$$

Further, according to Theorem 3.1b),

$$\|G_3\|_1 \le \nu_n \|y_0\|_1 \max_i |u_{ni}| \sum_{i\ge 1} |\beta_2|^{i-1} \to 0.$$

These three facts prove the lemma.

Proof of Theorem 2.1. Recall that the problem of convergence in distribution of w has been reduced to that of $\sum Y_{ni}$, for each $a \in R^2$ (see (2.7)). We consider two cases.

1) If $\lim \Sigma_n > 0$, then convergence of $\sum Y_{ni}$ is equivalent to that of $\sum X_{ni}$, where $X_{ni} = Y_{ni}/\Sigma_n$. It is seen from the definition of Y_{ni} that X_{ni} are martingale differences, and they satisfy the normalization condition from Theorem 3.4. Condition (a) from that theorem is equivalent to $\text{plim } q_n(X) = 0$. Because $\lim \Sigma_n > 0$, Lemmas 2.3, 2.4, and 2.5 show that the last equation is true. Lemma 2.6 provides condition (b) from Theorem 3.4. Thus, $\sum X_{ni}$ converges to a standard normal and $\sum Y_{ni}$ converges to a normal with mean 0 and variance $\sigma^2 a' Q_\infty a$ (see (2.19)). By the Cramér-Wold theorem, this proves the first relation in (2.5) in the case under consideration.

2) Let us prove convergence in distribution of w_1, the first coordinate of w. Choosing in all previous definitions $a_1 = 1,\ a_2 = 0$, we see from (2.19) that $\lim \Sigma_n > 0$. Hence, the first part of the proof applies and $w_{\infty,1} = \text{dlim } w_1$ exists and has variance σ^2.

Now suppose that $\lim \Sigma_n = 0$. Then (2.19) implies $a_1 + \gamma a_2 = 0,\ a_2 \lambda = 0$. If $\lambda > 0$, then $a_2 = 0$ and $a_1 = 0$. In this case convergence of $a'w$ is trivial. To avoid triviality, we assume that

$$\lambda = 0, \quad a_2 \neq 0, \quad a_1 = -\gamma a_2. \tag{2.33}$$

For a general a satisfying (2.33) we are going to show that $\text{plim } a'w = 0$. Along with (2.7) one has $\text{plim } a'w = \text{plim} \sum_{i=2}^n Y_{ni}$, if the limit at the right exists. From (2.8), (2.9), and (2.10) it follows that

$$\sum_{i=2}^n Y_{ni} = a_2 \Big[\sum_{i=2}^n \big(-\gamma h_{1i} + \frac{\beta_1}{1+\lambda_n} (\Psi_n h_1)_{i-1} \big) e_{ni} + \frac{1}{m_2} \sum_{i=2}^n (\Psi_n e)_{i-1} e_{ni} \Big] = \tag{2.34}$$

(using (2.11) with $a_2 = 1, a_1 = -\gamma$)

$$= a_2 \Big[\sum_{i=2}^n (-\gamma u_n + \mu_n L_n \Psi_n u_n)_i e_{ni} + \nu_n \sum_{i=2}^n (L_n \Psi_n e)_i e_{ni} \Big].$$

Choosing $a_1 = -\gamma$ and $a_2 = 1$ in Lemma 2.1, parts a) and b), we obtain by orthogonality

$$\lim_{n\to\infty} \Big\| \sum_{i=2}^n (-\gamma u_n + \mu_n L_n \Psi_n u_n)_i e_{ni} \Big\|_2 = \lim_{n\to\infty} \| -\gamma u_n + \mu_n L_n \Psi_n u_n \| = 0. \tag{2.35}$$

Since $(L_n\Psi_n e)_j$ is $F_{n,j-1}$-measurable, assumption A4) gives

$$E\Big(\nu_n\sum_{i=2}^n(L_n\Psi_n e)_i e_{ni}\Big)^2=\nu_n^2E\Big[\sum_{i=2}^n(L_n\psi_n e)_i^2e_{ni}^2+ \tag{2.36}$$

$$+2\sum_{2\le i<j\le n}(L_n\Psi_n e)_i(L_n\Psi_n e)_j e_{ni}e_{nj}\Big]=\sigma^2\nu_n^2\sum_{i=2}^n E(L_n\Psi_n e)_i^2\le$$

$$\le\sigma^2\nu_n^2E\|L_n\Psi_n e\|_2^2\le$$

(using (3.2) and Theorem 3.1a))

$$\le c_1\nu_n^2\sum_{i=1}^n Ee_{ni}^2=c_2\nu_n^2 n=c_2\Big(\frac{\lambda_n}{1+\lambda_n}\Big)^2\to 0.$$

This is because $\lambda_n\to 0$ (see (2.13) and (2.33)). (2.34), (2.35), and (2.36) prove that $\operatorname{plim} a'w=a_2\operatorname{plim}(-\gamma w_1+w_2)=0$. Because w_1 converges in distribution to $w_{\infty,1}\in N(0,\sigma^2)$, w_2 converges in distribution to $w_{\infty,2}=\gamma w_{\infty,1}\in N(0,\sigma^2\gamma^2)$. w converges in distribution to w_∞ whose variance is $\sigma^2\begin{pmatrix}1&\gamma\\ \gamma&\gamma^2\end{pmatrix}$. The proof of the first equation in (2.5) is complete.

The second equation in (2.5) is an immediate consequence of Lemmas 2.7 and 2.8.

3. APPENDIX

By definition, the interpolation operator $D_n: R^n\to L_2$ takes a vector $z\in R^n$ to a simple function

$$D_n z=\sqrt{n}\sum_{j=1}^n z_j 1(i_j).$$

Here $1(i_j)$ stands for the indicator of i_j. The lag operator $L_n: R^n\to R^n$ is defined by

$$(L_n z)_j=z_{j-1},\quad j=2,\dots,n;\quad (L_n z)_1=0. \tag{3.1}$$

It is easy to see that the operators d_n and L_n are uniformly bounded and D_n is isometric:

$$\|d_n f\|\le\|f\|,\ f\in L_2;\ \|L_n z\|\le\|z\|,\ \|D_n z\|=\|z\|,\ z\in R^n. \tag{3.2}$$

Let $\{\psi_j: j=0,\pm1,\dots\}$ be a summable sequence of real numbers. We define $\Psi_n: R^n\to R^n$ by

$$(\Psi_n z)_k=\sum_{j=1}^n\psi_{j-k}z_j,\quad k=1,\dots,n. \tag{3.3}$$

With the sequence $\{\psi_j\}$ we can also associate the number $\alpha_\psi = \sum_j |\psi_j| < \infty$. It is easy to check that

$$\|L_n \Psi_n z - \Psi_n L_n z\| = \Big[z_n^2 \sum_{k=2}^{n} \psi_{n-k+1}^2 + \Big(\sum_{j=1}^{n-1} \psi_j z_j\Big)^2\Big]^{1/2}. \tag{3.4}$$

The less obvious properties, which have been established in Mynbaev (2001), are gathered in the next theorem.

Theorem 3.1.

a) If $\alpha_\psi < \infty$, then

$$\|\Psi_n z\| \le \alpha_\psi \|z\|, \quad z \in R^n, \quad n \ge 1.$$

b) If $\{u_n\}$ is L_2-approximated by f, then

$$\lim_{n\to\infty} \|u_n\| = \|f\|, \quad \lim_{n\to\infty} \max_{1\le j\le n} |u_{nj}| = 0.$$

c) If $\alpha_\psi < \infty$ and $\{u_n\}$ is L_2-approximable, then

$$\lim_{n\to\infty} \Big\| \Big(\sum_{j=-\infty}^{\infty} \psi_j - \Psi_n \Big) u_n \Big\| = 0.$$

d) If $\{u_n\}$ is L_2-approximable, then

$$\lim_{n\to\infty} \|L_n u_n - u_n\| = 0.$$

The next three results can be found in Davidson (1994).

Theorem 3.2 (Chow-Davidson martingale WLLN). Let $\{X_{ni}, F_{ni}\}$ be a martingale difference array, $\{c_{ni}\}$ a positive constant array, and $\{k_n\}$ an increasing integer sequence with $k_n \uparrow \infty$. If

(a) $\{X_{ni}/c_{ni}\}$ is uniformly integrable,

(b) $\limsup_{n\to\infty} \sum_{i=1}^{k_n} c_{ni} < \infty$, and

(c) $\lim_{n\to\infty} \sum_{i=1}^{k_n} c_{ni}^2 = 0$,

then $\| \sum_{i=1}^{k_n} X_{ni} \|_1 \to 0$.

Let (Ω, F, P) be a probability space. The array $\{\{X_{ni}, F_{ni}\}_{i=-\infty}^{\infty}\}_{n=1}^{\infty}$ is called an L_1-mixingale, if

1) for each n, $\{F_{ni}\}$ is an increasing sequence of σ-subfields of F,

2) X_{ni} are integrable random variables, and

3) there exist an array of nonnegative constants $\{\{c_{ni}\}_{i=-\infty}^{\infty}\}_{n=1}^{\infty}$ and a nonnegative sequence $\{\zeta_m\}_{m=0}^{\infty}$ such that $\lim_{m\to\infty}\zeta_m = 0$ and

$$\|E(X_{ni}|F_{n,i-m})\|_1 \le c_{ni}\zeta_m, \tag{3.5}$$

$$\|X_{ni} - E(X_{ni}|F_{n,i+m})\|_1 \le c_{ni}\zeta_{m+1} \tag{3.6}$$

hold for all i, n, and $m \ge 0$.

Theorem 3.3 (Andrews-Davidson mixingale WLLN). Let the array $\{X_{ni}, F_{ni}\}$ be an L_1-mixingale with respect to a constant array $\{c_{ni}\}$. If for some increasing integer sequence with $k_n \uparrow \infty$ conditions (a), (b), and (c) from Theorem 3.2 are satisfied, then $\| \sum_{i=1}^{k_n} X_{ni} \|_1 \to 0$.

Theorem 3.4 (McLeish CLT). Let $\{X_{ni}, F_{ni}\}$ be a m.d. array with finite unconditional variances σ_{ni}^2, and $\sum_{i=1}^{n}\sigma_{ni}^2 = 1$. If

(a) $\operatorname{plim}\sum_{i=1}^{n} X_{ni}^2 = 1$, and

(b) $\operatorname{plim}\max_{1\le j\le n}|X_{nj}| = 0$,

then $\sum_{j=1}^{n} X_{nj}$ converges in distribution to an element of $N(0,1)$.

Let $\{X_{ni}, F_{ni}\}$ be a m.d. array and let r_i be $F_{n,i-1}$-measurable. Then

$$S_n = \sum_{i=1}^{n} r_i X_{ni}$$

is called a transform of $\{X_{ni}, F_{ni}\}$. The next theorem has been established by Burkholder (1973).

Theorem 3.5. Let $\{X_{ni}, F_{ni}\}$ satisfy

$$E(X_{ni}^2|F_{n,i-1}) = 1, \quad E(|X_{ni}||F_{n,i-1}) \ge c.$$

Then the martingale S_n satisfies

$$E\max_{1\le j\le n} S_j^2 \le c\sum_{j=1}^{n} Er_j^2.$$

BIBLIOGRAPHY

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

Anderson, T. W., and Kunitomo, N. (1992). Asymptotic distribution of regression and autoregression coefficients with martingale difference disturbances. *J. Multivariate Anal.*, **40**, 221–243.

Andrews, D. W. K. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, **4**, 458–467.

Andrews, D. W. K., and McDermott, C. J. (1995). Nonlinear econometric models with deterministically trending variables. *Rev. Econom. Stud.*, **62**, no. 3, 343–360.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.*, **1**, no. 1, 14–42.

Chow, Y. S. (1971). On the L_p-convergence for $n^{-1/p}S_n$, $0 < p < 2$. *Ann. Math. Statist.*, **36**, 393–394.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford Univ. Press, Oxford.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton Univ. Press, Princeton, NJ.

Leonenko, N. N., and Šilac-Benšić, M. (1997). On the asymptotic distributions of least square estimations in a regression model with singular errors. *Dopov. Nats. Akad. Nauk Ukr. Mat. Prirodozn. Tekh. Nauki*, no. 7, 26–31.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.*, **2**, 620–628.

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. In: *7 Escola de Séries Temporais e Econometria. Sociedade Brasileira de Econome-*

tria. August 6-8, 80–82.

Mynbaev, K. T. (2001). L_p-approximable sequences of vectors and limit distribution of quadratic forms of random variables. *Adv. Appl. Math.*, **26**, 302–329.

Mynbaev, K. T. (2003). Modeling deterministic regressors. In: *Advances in Statistical Inferential Methods: Theory and Applications.* Almaty, Kazakhstan, June 9-12, 133–148.

Mynbaev, K. T., and Castelar, I. (2001). *The Strengths and Weaknesses of L_2-approximable Regressors. Two Essays on Econometrics.* Fortaleza: Expressão Gráfica, v.1.

Nabeya, S. (2000). Asymptotic distributions for unit root test statistics in nearly integrated seasonal autoregressive models. *Econometric Theory*, **16**, no. 2, 200–230.

Ng, S., and Vogelsang, T. J. (2002). Forecasting autoregressive time series in the presence of deterministic components. *Econom. J.*, **5**, no. 1, 196–224.

Rahbek, A.; Kongsted, H. C., and Jorgensen, C. (1999). Trend stationarity in the $I(2)$ cointegration model. *J. Econometrics*, **90**, no. 2, 265–289.

Sibbertsen, Ph. (2001). S-estimation in the linear regression model with long-memory error terms under trend. *J. Time Ser. Anal.*, **22**, no. 3, 353–363.

Sims, C. A.; Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, **58**, no. 1, 113–144.

Tam, Wing-kuen, and Reinsel, G. C. (1998). Seasonal moving-average unit root tests in the presence of a linear trend. *J. Time Ser. Anal.*, **19**, no. 5, 609–625.