OLS Asymptotics for Vector Autoregressions with Deterministic Regressors

Mynbayev, Kairat

Kazakh-British Technical University

2007

OLS Asymptotics for Vector Autoregressions with Deterministic Regressors

Kairat T. Mynbaev
February 14, 2007
Kazakh-British Technical University
Almaty, Kazakhstan
kairat@fulbrightweb.org

Abstract

We consider a mixed vector autoregressive model with deterministic exogenous regressors and an autoregressive matrix whose characteristic roots are less than 1 in absolute value. The errors are 2+epsilon-integrable martingale differences with heterogeneous second-order conditional moments. The behavior of the OLS estimator depends on the rate of growth of the exogenous regressors. For bounded or slowly growing regressors we prove asymptotic normality. In case of quickly growing regressors (e.g., polynomial trends) the result is negative: the OLS asymptotics cannot be derived using the conventional scheme and any nonstochastic diagonal normalizer.

Keywords: vector autoregression, polynomial trend, deterministic regressor, OLS estimator, asymptotic distribution

JEL classification: C32

1 Introduction

Autoregressive models have a long and rich history. In economics, dependence of a variable on its own past values is a very plausible assumption. Economic applications necessitated introduction of vector autoregressions many of which can be formalized as

$$y_t = Ax_t + By_{t-1} + e_t,\ t = 1, ..., n, \tag{1.1}$$

where $y_t,\ x_t,\ e_t$ are random vectors and A and B are parameter matrices to be estimated from observed y_t's and x_t's. The x_t's are assumed to be exogenous (determined outside the system) and the minimal assumption about the unobserved errors e_t is that their mean is zero. The general theory of vector autoregressions is described in Lütkepohl (1991), Hamilton (1994) and Charemza and Deadman (1992), among others.

The case when the characteristic roots of B lie inside the unit circle $|\lambda| < 1$ is called *stable*. We consider only the stable case and refer to Tanaka (1996) and Nielsen (2005) regarding the unstable case. Conditions on the exogenous regressors critically depend on whether they are assumed deterministic or stochastic. We focus on deterministic regressors and suggest the reader to consult Anderson and Kunitomo (1992) about results for stochastic regressors.

Among the models with deterministic regressors, those with polynomial trends are of particular interest. The OLS asymptotics for autoregressions with polynomial trends has been a long-standing issue (see Hamilton (1994, Chapter 16) for more information). Sims et al. (1990) have proposed a linear transformation to investigate such a model. However, that transformation uses unknown coefficients and therefore is not feasible. Our solution to the problem is, in a sense, negative. To explain the format of the results, we need several definitions.

By putting equations (1.1) side by side we can write them in a matrix form

$$Y_n = AX_n + BY_n^- + \mathcal{E}_n \tag{1.2}$$

where

$$Y_n = (y_1, ..., y_n),\ X_n = (x_1, ..., x_n),\ Y_n^- = (y_0, ..., y_{n-1}),\ \mathcal{E}_n = (e_1, ..., e_n).$$

Denoting $\Gamma = (A, B),\ Z_n = \begin{pmatrix} X_n \\ Y_n^- \end{pmatrix}$ we write (1.2) as

$$Y_n = \Gamma Z_n + \mathcal{E}_n \tag{1.3}$$

and the OLS estimator of Γ is given by

$$\widehat{\Gamma}_n = Y_n Z_n'(Z_n Z_n')^{-1} \tag{1.4}$$

(see, e.g., Lütkepohl, 1991).

A basic fact about OLS estimators is that they should be centered and scaled to obtain convergence in distribution. Centering means passing from (1.3)+(1.4) to

$$\widehat{\Gamma}_n - \Gamma = \mathcal{E}_n Z_n'(Z_n Z_n')^{-1}. \tag{1.5}$$

As for the scaling, we follow Anderson's (1971) suggestion. See Mynbaev and Castelar (2001) for discussion of its advantages. Let D_n be some nonsingular diagonal matrix, called a *normalizer*. Then (1.5) implies

$$(\widehat{\Gamma}_n - \Gamma)D_n = \mathcal{E}_n Z_n' D_n^{-1}(D_n^{-1} Z_n Z_n' D_n^{-1})^{-1}. \tag{1.6}$$

We use the name *N-factor* (numerator) for $\mathcal{E}_n Z_n' D_n^{-1}$ and *D-factor* (denominator) for $D_n^{-1} Z_n Z_n' D_n^{-1}$. By the *conventional scheme* of deriving the OLS asymptotics we mean the procedure consisting of three steps:

(1) choose an appropriate normalizer D_n,

(2) prove convergence of the N-factor in distribution to some normal vector,

(3) prove convergence of the D-factor in probability to some nonsingular matrix Q.

Convergence in distribution of $(\widehat{\Gamma}_n - \Gamma)D_n$ follows trivially from (1.6) and the conventional scheme.

Lately there have been attempts to design unified approaches to modeling deterministic regressors. The first approach has been used to study consistency of the OLS estimator. The details and history can be found in Nielsen (2005). The second has been undertaken by Andrews and McDermott (1995) in the context of nonlinear models. A third approach, more suited for linear models, has been developed by Mynbaev (2006a) in the scalar case with just one exogenous regressor and one lag:

$$y_t = \alpha x_t + \beta y_{t-1} + e_t,\ t = 1, ..., n. \tag{1.7}$$

Here we follow Mynbaev's (2001) methodology of approximating infinite sequences of vectors with functions of a continuous argument. In the rest of the Introduction we use (1.7) for simplicity and denote $||x||_2 = \left(\sum_{t=1}^n x_t^2\right)^{1/2}$. Under some regularity conditions the situation with the OLS asymptotics can be described qualitatively by two statements:

I. Let $\kappa_0 = \lim_{n\to\infty} \sqrt{n}/||x||_2$. If $\kappa_0 > 0$, then $\det Q \neq 0$ and the conventional scheme provides asymptotic normality of the OLS estimator.

II. If $\kappa_0 = 0$, then $\det Q = 0$ and the conventional scheme with our normalizer does not work. Moreover, there is no diagonal nonstochastic normalizer which would render the conventional scheme feasible.

Example. An autoregression with a linear trend $y_t = \alpha + \beta t + \rho y_{t-1} + e_t,\ t = 1, ..., n$, has been extensively studied and applied in the literature (one of the most recent references is Kim et al., 2003). Assume that $|\rho| < 1$, the initial condition y_0 is square-integrable and, for simplicity, that the errors are i.i.d normal. The number κ_0 from Assumption 4 (Section 2.3) is zero, so by Theorem 3.1 and 3.2 the D-factor converges in probability to Q with $\det Q = 0$ and the N-factor converges in distribution to a degenerate normal vector with our normalization. Further, if $\alpha\beta \neq 0$, then the asymptotics of the OLS estimator cannot be derived using the conventional scheme and any diagonal nonstochastic normalization. □

In this example the trend pushes the dependent variable to the extent that the lag becomes asymptotically collinear with the trend. It is always like this if the exogenous regressor is a polynomial. If x_t is a polynomial and one relies on the conventional scheme in studying the asymptotical properties, it is incorrect to include the lag y_{t-1} in the right-hand side. Our result does not exclude the possibility of convergence in distribution of some linear functionals of the OLS estimator.

Anderson and Kunitomo (1992) impose three infinite sets of conditions: one on the errors

$$\frac{1}{n}\sum_{t=\max\{r,s\}+2}^{n} \sigma_t e_{t-1-r} e_{t-1-s} \xrightarrow{p} \delta_{rs}\sigma^2,\ \ r,s=0,1,2,... \tag{1.8}$$

where $\delta_{ss}=1$ and $\delta_{rs}=0$ for $r\neq s$, another

$$\frac{1}{n}\sum_{t=1}^{n-h} u_{t+h}u_t \xrightarrow{p} M_h = M_{-h},\ \ h=0,1,2,... \tag{1.9}$$

on the normalized regressor $(u_1,...,u_n)=(x_1,...,x_n)/||x||_2$ and the last set on the interaction of the errors with the normalized regressor

$$\frac{1}{n}\sum_{t=1}^{n-h} u_{t+h}e_t \xrightarrow{p} 0,\ \ h=1,2,... \tag{1.10}$$

Our method allows us to avoid (1.8), (1.9) and (1.10). The method also improves upon Mynbaev (2006a): the Andrews (1988) weak law of large numbers for mixingales and Burkholder (1973) theorem on transforms of martingales are not used, while the errors integrability requirement is lower and heterogeneous errors are allowed.

Of course, the choice of the normalizer is of paramount importance. It is $||x||_2$ for $y_t=\alpha x_t+e_t$ and $\sqrt{n}$ for a pure autoregression $y_t=\beta y_{t-1}+e_t$ (see Anderson, 1971). Mynbaev (2006a) has shown that

$$D_n=\begin{pmatrix} ||x||_2 & 0 \\ 0 & ||x||_2+\sqrt{n}\end{pmatrix}$$

works for (1.7) with possibly growing exogenous regressor. Note that $||x||_2+\sqrt{n}$ is equivalent to $\max\{||x||_2,\sqrt{n}\}$ which explains our choice of D_n. Our experience warrants a general principle: if you have normalizers for two extreme cases (for models with only exogenous regressors or only lags), then set the normalizer for the autoregressive part of the combined model to the maximum of the normalizers for the extreme cases.

In Section 2 we explain the terminology and list some auxiliary statements to be used later. The main assumptions are gathered in Section 2.3. Section 3 contains the main results and their proofs, see especially Theorem 3.3. Theorem 3.3 can be applied in the spirit of Anderson and Kunitomo (1994).

2 Assumptions and Auxiliary Statements

2.1 Operators Arising in the Theory of Autoregressive Models

For $1 \le p \le \infty$ denote L_p the space of measurable on $(0,1)$ functions F provided with the norm

$$\|F\|_p = \left(\int_0^1 |F(x)|^p dx\right)^{1/p}, \ 1 \le p < \infty; \ \|F\|_\infty = \operatorname{ess\,sup}_{x \in (0,1)} |F(x)|.$$

Its discrete analogue l_p consists of sequences $\{z_i : \ i \in I\}$ having a finite norm

$$\|z\|_p = \left(\sum_{i \in I} |z_i|^p\right)^{1/p}, \quad 1 \le p < \infty; \quad \|z\|_\infty = \sup_{i \in I} |z_i|.$$

The set of indices I depends on the context. In particular, we use $\mathbb{R}^n_p$ (the set of n-dimensional vectors) and $\mathbb{M}_p$ (the set of matrices of all sizes). For $p \in [1, \infty]$ the number or symbol q is defined from $1/p + 1/q = 1$. A *discretization operator* $d_{np} : L_p \to \mathbb{R}^n_p$ acts on $F \in L_p$ according to

$$(d_{np}F)_i = n^{1/q} \int_{(i-1)/n}^{i/n} F(x)dx, \quad i = 1, \ldots, n.$$

Let $\{z_n\}$ be a sequence of vectors such that $z_n \in \mathbb{R}^n$ for all n. Following Mynbaev (2001) we say that $\{z_n\}$ is L_p-*approximable* if there exists a function $z^c \in L_p$ such that

$$\|z_n - d_{np} z^c\|_p \to 0, \ n \to \infty. \tag{2.1}$$

The superscript c emphasizes that z^c is considered a continuous proxy for $\{z_n\}$. In this case we also say that $\{z_n\}$ is L_p-*close* to z^c. This notion is designed for modeling deterministic regressors in linear models and should be distinguished from L_p-approximability introduced in Pötcher and Prucha (1991) for modeling stochastic regressors in nonlinear models. For reader's convenience some properties of L_p-approximable sequences are listed in Theorem 2.1.

We need generalizations of the above definitions to matrix-valued functions. Denote $\tau_n = \{1, ..., n\}$. For a matrix-valued function $F : \tau_n \to \mathbb{M}_p$ its norm is defined by

$$||F; l_p(\tau_n, \mathbb{M}_p)|| = \begin{cases} \left(\sum_{t=1}^n ||F_t||_p^p\right)^{1/p}, \ p < \infty, \\ \max_{1\le t\le n} ||F_t||_\infty, \ p = \infty. \end{cases}$$

We always assume that such a function has values of the same size. By definition, the discretization operator is applied to matrices element-wise. Let $s(M)$ denote the size of a matrix A (a product of its dimensions). A sequence $\{F_n\}$ such that $F_n \in l_p(\tau_n, \mathbb{M}_p)$ for all n and $s(F_1) = s(F_2) = ...$ is called L_p-*approximable* if there is a matrix F^c with components from L_p such that $||F_n - d_{np}F^c; l_p(\tau_n, \mathbb{M}_p)|| \to 0$, $n \to \infty$. If this is true we also say that $\{F_n\}$ is L_p-*close* to F^c. Obviously, uniform boundedness of norms

$$\sup_n ||F_n; l_p(\tau_n, \mathbb{M}_p)|| < \infty \tag{2.2}$$

is necessary for L_p-approximability. We write $F^c \in L_p$ to mean that all components of F^c belong to L_p. $F^c \in C[0, 1]$ has a similar meaning where $C[0, 1]$ is the set of continuous functions on $[0, 1]$.

A matrix F with n columns is considered a function on τ_n with values F_t equal to its columns, $t = 1, ..., n$. If A, B are two matrices, the function with values AF_1B, ..., AF_nB should be distinguished from the function with values $A_1F_1B_1$, ..., $A_nF_nB_n$ where A, B are two functions. In both cases we denote the product by AFB indicating whether A, B are matrices or functions. Let F be a matrix-valued function. With two square matrices A, B we can associate three operators:

$$\begin{aligned} (P_AF)_t &= \sum_{s=1}^{t-1} A^{t-1-s}F_s, \ (Q_AF)_t = \sum_{s=t+1}^{n} F_sA^{s-1-t}, \\ (R_{A,B}F)_t &= \sum_{s=1}^{t-1} A^{t-1-s}F_sB^{t-1-s}, \ t = 1, ..., n, \end{aligned} \tag{2.3}$$

where by definition the corresponding matrix is null if the summation set is empty: $(P_AF)_1 = 0$, $(Q_AF)_n = 0$, $(R_{A,B}F)_1 = 0$. Note that along with the sum $(P_AF)_t = A^{t-2}F_1 + ... + A^0F_{t-1}$ with descending orders of A one can think of ascending orders as in $A^0F_1 + ... + A^{t-2}F_{t-1}$. Observe also that in P_A the summation set increases with t, whereas in Q_A it decreases. We use the same notation P_A for such modalities because the corresponding operators have the same limits. The same agreement applies to the other two operators.

Theorem 2.1 (Mynbaev, 2006b) (i) If $\{X_n\}$ is L_p-approximable and $p < \infty$, then $\lim_{n\to\infty} \max_{1\le t\le n} ||X_{nt}||_p = 0$.

(ii) Let $1 < p < \infty$. Consider sequences of matrix-valued functions $\{X_n\}$, $\{Y_n\}$, $\{Z_n\}$ such that X_n, Y_n, Z_n are defined on τ_n, $n = 1, 2, ...$ If $\{X_n\}$ is L_p-close to $X^c \in L_p$, $\{Y_n\}$ is L_q-close to $Y^c \in L_q$ and $\{Z_n\}$ is L_∞-close to $Z^c \in C[0,1]$, then

$$\lim_{n\to\infty} \sum_{t=1}^{n} X_{nt}Y_{nt}Z_{nt} = \int_0^1 X^c(x)Y^c(x)Z^c(x)dx.$$

(iii) Suppose B is a square matrix with eigenvalues satisfying $|\lambda| < 1$ and let $1 \le p \le \infty$. Then

$$\max\{||P_B||, ||Q_B||, ||R_{B,B'}||\} < \infty$$

uniformly in $n = 1, 2, ...$ where the norms of operators are from $l_p(\tau_n, \mathbb{M}_p)$ to itself. Suppose, further, that $p < \infty$ and $\{X_n\}$ is L_p-close to $X^c \in L_p$. Then $\{P_BX_n\}$ is L_p-close to $(I-B)^{-1}X^c$, $\{Q_BX_n\}$ is L_p-close to $X^c(I-B)^{-1}$ and $\{R_{B,B'}X_n\}$ is L_p-close to $\sum_{s=0}^{\infty} B^sX^cB'^s$.

(iv) If $\{X_n\}$ is L_p-close to $X^c \in L_p$ and $\{Y_n\}$ is L_p-close to $Y^c \in L_p$, then $\{X_n + Y_n\}$ is L_p-close to $X^c + Y^c$.

(v) If $\{X_n\}$ is L_p-close to $X^c \in L_p$, $p < \infty$, and $\{Y_n\}$ is L_∞-close to $Y^c \in C[0,1]$, then $\{X_nY_n\}$ is L_p-close to X^cY^c. In particular, if $\{A_n\}$ is a sequence of matrices converging to A, then $\{A_nX_n\}$ is L_p-close to AX^c.

(vi) If $\{X_n\}$ is L_∞-close to $X^c \in L_\infty$, then $\{n^{-1/p}X_n\}$ is L_p-close to X^c.

2.2 Elements of the conventional scheme

Everywhere we abide by the usual matrix algebra conventions: all vectors are column-vectors and all matrices in the same formula are compatible. All properties of the Kronecker product, traces and vectorization we use can be found in Lütkepohl (1991). $\det A$ is also denoted $|A|$.

The system of vector equations (1.1) with just one lag of the dependent variable encompasses a variety of cases we do not want to consider. For example, Lütkepohl (1991, Section 10.5.1) has a system in which y_t includes lagged exogenous regressors. The application we have in mind is the system of scalar equations

$$y_t = \alpha_1 x_{1t} + ... + \alpha_r x_{rt} + \beta_1 y_{t-1} + ... + \beta_s y_{t-s} + e_t \tag{2.4}$$

which can be written in form (1.1) with matrices

$$A = \begin{pmatrix} \alpha_1 & ... & \alpha_r \\ 0 & ... & 0 \\ ... & ... & ... \\ 0 & ... & 0 \end{pmatrix}_{s\times r}, \ B = \begin{pmatrix} \beta_1 & ... & \beta_{s-1} & \beta_s \\ 1 & ... & 0 & 0 \\ ... & ... & ... & ... \\ 0 & ... & 1 & 0 \end{pmatrix}_{s\times s}$$

if the x_t, y_t and e_t in (1.1) are

$$\begin{pmatrix} x_{1t} \\ ... \\ x_{rt} \end{pmatrix}_{r\times 1}, \begin{pmatrix} y_t \\ y_{t-1} \\ ... \\ y_{t-s+1} \end{pmatrix}_{s\times 1}, \begin{pmatrix} e_t \\ 0 \\ ... \\ 0 \end{pmatrix}_{s\times 1},$$

respectively. Here we have r different exogenous regressors and just one (scalar) dependent variable, even though in the vector form there is an s-dimensional dependent vector. Then in (1.2) the sizes are $s(Y_n) = s(\mathcal{E}_n) = s \times n$, $s(X_n) = r \times n$; in (1.3) $s(\Gamma) = s \times (r + s)$, $s(Z_n) = (r + s) \times n$.

Making a start from Anderson (1971, Theorem 2.6.1) we use as a normalizer for X_n the matrix

$$d_n = \text{diag}\,[d_{n1}, ..., d_{nr}]$$

with Euclidean norms $d_{ni} = \left(\sum_{t=1}^n x_{it}^2\right)^{1/2}$ of rows of X_n on the main diagonal. Following Mynbaev (2006a) we choose $\Delta_n I_s$ as a normalizer for Y_n^-, where

$$\Delta_n = \max\{d_{n1}, ..., d_{nr}, \sqrt{n}\}.$$

Denoting

$$H_n = d_n^{-1} X_n, \ D_n = \begin{pmatrix} d_n & 0 \\ 0 & \Delta_n I_s \end{pmatrix}_{(s+r)\times(s+r)} \tag{2.5}$$

we finalize the definition of the elements of the conventional scheme with

$$D_n^{-1} Z_n = \begin{pmatrix} H_n \\ \frac{1}{\Delta_n} Y_n^- \end{pmatrix}. \tag{2.6}$$

It is easy to obtain from (1.1) by induction

$$y_t = \sum_{s=1}^{t} B^{t-s}(Ax_s + e_s) + B^t y_0, \ t \geq 1.$$

This equation, (2.5) and (2.3) give $Y_n^- = M_n + \rho_n$ where

$$M_n = P_B(AX_n + \mathcal{E}_n) = P_B(Ad_nH_n + \mathcal{E}_n), \ \rho_n = (y_0, By_0, ..., B^{n-1}y_0) \quad (2.7)$$

are the *main part* and *residual*, respectively. From (2.6) we see that the N-factor equals

$$\mathcal{E}_n Z_n' D_n^{-1} = (\mathcal{E}_n H_n', \frac{1}{\Delta_n}\mathcal{E}_n M_n') + (0, \frac{1}{\Delta_n}\mathcal{E}_n \rho_n') \quad (2.8)$$

and the D-factor is

$$\begin{aligned} D_n^{-1} Z_n Z_n' D_n^{-1} \quad &= \begin{pmatrix} H_nH_n' & \frac{1}{\Delta_n}H_nM_n' \\ \frac{1}{\Delta_n}M_nH_n' & \frac{1}{\Delta_n^2}M_nM_n' \end{pmatrix} \\ &+ \begin{pmatrix} 0 & \frac{1}{\Delta_n}H_n\rho_n' \\ \frac{1}{\Delta_n}\rho_nH_n' & \frac{1}{\Delta_n^2}(M_n\rho_n' + \rho_nM_n' + \rho_n\rho_n') \end{pmatrix}. \quad (2.9) \end{aligned}$$

All terms containing the residual will be shown to be asymptotically negligible.

2.3 Assumptions

Here we list and discuss the main assumptions. Additional assumptions are made when the analysis reveals their necessity.

Assumption 1 (stability) All eigenvalues of B satisfy $|\lambda| < 1$.

This condition implies existence of $c > 0$ and $\lambda \in (0, 1)$ such that

$$||B^k|| \leq c\lambda^k, \ k = 0, 1, ... \quad (2.10)$$

(see, for example, Anderson 1971, Lemma 5.5.1).

Assumption 2 (on normalized regressors) The sequence $\{H_n\}$ (see (2.5)) is L_2-close to some vector $H^c \in L_2$.

Recall that H_n is considered a function on τ_n, its columns $H_{n1}, ..., H_{nn}$ being its values. Mynbaev and Castelar (2001) have shown that this condition is satisfied for constants, logarithmic and polynomial trends and is not satisfied for exponential trends.

The error matrices $\mathcal{E}_n$ can be more general that it is implied by (2.4). The columns e_{nt} of $\mathcal{E}_n$ may depend on n and its rows, starting from the second, don't have to be null. By $\stackrel{d}{\rightarrow}$ and dlim ($\stackrel{p}{\rightarrow}$ and plim) we denote convergence in distribution (in probability, respectively). $I(A)$ denotes the indicator of a set A.

Assumption 3 (on errors) (i) For each n, the columns e_{nt} are martingale differences with respect to nested σ-fields $F_{n0} \subset F_{n1} \subset ... \subset F_{nn}$, that is, e_{nt} is F_{nt}-measurable and $E(e_{nt}|F_{n,t-1}) = 0$.

(ii) $\sup_{n,t}||e_{nt}||_2^p < \infty$ for some $p > 2$ and conditional expectations $\Sigma_{nt} = E(e_{nt}e'_{nt}|F_{n,t-1})$ are constant matrices.

(iii) Denote Σ_n a function on τ_n with values $\Sigma_{n1}, ..., \Sigma_{nn}$. The sequence $\{\Sigma_n\}$ is assumed to be L_∞-close to some $\Sigma^c \in C[0,1]$.

(iv) $\text{plim}_{A\to\infty} \sup_{n,t} E(||e_{nt}||_2^2 I(||e_{nt}||_2 > A)|F_{n,t-1}) = 0$ and the σ-fields are nested: $F_{nt} \subset F_{n+1,t}$ for $1 \le t \le n$, $n \ge 1$.

The standard implication of condition (ii) is that $||e_{nt}||_2^2$ are uniformly integrable (u.i.) and

$$E(e_{ns}e'_{nt}|F_{n,\max\{s,t\}-1}) = \begin{cases} 0, & s \neq t, \\ \Sigma_{nt}, & s = t. \end{cases} \tag{2.11}$$

Normally this equation will be used in conjunction with the law of iterated expectations, without explicitly mentioning it. One of conditions in Anderson and Kunitomo (1992) is

$$\frac{1}{n}\sum_{t=1}^{n}\Sigma_{tn} \stackrel{p}{\rightarrow} \Sigma.$$

Here the information about heterogeneity contained in Σ_{tn} is forgotten in the limit matrix Σ. Assumption 3(iii) and Theorem 2.1 allow us to prove

$$\frac{1}{n}\sum_{t=1}^{n}\Sigma_{tn} \rightarrow \int_0^1 \Sigma^c(x)dx$$

where the limit expression retains the heterogeneity information. Assumption 3 allows the errors to degenerate in the limit, as in the following example.

Example. Let $e_1,\ e_2,\ ...$ be i.i.d. variables satisfying $\sup_t E|e_t|^p < \infty$ for some $p > 2$. Take any sequence $\{f_n\}$ of vectors $f_n \in \mathbb{R}^n$ such that $\{f_n\}$ is L_∞-close to some $f \in C[0,1]$ (for example, one can take $f_n = (f(1/n), f(2/n), ..., f(1))'$, see Theorem 3.3(b) in Mynbaev (2001)), and put $e_{nt} = f_{nt}e_t$. Let $F_t = \sigma(e_j : j \le t)$ be the least σ-field such that $e_1, ..., e_t$

are F_t-measurable. Then e_{nt} is F_t-measurable, $E(e_{nt}|F_{t-1}) = 0$ by independence, $E|e_{nt}|^p = |f_{nt}|^p E|e_t|^p \le c$, $\Sigma_{nt} = E(e_{nt}^2|F_{t-1}) = f_{nt}^2 Ee_t^2 = \sigma^2 f_{nt}^2$ where $\sigma^2 = Ee_t^2$. It is easy to show that $\{\Sigma_n\}$ is L_∞-close to $\sigma^2 f^2$. Thus, in the limit Σ_n vanishes where f^2 vanishes. $\square$

Assumption 4 (stabilization of relative growth rates of regressors) The limits

$$\kappa_i = \lim_{n\to\infty} \frac{1}{\Delta_n} d_{ni} \in [0,1],\ i = 1, ..., r,\ \kappa_0 = \lim_{n\to\infty} \frac{\sqrt{n}}{\Delta_n} \in [0,1]$$

exist.

Denoting $b_n = \frac{1}{\Delta_n} d_n$, $b =$diag$[\kappa_1, ..., \kappa_r]$, under this assumption one has

$$b = \lim b_n.$$

Assumption 5 (on the initial value) $E||y_0||_2^2 < \infty$.

Our "negative result" requires one more condition:

Assumption 6. $\alpha_1 ... \alpha_r \neq 0$ and $|\int_0^1 H^c (H^c)' dx| \neq 0$.

(2.8) explains why we are interested in studying the vector

$$W_n = (\mathcal{E}_n H_n', \frac{1}{\Delta_n} \mathcal{E}_n M_n'). \tag{2.12}$$

The problem of convergence of W_n in distribution is reduced to a one-dimensional case using a known device (cf. Anderson, 1971, Theorem 7.7.7).

Lemma 2.1. Convergence in distribution

$$\text{vec}W_n \xrightarrow{d} N\left(0, \int_0^1 \Omega_1(x) \otimes \Omega_2(x) dx\right),$$

where $\Omega_1,\ \Omega_2$ are symmetric matrices with square-integrable components, takes place if and only if for any constant matrix C

$$\text{tr}(W_n C) \xrightarrow{d} N\left(0, \int_0^1 \text{tr}[C'\Omega_1(x) C \Omega_2(x)] dx\right). \tag{2.13}$$

Proof. Using

$$\text{tr}(ABC) = (\text{vec}B')'(I \otimes C)\text{vec}A \tag{2.14}$$

we get

$$\text{tr}(W_n C) = c' \text{vec} W_n \tag{2.15}$$

where $c = \mathrm{vec}(C')$. From (2.14) and

$$\mathrm{vec}(AB) = (B' \otimes I)\mathrm{vec}A, \ (A \otimes B)(C \otimes D) = (AC) \otimes (BD) \tag{2.16}$$

we see that

$$\begin{aligned}\int_0^1 \mathrm{tr}[(C'\Omega_1)C\Omega_2]dx &= \int_0^1 c'(I \otimes \Omega_2)\mathrm{vec}(C'\Omega_1)dx \\ &= c'\int_0^1 (I \otimes \Omega_2)(\Omega_1' \otimes I)dxc \\ &= c'\int_0^1 \Omega_1(x) \otimes \Omega_2(x)dxc. \end{aligned} \tag{2.17}$$

(2.15), (2.17) and the Cramér-Wold theorem prove the lemma. □

Partitioning C conformably, $C' = (C_1', C_2')$, and utilizing (2.12) we get

$$\begin{aligned}\mathrm{tr}(W_nC) &= \mathrm{tr}\left(\mathcal{E}_n H_n' C_1 + \frac{1}{\Delta_n}\mathcal{E}_n M_n' C_2\right) \\ &= \mathrm{tr}\left[\sum_{t=1}^n e_{nt}H_{nt}'C_1 + \frac{1}{\Delta_n}\sum_{t=1}^n e_{nt}(P_BAd_nH_n + P_B\mathcal{E}_n)_t'C_2\right] \\ &= \sum_{t=1}^n \left[H_{nt}'C_1 + (P_BAb_nH_n)_t'C_2\right]e_{nt} + \frac{1}{\Delta_n}\sum_{t=1}^n (P_B\mathcal{E}_n)_t'C_2e_{nt}.\end{aligned}$$

Hence, denoting

$$G_{nt} = C_1'H_{nt} + C_2'(P_BAb_nH_n)_t, \ S_{nt} = G_{nt}'e_{nt}, \tag{2.18}$$

$$T_{nt} = \frac{1}{\Delta_n}(P_B\mathcal{E}_n)_t'C_2e_{nt} = \frac{1}{\Delta_n}\sum_{s=1}^{t-1} e_{ns}'B'^{t-1-s}C_2e_{nt} \tag{2.19}$$

we have the decomposition

$$\mathrm{tr}(W_nC) = \sum_{t=1}^n (S_{nt} + T_{nt}). \tag{2.20}$$

S_{nt} and T_{nt} are real-valued m.d.s because $(P_B\mathcal{E}_n)_t$ is $F_{n,t-1}$- measurable.

Important additional notation is introduced before Lemmas 3.3 and 3.4 and Theorems 3.1 and 3.2. For proving convergence in mean the following

Chow-Davidson weak law of large numbers for martingales is useful (see Davidson, 1994, Theorem 19.7).

Theorem 2.2. If $\{X_{nt}, F_{nt}\}$ is an m.d. array, positive constants c_{nt} satisfy

$$\text{(i)} \ \sup_n \sum_{t=1}^{n} c_{nt} < \infty \text{ and (ii) } \lim_{n\to\infty} \sum_{t=1}^{n} c_{nt}^2 = 0 \tag{2.21}$$

and variables X_{nt}/c_{nt} are uniformly integrable, then $E|\sum_{t=1}^{n} X_{nt}| \to 0$.

Note that this theorem trivially extends to vector m.d. arrays.

In comparison with ours, the approach of Anderson and Kunitomo (1992) is more statistical (some characteristics of the limiting distribution are estimated from data, while in our approach they follow from the assumptions) and probabilistic (they use truncation of variables which is a nonlinear operation and does not go along with the functional-theoretical tools used here). Following their lead, among different versions of martingale central limit theorems (CLT) we choose the format suggested by Dvoretzky (1972), for the simple reason that it allows $\sigma^2 = 0$. However, this technical simplification does not make redundant the analysis of the singular case (see our Theorems 3.1-3.3). Anderson and Kunitomo do not do such analysis.

Theorem 2.3 (Dvoretzky CLT) If $\{X_{nt}, F_{nt}\}$ is an m.d. array, σ_{nt}^2 denotes $E(X_{nt}^2|F_{n,t-1})$,

$$\text{plim} \sum_{t=1}^{n} \sigma_{nt}^2 = \sigma^2, \tag{2.22}$$

where $\sigma^2 \geq 0$ is a constant, the σ-fields are nested: $F_{nt} \subset F_{n+1,t}$ for $1 \leq t \leq n$, $n \geq 1$ and for any $\varepsilon > 0$

$$\text{plim} \sum_{t=1}^{n} E(X_{nt}^2 I(|X_{nt}| > \varepsilon)|F_{n,t-1}) = 0, \tag{2.23}$$

then

$$\sum_{t=1}^{n} X_{nt} \xrightarrow{d} N(0, \sigma^2).$$

The original Dvoretzky paper misses the requirement that σ-fields should be nested, see Hall and Heyde (1980) for details.

3 Main Results

The plan is, naturally, to study convergence of the N- and D-factors. In the next lemma ν and μ are arbitrary sets of indices and, as before, matrices in a sequence are of the same size. Recall that "u.i." means uniformly integrable.

Lemma 3.1. (i) For a sequence $\{X_n : n \in \nu\}$ of random matrices uniform integrability of (i,j)th elements $\{X_{nij}\}$ for all i,j is equivalent to uniform integrability of $\{||X_n||_2\}$.

(ii) If variables $||X_n||_2^p$, $n \in \nu$, are u.i., $||Y_m||_2^q$, $m \in \mu$, have uniformly bounded L_1-norms and $p < \infty$, then a double-index family $\{X_n Y_m : n \in \nu,\ m \in \mu\}$ is u.i.

(iii) If vectors X_m, $m \in \mu$, are u.i. and for each $n \in \nu$ $\{\alpha_{nm} : m \in \mu_n\}$ is a set of constant matrices satisfying $\mu_n \subset \mu$, $\alpha = \sup_n \sum_{m \in \mu_n} ||\alpha_{nm}||_2 < \infty$, then the family $\left\{\sum_{m \in \mu_n} \alpha_{nm} X_n : n \in \nu\right\}$ is u.i.

(iv) For $\{X_n\}$ a sequence of random matrices the following conditions are equivalent: (1) all elements of $X_n' X_n$ are u.i., (2) variables $||X_n||_2^2 = \mathrm{tr}(X_n' X_n)$ are u.i.

Proof. It is easy to prove the lemma using the next characterization (see Davidson, 1994, Theorem 12.9): $\{X_n\}$ is u.i. if and only if $\sup_n E|X_n| < \infty$ and for any $\varepsilon > 0$ there is a $\delta > 0$ such that for all events A of probability $P(A) < \delta$ one has $\sup_n E|X_n| I(A) < \varepsilon$.

By equivalence of any two norms on a finite-dimensional space

$$c_1 E||X_n||_2 I(A) \le \sum_{i,j} E|X_{nij}| I(A) \le c_2 E||X_n||_2 I(A)$$

which implies (i). To prove (ii), one has to apply the above characterization to $||X_n||_2^p$ and use the Hölder inequality:

$$E||X_n Y_m||_2 I(A) \le (E||X_n||_2^p I(A))^{1/p} (E||Y_m||_2^q)^{1/q} \le \varepsilon \sup_m (E||Y_m||_2^q)^{1/q}.$$

(iii) follows from

$$E \left\| \sum_{m \in \mu_n} \alpha_{nm} X_m \right\|_2 I(A) \le \alpha \sup_m E \left\| X_m \right\|_2 I(A) \le \alpha \varepsilon.$$

Let us prove (iv). If all elements of $X_n' X_n$ are u.i., then so are the elements of the main diagonal and by (iii) $||X_n||_2^2 = \mathrm{tr}(X_n' X_n)$ is u.i. Conversely, let

$||X_n||_2^2$ be u.i. and let $\delta > 0$ be such that $E||X_n||_2^2 I(A) < \varepsilon$ for all A satisfying $P(A) < \delta$. Then for the (i,j)th element of $X_n' X_n$ we have

$$E\left|\sum_l X_{nli} X_{nlj}\right| I(A) \le E||X_n||_2^2 I(A) < \varepsilon$$

which is what we want. $\square$

In the next lemma we study the behavior of two auxiliary random vectors

$$U_n = \frac{1}{\Delta_n} \sum_{t=1}^n X_{nt} (P_B \mathcal{E}_n)_t'$$

and

$$V_n = \frac{1}{n} \sum_{t=1}^n X_{nt} (P_B \mathcal{E}_n)_t (P_B \mathcal{E}_n)_t' = \frac{1}{n} \sum_{t=1}^n X_{nt} \sum_{k,l=1}^{t-1} B^{t-1-k} e_{nk} e_{nl}' B'^{t-1-l}. \quad (3.1)$$

L_p- lim denotes the limit in mean of order p.

Lemma 3.2. Let Assumptions 1 and 3 hold. Then

(a) If $\{X_n\}$ is vector-valued and L_2-close to $X^c \in L_2$, then

$$L_2\text{- }\lim U_n = 0. \quad (3.2)$$

(b) If $\{X_n\}$ is L_∞-close to $X^c \in C[0,1]$, then

$$L_1\text{- }\lim V_n = \lim EV_n = \int_0^1 X^c(x) \Xi(x) dx \quad (3.3)$$

where $\Xi(x) = \sum_{s=0}^{\infty} B^s \Sigma^c(x) B'^s$.

Proof. (a) Since $X_{nt}' X_{nt}$ is a scalar, we have

$$\begin{aligned} E||U_n||_2^2 &= E\mathrm{tr}(U_n' U_n) = \frac{1}{\Delta_n^2} \sum_{s,t=1}^n E\mathrm{tr}[(P_B \mathcal{E}_n)_t X_{nt}' X_{ns} (P_B \mathcal{E}_n)_s'] \\ &= \frac{1}{\Delta_n^2} \sum_{s,t=1}^n X_{nt}' X_{ns} E\mathrm{tr}\left(\sum_{k=1}^{t-1} B^{t-1-k} e_{nk} \sum_{l=1}^{s-1} e_{nl}' B'^{t-1-l}\right) \\ &= \frac{1}{\Delta_n^2} \sum_{t=1}^n ||X_{nt}||_2^2 \mathrm{tr}\left(\sum_{k=1}^{t-1} B^{t-1-k} \Sigma_{nk} B'^{t-1-k}\right) \\ &= \frac{1}{\Delta_n^2} \sum_{t=1}^n ||X_{nt}||_2^2 \mathrm{tr}(R_{B,B'} \Sigma_n)_t. \end{aligned}$$

By Theorem 2.1(iii) and Assumption 3(ii) $||R_{B,B'}\Sigma_n||_\infty \le c$, so

$$E||U_n||_2^2 \le \frac{c}{\Delta_n^2}\sum_{t=1}^n ||X_{nt}||_2^2 \to 0$$

which proves (3.2).

(b) By orthogonality (2.11)

$$EV_n = \frac{1}{n}\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1} B^{t-1-k}\Sigma_{nk}B'^{t-1-k} = \sum_{t=1}^n \frac{1}{\sqrt{n}}X_{nt}\left(R_{B,B'}\frac{1}{\sqrt{n}}\Sigma_n\right)_t. \tag{3.4}$$

By Theorem 2.1(vi) $\left\{\frac{1}{\sqrt{n}}X_n\right\}$ is L_2-close to X^c and $\left\{\frac{1}{\sqrt{n}}\Sigma_n\right\}$ is L_2-close to Σ^c. By Theorem 2.1(iii) $\left\{R_{B,B'}\frac{1}{\sqrt{n}}\Sigma_n\right\}$ is L_2-close to Ξ. Thus the second equation in (3.3) follows from (3.4) and Theorem 2.1(ii).

Before proving the other part of (3.2), we need to reveal the m.d. structure of the difference $V_n - EV_n$. From (3.1) and (3.4) we have

$$\begin{aligned} V_n - EV_n &= \frac{1}{n}\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1}B^{t-1-k}(e_{nk}e'_{nk}-\Sigma_{nk})B'^{t-1-k} \\ &+\frac{1}{n}\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1}\sum_{l=1}^{k-1}[B^{t-1-k}e_{nk}e'_{nl}B'^{t-1-l} \\ &+B^{t-1-l}e_{nl}e'_{nk}B'^{t-1-k}]. \end{aligned} \tag{3.5}$$

Here each pair (k,l) such that $1 \le l < k \le t-1$ is matched by another pair with $1 \le k < l \le t-1$. In the second pair k and l are switched places. Changing summation order in the first big sum in (3.5) we get

$$\begin{aligned} &\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1}B^{t-1-k}(e_{nk}e'_{nk}-\Sigma_{nk})B'^{t-1-k} \\ =& \sum_{k=1}^{n-1}\sum_{t=k+1}^n X_{nt}B^{t-1-k}(e_{nk}e'_{nk}-\Sigma_{nk})B'^{t-1-k}. \end{aligned} \tag{3.6}$$

A part of the second sum in (3.5) can be rearranged as follows:

$$
\begin{aligned}
& \sum_{t=1}^{n} X_{nt} \sum_{k=1}^{t-1} \sum_{l=1}^{k-1} B^{t-1-k} e_{nk} e'_{nl} B'^{t-1-l} \\
= & \sum_{k=1}^{n-1} \sum_{t=k+1}^{n} X_{nt} B^{t-1-k} e_{nk} \left(\sum_{l=1}^{k-1} e'_{nl} B'^{k-1-l} \right) B'^{t-k} \\
= & \sum_{k=1}^{n-1} \sum_{t=k+1}^{n} X_{nt} B^{t-1-k} e_{nk} (P_B \mathcal{E}_n)'_k B'^{t-k}. \qquad (3.7)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \sum_{t=1}^{n} X_{nt} \sum_{k=1}^{t-1} \left(\sum_{l=1}^{k-1} B^{t-1-l} e_{nl} \right) e'_{nk} B'^{t-1-k} \\
= & \sum_{k=1}^{n-1} \sum_{t=k+1}^{n} X_{nt} B^{t-k} \left(\sum_{l=1}^{k-1} B^{k-1-l} e_{nl} \right) e'_{nk} B'^{t-1-k} \\
= & \sum_{k=1}^{n-1} \sum_{t=k+1}^{n} X_{nt} B^{t-k} (P_B \mathcal{E}_n)_k e'_{nk} B'^{t-1-k}. \qquad (3.8)
\end{aligned}
$$

(3.6), (3.7) and (3.8) are summarized in

$$
V_n - EV_n = \sum_{k=1}^{n-1} Y_{nk}
$$

where

$$
\begin{aligned}
Y_{nk} = & \frac{1}{n} \sum_{t=k+1}^{n} X_{nt} [B^{t-1-k} (e_{nk} e'_{nk} - \Sigma_{nk}) B'^{t-1-k} \\
& + B^{t-1-k} e_{nk} (P_B \mathcal{E}_n)'_k B'^{t-k} + B^{t-k} (P_B \mathcal{E}_n)_k e'_{nk} B'^{t-1-k}]. \qquad (3.9)
\end{aligned}
$$

Since $(P_B \mathcal{E}_n)_k$ is $F_{n,k-1}$-measurable, $\{Y_{nk}\}$ is clearly a vector m.d. array. The numbers $c_{nt} = 1/n$, $t = 1, ..., n$, satisfy conditions of Theorem 2.2.

By Assumption 3(ii) the family $\{||e_{nt}||_2^2\}$ is u.i., so by the equivalent characterization

$$
\lim_{P(A) \to 0} \sup_{n,t} E||e_{nt}||_2^2 I(A) = 0, \ \sup_{n,t} ||e_{nt}||_2^2 < \infty.0
$$

Hence, by Lemma 3(ii) the family $\{e_{nk}e'_{nl}\}$ is u.i. Next we apply Assumption 1 and Lemma 3(iii) to conclude that the products

$$(P_B\mathcal{E}_n)_k e'_{nk} = \left(\sum_{l=1}^{k-1} B^{k-1-l} e_{nl}\right) e'_{nk}.$$

are u.i. Therefore the family consisting of those products and $e_{nk}e'_{nk} - \Sigma_{nk}$ is u.i. Finally, the variables Y_{nk}/c_{nt} are u.i. by Lemma 3(iii), because $||X_{nt}||_\infty \leq \infty$ and

$$\sum_{t=k+1}^{n} ||X_{nt}||_2[||B^{t-1-k}||_2^2 + 2||B^{t-1-k}||_2||B^{t-k}||_2] \leq c\sum_{s=0}^{\infty} ||B^s||_2^2 < \infty.$$

Thus Theorem 2.2 yields $E||V_n - EV_n||_2 \to 0$ which completes the proof. □

Denote $J = (I - B)^{-1}Ab$ and

$$\Omega_0(x) = \begin{pmatrix} H^c(H^c)' & H^c(H^c)'J' \\ JH^c(H^c)' & JH^c(H^c)'J' \end{pmatrix}.$$

Mynbaev's (2006a) explanation of the next lemma is that $\sum_t S_{nt}^2$ is responsible mainly for the contribution of the exogenous part, $\sum_t T_{nt}^2$ is responsible mainly for the contribution of the autoregressive part, and $\sum_t S_{nt}T_{nt}$ controls interaction between the two.

Lemma 3.3. Under Assumptions 1-4

$$\lim \sum_{t=1}^{n} E(S_{nt}^2|F_{n,t-1}) = \text{tr}\int_0^1 C'\Omega_0(x)C\Sigma^c(x)dx, \quad (3.10)$$

$$L_2\text{-}\lim \sum_{t=1}^{n} E(S_{nt}T_{nt}|F_{n,t-1}) = 0, \quad (3.11)$$

$$L_1\text{-}\lim \sum_{t=1}^{n} E(T_{nt}^2|F_{n,t-1}) = \kappa_0^2\text{tr}\int_0^1 C_2'\Xi(x)C_2\Sigma^c(x)dx. \quad (3.12)$$

Proof. From (2.11) and (2.18) we see that

$$\sum_{t=1}^{n} E(S_{nt}^2|F_{n,t-1}) = \sum_{t=1}^{n} G'_{nt}E(e_{nt}e'_{nt}|F_{n,t-1})G_{nt} = \sum_{t=1}^{n} G'_{nt}\Sigma_{nt}G_{nt}.$$

By Assumptions 2, 4 and Theorem 2.1(v) $\{Ab_nH_n\}$ is L_2-close to AbH^c. Assumption 1 and Theorem 2.1(iii) therefore imply

$$P_BAb_nH_n \text{ is } L_2\text{-close to } (I-B)^{-1}AbH^c = JH^c. \tag{3.13}$$

Hence, by Theorem 2.1, items (iv) and (v)

$$G_n = C_1'H_n + C_2'P_BAb_nH_n \text{ is } L_2\text{-close to } G^c \equiv (C_1' + C_2'J)H^c. \tag{3.14}$$

Since $\{\Sigma_n\}$ is L_∞-close to Σ^c, by Theorem 2.1(ii)

$$\sum_{t=1}^{n} G_{nt}'\Sigma_{nt}G_{nt} \to \int_0^1 (G^c)'\Sigma^c G^c dx.$$

Note that

$$G^c = (C_1', C_2')\begin{pmatrix} H^c \\ JH^c \end{pmatrix} = C'\begin{pmatrix} H^c \\ JH^c \end{pmatrix},\ G^c(G^c)' = C'\Omega_0 C,$$

so

$$\int_0^1 (G^c)'\Sigma^c G^c dx = \operatorname{tr}\int_0^1 G^c(G^c)'\Sigma^c dx = \operatorname{tr}\int_0^1 C'\Omega_0 C\Sigma^c dx$$

and (3.10) follows.

Using definitions (2.18), (2.19) and (2.3) rearrange

$$\begin{aligned}\sum_{t=1}^{n} E(S_{nt}T_{nt}|F_{n,t-1}) &= \frac{1}{\Delta_n}\sum_{t=1}^{n} G_{nt}'E(e_{nt}e_{nt}'|F_{n,t-1})C_2'(P_B\mathcal{E}_n)_t \\ &= \frac{1}{\Delta_n}\sum_{t=1}^{n} G_{nt}'\Sigma_{nt}C_2'(P_B\mathcal{E}_n)_t \\ &= \frac{1}{\Delta_n}\operatorname{tr}\sum_{t=1}^{n} C_2\Sigma_{nt}G_{nt}(P_B\mathcal{E}_n)_t'.\end{aligned}$$

This type of variable appeared in Lemma 3.2(a) with $X_{nt} = C_2\Sigma_{nt}G_{nt}$. By (3.14), Assumption 3 and Theorem 2.1(v) $\{X_n\}$ is L_2-close to $C_2\Sigma^cG^c$, so by Lemma 3.2(a) (3.11) is true.

Since $(P_B\mathcal{E}_n)_t$ is $F_{n,t-1}$-measurable, (2.19) implies

$$\begin{aligned}\sum_{t=1}^{n} E(T_{nt}^2|F_{n,t-1}) &= \frac{1}{\Delta_n^2}\sum_{t=1}^{n}(P_B\mathcal{E}_n)_t' C_2 E(e_{nt}e_{nt}'|F_{n,t-1})C_2'(P_B\mathcal{E}_n)_t \\ &= \frac{1}{\Delta_n^2}\mathrm{tr}\sum_{t=1}^{n}(P_B\mathcal{E}_n)_t' C_2\Sigma_{nt}C_2'(P_B\mathcal{E}_n)_t \\ &= \frac{n}{\Delta_n^2}\mathrm{tr}\frac{1}{n}\sum_{t=1}^{n} C_2\Sigma_{nt}C_2'(P_B\mathcal{E}_n)_t(P_B\mathcal{E}_n)_t'.\end{aligned}$$

Here $X_n = C_2\Sigma_n C_2'$ is L_∞-close to $C_2\Sigma^c C_2'$ by Assumption 3 and Theorem 2.1(v). Therefore (3.12) follows from Lemma 3.2(b) and Assumption 4.□

Put

$$\Omega_1(x) = \Omega_0(x) + \begin{pmatrix} 0 & 0 \\ 0 & \kappa_0^2\Xi \end{pmatrix} = \begin{pmatrix} H^c(H^c)' & H^c(H^c)'J' \\ JH^c(H^c)' & JH^c(H^c)'J' + \kappa_0^2\Xi \end{pmatrix},$$

$$\sigma^2 = \mathrm{tr}\int_0^1 C'\Omega_1 C\Sigma^c dx. \tag{3.15}$$

Lemma 3.4. If Assumptions 1 through 4 hold, then for any constant matrix C

$$\mathrm{tr}(W_n C) \xrightarrow{d} N(0,\sigma^2). \tag{3.16}$$

Proof. We are going to apply Theorem 2.3. According to (2.20) we need to consider $X_{nt} = S_{nt} + T_{nt}$. By Lemma 3.3 we have a stronger statement than (2.22):

$$L_1\text{- }\lim\sum_{t=1}^{n}\sigma_{nt}^2 = \mathrm{tr}\int_0^1 [C'\Omega_0 C\Sigma^c + \kappa_0^2 C_2'\Xi C_2\Sigma^c]dx = \sigma^2. \tag{3.17}$$

The proof of (2.23) is a little longer. We need to study properties of $\phi_{nt} = \frac{1}{\Delta_n}||(P_B\mathcal{E}_n)_t||_2$. Obviously, ϕ_{nt} is $F_{n,t-1}$-measurable and by Lemma 3.2(b) and Assumption 4

$$\begin{aligned}L_1\text{- }\lim\sum_{t=1}^{n}\phi_{nt}^2 &= L_1\text{- }\lim\frac{n}{\Delta_n^2}\mathrm{tr}\frac{1}{n}\sum_{t=1}^{n}(P_B\mathcal{E}_n)_t(P_B\mathcal{E}_n)_t' \\ &= \kappa_0^2\mathrm{tr}\int_0^1 \Xi(x)dx.\end{aligned} \tag{3.18}$$

By the Chebyshev inequality for any $\delta > 0$

$$EI(\phi_{nt} > \delta) \le \frac{1}{\delta\Delta_n} E\left\|\sum_{k=1}^{t-1} B^{t-1-k} e_{nk}\right\|_2 \le \frac{c_1}{\delta\Delta_n}. \qquad (3.19)$$

By Minkowski inequality and Assumption 3(ii)

$$\begin{aligned}(E|\phi_{nt}|^p)^{1/p} &= \frac{1}{\Delta_n}\left(E\left\|\sum_{k=1}^{t-1} B^{t-1-k} e_{nk}\right\|_2^p\right)^{1/p} \\ &\le \frac{1}{\Delta_n}\sum_{k=1}^{t-1}\left\|B^{t-1-k}\right\|_2 \sup_{n.k}(E||e_{nk}||_2^p)^{1/p} \le \frac{c_2}{\Delta_n}.\end{aligned}$$

With $p_1 = p/2$ from the last two bounds we get

$$\begin{aligned}E\phi_{nt}^2 I(\phi_{nt} > \delta) &\le (EI(\phi_{nt} > \delta))^{1/q_1}(E|\phi_{nt}|^{2p_1})^{1/p_1} \\ &\le \frac{c_3}{(\delta\Delta_n)^{1/q_1}\Delta_n^2}. \qquad (3.20)\end{aligned}$$

Since $\{G_n\}$ is L_2-approximable (see (3.14)), there exists $n_0 = n_0(\delta)$ such that

$$\sup_{n\ge 1}||G_n; l_2(\tau_n, M_2)|| < \infty, \ \sup_{n\ge 1}\max_{1\le t\le n}||G_{nt}||_2 \le \delta. \qquad (3.21)$$

Using the last estimate and $|X_{nt}| \le c(||G_{nt}||_2 + \phi_{nt})||e_{nt}||_2$, for any $\delta > 0$ and $n \ge n_0$ we have

$$\begin{aligned}I(|X_{nt}| > \varepsilon) &\le I\left((||G_{nt}||_2 + \phi_{nt})||e_{nt}||_2 > \frac{\varepsilon}{c}\right)[I(||G_{nt}||_2 + \phi_{nt} \le 2\delta) \\ &\quad +I(||G_{nt}||_2 + \phi_{nt} > 2\delta)] \le I\left(||e_{nt}||_2 > \frac{\varepsilon}{2\delta c}\right) + I(\phi_{nt} > \delta).\end{aligned}$$

This together with

$$X_{nt}^2 \le 2(S_{nt}^2 + T_{nt}^2) \le c(||G_{nt}||_2^2 + \phi_{nt}^2)||e_{nt}||_2^2$$

allows us to proceed with proving (2.23):

$$\begin{aligned}&\sum_{t=1}^{n} E(X_{nt}^2 I(|X_{nt}| > \varepsilon)|F_{n,t-1}) \\ \le\ & c\sum_{t=1}^{n}(||G_{nt}||_2^2 + \phi_{nt}^2)E\left(||e_{nt}||_2^2 I\left(||e_{nt}||_2 > \frac{\varepsilon}{2\delta c}\right)|F_{n,t-1}\right) \\ &+c\sum_{t=1}^{n}(||G_{nt}||_2^2 + \phi_{nt}^2)I(\phi_{nt} > \delta)E\left(||e_{nt}||_2^2|F_{n,t-1}\right). \qquad (3.22)\end{aligned}$$

By (3.18) and (3.21)

$$\alpha_n \equiv \sum_{t=1}^{n}(||G_{nt}||_2^2 + \phi_{nt}^2) = O(1)$$

which in combination with Assumption 3(iv) leads to

$$\begin{aligned} & \sum_{t=1}^{n}(||G_{nt}||_2^2 + \phi_{nt}^2)E\left(||e_{nt}||_2^2 I\left(||e_{nt}||_2 > \frac{\varepsilon}{2\delta c}\right)|F_{n,t-1}\right) \\ \leq\ & \alpha_n \sup_{n,t} E\left(||e_{nt}||_2^2 I\left(||e_{nt}||_2 > \frac{\varepsilon}{2\delta c}\right)|F_{n,t-1}\right) \xrightarrow{p} 0,\ \delta \to 0. \end{aligned} \quad (3.23)$$

Further, application of (3.19), (3.20) and (3.21) results in

$$\begin{aligned} & E\sum_{t=1}^{n}(||G_{nt}||_2^2 + \phi_{nt}^2)I(\phi_{nt} > \delta)E\left(||e_{nt}||_2^2|F_{n,t-1}\right) \\ =\ & \sum_{t=1}^{n}||G_{nt}||_2^2 EI(\phi_{nt} > \delta)\mathrm{tr}\Sigma_{nt} + \sum_{t=1}^{n}E\phi_{nt}^2 I(\phi_{nt} > \delta)\mathrm{tr}\Sigma_{nt} \\ \leq\ & \frac{c_1}{\delta\Delta_n} + \frac{c_2 n}{(\delta\Delta_n)^{1/q_1}\Delta_n^2} \to 0,\ n \to \infty, \end{aligned} \quad (3.24)$$

for any $\delta > 0$, because $q_1 < \infty$. The left side of (3.23) can be made small uniformly in n by choosing a small δ. For the selected δ, the left side of (3.24) can be made small by taking n sufficiently large. Then (3.22), (3.23) and (3.24) prove (2.23). By Theorem 3.3 (3.16) follows. □

The next lemma establishes the standard fact that the influence of the initial value in (2.8) and (2.9) is asymptotically negligible.

Lemma 3.5. If Assumptions 1 through 5 hold, then

$$\mathrm{dlim}\mathcal{E}_n Z_n' D_n^{-1} = \mathrm{dlim}(\mathcal{E}_n H_n', \frac{1}{\Delta_n}\mathcal{E}_n M_n'), \quad (3.25)$$

$$L_1\text{- }\lim D_n^{-1}Z_n Z_n' D_n^{-1} = L_1\text{- }\lim\begin{pmatrix} H_n H_n' & \frac{1}{\Delta_n}H_n M_n' \\ \frac{1}{\Delta_n}M_n H_n' & \frac{1}{\Delta_n^2}M_n M_n' \end{pmatrix}, \quad (3.26)$$

assuming that the limits on the right exist.

Proof. (3.25) follows from (2.8), Assumptions 1, 3 and 5 and the bound

$$\begin{aligned} E\left\|\frac{1}{\Delta_n}\mathcal{E}_n\rho_n'\right\|_2 \ &\leq\ \frac{1}{\Delta_n}\sum_{t=1}^{n}E||e_{nt}||_2||y_0||_2||B^{t-1}||_2 \\ &\leq\ \frac{1}{\Delta_n}\sup_{n,t}(E||e_{nt}||_2^2)^{1/2}(E||y_0||_2^2)^{1/2}\sum_{t=1}^{\infty}||B^{t-1}||_2 \to 0. \end{aligned}$$

$\{H_n\}$ satisfies a condition of type (2.2), so

$$\begin{aligned} E\left\|\frac{1}{\Delta_n}H_n\rho_n'\right\|_2 &\leq \frac{1}{\Delta_n}\sum_{t=1}^{n}||H_{nt}||_2E||y_0||_2||B^{t-1}||_2 \\ &\leq \frac{1}{\Delta_n}\left(\sum_{t=1}^{n}||H_{nt}||_2^2\right)^{1/2}\left(\sum_{t=1}^{\infty}||B^{t-1}||_2^2\right)^{1/2}E||y_0||_2 \to 0. \end{aligned}$$

Obviously,

$$E\left\|\frac{1}{\Delta_n^2}\rho_n\rho_n'\right\|_2 \leq \frac{1}{\Delta_n^2}\sum_{t=1}^{\infty}||B^{t-1}||_2^2E||y_0||_2^2 \to 0.$$

By (2.7) and (3.13)

$$\begin{aligned} E\left\|\frac{1}{\Delta_n^2}M_n\rho_n'\right\|_2 &= E\left\|\frac{1}{\Delta_n^2}\sum_{t=1}^{n}[(P_BAd_nH_n)_t+(P_B\mathcal{E}_n)_t]y_0'B'^{t-1}\right\|_2 \\ &\leq \frac{1}{\Delta_n}\sum_{t=1}^{n}\left\|(P_BAb_nH_n)_t||_2E||y_0||_2||B^{t-1}\right\|_2 \\ &\quad+\frac{1}{\Delta_n^2}\sum_{t=1}^{n}\sum_{k=1}^{t-1}||B^{t-1-k}||_2E||e_{nk}y_0||_2||B^{t-1}||_2 \\ &\leq \frac{1}{\Delta_n}\left(\sum_{t=1}^{n}||(P_BAb_nH_n)_t||_2^2\sum_{t=1}^{\infty}||B^{t-1}||_2^2\right)^{1/2}E||y_0||_2 \\ &\quad+\frac{1}{\Delta_n^2}\sup_{n,t}(E||e_{nt}||_2^2)^{1/2}\left(E||y_0||_2^2\right)^{1/2}\left(\sum_{t=1}^{\infty}||B^{t-1}||_2\right)^2 \to 0. \end{aligned}$$

Now (3.26) follows from (2.9) and the last three bounds. □

Denote

$$G=\int_0^1 H^c(H^c)'dx,\ Q=\begin{pmatrix} G & GJ' \\ JG & JGJ'+\kappa_0^2\int_0^1\Xi(x)dx \end{pmatrix}$$

Theorem 3.1 (convergence of the D-factor) If Assumptions 1-5 hold, then

(i) The D-factor converges in $L_1(\Omega)$

$$L_1\text{-}\lim D_n^{-1}Z_nZ_n'D_n^{-1}=Q. \tag{3.27}$$

(ii) Condition $|Q| \neq 0$ is equivalent to a combination of three conditions:

(a) $\kappa_0 > 0$, (b) $|G| \neq 0$, (c) $|\int_0^1 \Xi(x)dx| \neq 0$.

Proof. (i) We consider the blocks of the matrix in (3.26) one by one. By Assumption 2 and Theorem 2.3(ii)

$$\lim_{n\to\infty} H_n H_n' = \lim_{n\to\infty} \sum_{t=1}^{n} H_{nt} H_{nt}' = \int_0^1 H^c (H^c)' dx = G. \qquad (3.28)$$

Denote $F_n = P_B A b_n H_n$. From (2.7), (3.13), Theorem 2.1(ii) and Lemma 3.2(a)

$$\begin{aligned} L_2\text{- lim} \frac{1}{\Delta_n} H_n M_n' &= L_2\text{- lim} \left[H_n F_n' + \frac{1}{\Delta_n} H_n (P_B \mathcal{E}_n)' \right] \\ &= \int_0^1 H^c (H^c)' J' dx = GJ'. \end{aligned}$$

The block in the lower right corner of (3.26) equals

$$\frac{1}{\Delta_n^2} M_n M_n' = F_n F_n' + \frac{1}{\Delta_n} F_n (P_B \mathcal{E}_n)' + \frac{1}{\Delta_n} (P_B \mathcal{E}_n) F_n' + \frac{1}{\Delta_n^2} (P_B \mathcal{E}_n)(P_B \mathcal{E}_n)'.$$

Here by (3.13) and Lemma 3.2(a)

$$\lim F_n F_n' = JGJ', \ L_2\text{- lim} \frac{1}{\Delta_n} F_n (P_B \mathcal{E}_n)' = 0$$

and by Lemma 3.2(b)

$$L_1\text{- lim} \frac{1}{\Delta_n^2} (P_B \mathcal{E}_n)(P_B \mathcal{E}_n)' = \kappa_0^2 \int_0^1 \Xi(x) dx.$$

The proof is complete.

(ii) Suppose $|Q| \neq 0$. If $|G| = 0$, then some row of G is a linear combination of others. Denote the rows $(G)_1, ..., (G)_r$ and suppose $(G)_i = \sum_{j\neq i} c_j (G)_j$. Then $(GJ')_i = \sum_{j\neq i} c_j (GJ')_j$. This means that among the first rows of Q one is a linear combination of others and hence $|Q| = 0$. This proves necessity of (b).

When proving necessity of (a) and (c), we can assume that (b) is true, without loss of generality. By the determinant of a partitioned matrix rule

$$|Q| = |G| \left| JGJ' + \kappa_0^2 \int_0^1 \Xi dx - JGG^{-1}GJ' \right| = \kappa_0^2 |G| \left| \int_0^1 \Xi dx \right|.$$

This equation implies (a) and (c). Sufficiency of (a), (b) and (c) also follows from this equation.□

Theorem 3.2 (convergence of the N-factor) Let Assumptions 1-5 hold. Then

(i) the N-factor converges in distribution

$$\text{vec}(\mathcal{E}_n Z_n' D_n^{-1}) \xrightarrow{d} N\left(0, \int_0^1 \Omega_1(x)\otimes\Sigma^c(x)dx\right). \tag{3.29}$$

(ii) Condition $\left|\int_0^1 \Omega_1(x)\otimes\Sigma^c(x)dx\right| \neq 0$ is equivalent to a set of 3 conditions:

(a) $\kappa_0 > 0$, (b) $|\int_0^1 [H^c(H^c)']\otimes\Sigma^c dx| \neq 0$, (c) $|\int_0^1 \Xi(x)\otimes\Sigma^c(x)dx| \neq 0$.

Proof. (i) Lemma 3.5 reduces convergence of the N-factor to that of W_n. By Lemma 2.1 W_n converges if $tr(W_n C)$ converges for any C. This last convergence has been proved in Lemma 3.4. Lemma 2.1 provides the expression for the variance of the limit because if we denote $\mathcal{H} = H^c(H^c)'$ then

$$\Omega_0\otimes\Sigma^c = \begin{pmatrix} \mathcal{H}\otimes\Sigma^c & (\mathcal{H}\otimes\Sigma^c)(J'\otimes I) \\ (J\otimes I)(\mathcal{H}\otimes\Sigma^c) & (J\otimes I)(\mathcal{H}\otimes\Sigma^c)(J'\otimes I) \end{pmatrix}.$$

The proof of part (ii) is similar to the proof of part (ii) of Theorem 3.1.□

Corollary 3.1. If the conditions of Theorem 3.2 are satisfied and $\kappa_0 = 0$, then, in addition to convergence (3.29), for the partitioning $\mathcal{E}_n Z_n' D_n^{-1} = (U_n, V_n)$, where $U_n = \mathcal{E}_n H_n'$ and $V_n = \frac{1}{\Delta_n}\mathcal{E}_n(Y_n^-)'$, we can assert convergence $U_n \xrightarrow{d} U$, $V_n \xrightarrow{d} V$ where $vecU \sim N\left(0, \int_0^1 [H^c(H^c)']\otimes\Sigma^c dx\right)$and V is proportional to U, $V = UJ'$.

Proof. Convergence of $vecU_n$ and $vecV_n$ follows from $vec\,(\mathcal{E}_n Z_n' D_n^{-1}) = \begin{pmatrix} vecU_n \\ vecV_n \end{pmatrix}$ and (3.29). Denoting $\mathcal{G} = \int_0^1 [H^c(H^c)']\otimes\Sigma^c dx$, we can write the variance matrix in (3.29) as

$$\int_0^1 \Omega_1(x)\otimes\Sigma^c(x)dx = \begin{pmatrix} \mathcal{G} & \mathcal{G}(J'\otimes I) \\ (J\otimes I)\mathcal{G} & (J\otimes I)\mathcal{G}(J'\otimes I) \end{pmatrix}.$$

Equation $V = UJ'$ implies $vecU = (J\otimes I)vecU$, so that $\begin{pmatrix} vecU \\ vecV \end{pmatrix}$ has the same variance. Since a normal vector is uniquely defined by its mean and variance, this proves the corollary.□

Theorem 3.3 Let Assumptions 1-5 hold.

(i) (Convergence of the OLS estimator, case $\kappa_0 > 0$) If $|Q| \neq 0$, then

$$\text{vec}\left((\widehat{\Gamma}_n - \Gamma)D_n\right) \xrightarrow{d} N\left(0, \int_0^1 (Q^{-1}\Omega_1 Q^{-1}) \otimes \Sigma^c dx\right).$$

(b) (Inapplicability of the conventional scheme, case $\kappa_0 = 0$) Let also Assumption 6 hold and $\kappa_0 = 0$. Then by Theorem 3.1 the D-factor converges in mean to Q with $|Q| = 0$, so that the conventional scheme does not work. Moreover, there is no diagonal nonstochastic normalizer $\widetilde{D}_n$ for which the D-factor would converge in probability to a nonsingular matrix:

$$\text{plim}\widetilde{D}_n^{-1} Z_n Z_n' \widetilde{D}^{-1} = \widetilde{Q},\ |\widetilde{Q}| \neq 0. \tag{3.30}$$

Proof. Part (a) is an immediate consequence of Theorems 3.1 and 3.2 and the equation

$$\text{vec}\left((\widehat{\Gamma}_n - \Gamma)D_n\right) = [(D_n^{-1} Z_n Z_n' D_n^{-1})^{-1} \otimes I] vec(\mathcal{E}_n Z_n' D_n^{-1}).$$

Now we address the negative statement of part (b). Suppose that $JGJ' = 0$. Since $I - B$ is one-to-one, we have $AbG(Ab)' = 0$. Denote $\zeta = (\alpha_1\kappa_1, ..., \alpha_r\kappa_r)'$. Direct calculation shows that $AbG(Ab)' = 0$ implies $\zeta G \zeta' = 0$. Since G is positive definite, it follows that $\zeta = 0$. Assumption 6 then implies $b = 0$ which means that all d_{ni} are $o(\sqrt{n})$, $i = 1, ..., r$. But this is possible only when $\kappa_0 = 1$ – which is not what we assume in (b). Thus, our assumption is wrong and in the sequel we can use

$$JGJ' \neq 0. \tag{3.31}$$

Suppose that, contrary to the assertion, the normalizer $\widetilde{D}_n$ exists. To avoid notational clutter, denote Q_n the D-factor from (1.6) and let $K_n = D_n^{-1} Z_n$. Then the N-factor and D-factor become $\mathcal{E}_n Z_n' D_n^{-1} = \mathcal{E}_n K_n'$ and $Q_n = D_n^{-1} Z_n Z_n' D_n^{-1} = K_n K_n'$, respectively. We know that with our normalization

$$\text{plim}Q_n = \text{plim}\begin{pmatrix} H_n H_n' & \frac{1}{\Delta_n} H_n M_n' \\ \frac{1}{\Delta_n} M_n H_n' & \frac{1}{\Delta_n^2} M_n M_n' \end{pmatrix} = Q,\ |Q| = 0. \tag{3.32}$$

All objects in the parallel world (with an alternative normalizer $\widetilde{D}_n$) will be capped with a tilde. By assumption $\widetilde{D}_n$ is diagonal,

$$\widetilde{D}_n = \begin{pmatrix} \widetilde{d}_n & 0 \\ 0 & \widetilde{\Delta}_n I_s \end{pmatrix},\ \widetilde{d}_n = \text{diag}\,[\widetilde{d}_{n1}, ..., \widetilde{d}_{nr}].$$

Letting $C_n = D_n^{-1}\widetilde{D}_n$ we note the relationship between K_n and $\widetilde{K}_n$, $K_n = D_n^{-1}\widetilde{D}_n\widetilde{D}_n^{-1}Z_n = C_n\widetilde{K}_n$, which implies $Q_n = C_n\widetilde{K}_n\widetilde{K}_n'C_n = C_n\widetilde{Q}_nC_n$. We can invoke (3.32) and (3.30) to conclude that

$$\left(\frac{\widetilde{d}_{n1}...\widetilde{d}_{nr}\widetilde{\Delta}_n^s}{d_{n1}...d_{nr}\Delta_n^s}\right)^2 = |C_n|^2 = \frac{|Q_n|}{|\widetilde{Q}_n|} \xrightarrow{p} 0. \tag{3.33}$$

Partitioning $\widetilde{Q}$ conformably, $\widetilde{Q} = \begin{pmatrix} \widetilde{Q}_{11} & \widetilde{Q}_{12} \\ \widetilde{Q}_{21} & \widetilde{Q}_{22} \end{pmatrix}$, from (3.30) we have

$$\text{plim}\widetilde{H}_n\widetilde{H}_n' = \widetilde{Q}_{11}. \tag{3.34}$$

Let $h_n^{(1)}, ..., h_n^{(r)}$ denote the rows of H_n. By construction

$$||h_n^{(1)}||_2 = ... = ||h_n^{(r)}||_2 = 1.$$

If $\widetilde{\lambda}_1, ..., \widetilde{\lambda}_r$ denote the diagonal elements of $\widetilde{Q}_{11}$, then by (3.34)

$$\text{plim}_{n\to\infty}||\widetilde{h}_n^{(i)}||_2^2 = \widetilde{\lambda}_i,\ i = 1, ..., r.$$

There is a link between $\widetilde{h}_n^{(i)}$ and $h_n^{(i)}$, $\widetilde{h}_n^{(i)} = \frac{d_{ni}}{\widetilde{d}_{ni}}h_n^{(i)}$, so the last two equations imply

$$\text{plim}_{n\to\infty}\frac{d_{ni}}{\widetilde{d}_{ni}} = \sqrt{\widetilde{\lambda}_i}. \tag{3.35}$$

After this preparatory work let us suppose that

$$\widetilde{\lambda}_i = 0 \text{ for some } i \tag{3.36}$$

and show that this assumption leads to a contradiction. (3.35) and (3.36) show that

$$\frac{d_{ni}}{\widetilde{d}_{ni}} = o(1). \tag{3.37}$$

Besides, by the Cauchy-Schwartz inequality $\left|\widetilde{h}_n^{(i)}\widetilde{h}_n^{(j)\prime}\right| \le \left\|\widetilde{h}_n^{(i)}\right\|_2 \left\|\widetilde{h}_n^{(j)}\right\|_2 \to 0$ which means that a whole row in $\widetilde{Q}_{11}$ is zero:

$$\lim \widetilde{h}_n^{(i)}\left(\widetilde{h}_n^{(1)\prime}, ..., \widetilde{h}_n^{(r)\prime}\right) = \lim \widetilde{h}_n^{(i)}\widetilde{H}_n' = 0. \tag{3.38}$$

Now we consider two cases.

(1) Suppose that

$$\frac{\Delta_n}{\widetilde{\Delta}_n} \le c. \tag{3.39}$$

There is a link between the ith rows of $\frac{1}{\widetilde{\Delta}_n}\widetilde{H}_n' Y_n^-$ and $\frac{1}{\Delta_n}H_n' Y_n^-$:

$$\frac{1}{\widetilde{\Delta}_n}\widetilde{h}_n^{(i)} Y_n^- = \frac{d_{ni}}{\widetilde{d}_{ni}}\frac{\Delta_n}{\widetilde{\Delta}_n}\frac{1}{\Delta_n} h_n^{(i)} Y_n^-.$$

$\frac{1}{\Delta_n} h_n^{(i)} Y_n^-$ converges in probability (as a part of Q_n), so (3.37) and (3.39) lead to the conclusion that a whole row in $\widetilde{Q}_{12}$ is zero:

$$\text{plim}\frac{1}{\widetilde{\Delta}_n}\widetilde{h}_n^{(i)} Y_n^- = 0. \tag{3.40}$$

Clearly (3.38) and (3.40) contradict (3.30).

(2) Suppose that the opposite of (3.39) is true:

$$\frac{\Delta_{n_k}}{\widetilde{\Delta}_{n_k}} \to \infty$$

along some subsequence $\{n_k\}$. Then by Theorem 3.1 and (3.30)

$$\text{plim}\frac{1}{\Delta_{n_k}^2} Y_{n_k}^- Y_{n_k}^{-\prime} = \text{plim}\left(\frac{\widetilde{\Delta}_{n_k}}{\Delta_{n_k}}\right)^2 \frac{1}{\widetilde{\Delta}_{n_k}^2} Y_{n_k}^- Y_{n_k}^{-\prime} = 0.$$

Since the limit along a subsequence is zero, the limit $JGJ' = \text{plim}\frac{1}{\Delta_n^2} Y_n Y_n^{-1}$ along the whole sequence is also zero. This contradicts (3.31).

The contradiction stems from (3.36). Hence all diagonal elements of $\widetilde{Q}_{11}$ are positive which means that for each i d_{ni} is of the same order as $\widetilde{d}_{ni}$. Then by (3.33) $\frac{\widetilde{\Delta}_n}{\Delta_n} \to 0$ which can be used as above to show that $JGJ' = 0$, again contradicting (3.31). Thus (3.30) is impossible. $\square$

Remark 3.1. Denote $f_n = \max\{d_{n1}, ..., d_{nr}\}$. It is easy to prove that equality $\kappa_0 = 0$ is equivalent to $\sqrt{n} = o(f_n)$.

Remark 3.2. Even in the situation of Theorem 3.3(i) it is possible for the limiting distribution to be degenerate. Let there be only two regressors. Take functions H_1^c and H_2^c and a matrix Σ^c with nonoverlappings supports on $[0,1]$. Then the product $[H^c(H^c)'] \otimes \Sigma^c$ is a null matrix and it can be seen from Theorem 3.2(ii) that the N-factor converges to a degenerate normal vector. If we take the regressors to be of form $\{d_{n2}H_1^c\}$, $\{d_{n2}H_2^c\}$, then the norms of the rows of X_n will have finite limits, Δ_n will equal $\sqrt{n}$ for all large n, leading to $\kappa_0 = 1$. Then using Theorem 3.1(ii) it is straightforward to show that $|Q| \neq 0$.

References

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

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

[3] Anderson, T. W. and N. Kunitomo (1994) Asymptotic robustness of tests of overidentification and predeterminedness, J. of Econometrics, 62, 383-414.

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

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

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

[7] Charemza, W. and D. F. Deadman (1992) New Directions in Econometric Practice: General to Specific Modelling, Cointegration, and Vector Autoregression. Aldershot, Hants, England: E. Elgar; Brookfield, VT, USA.

[8] Chow, Y. S. and H. Teicher (1980) Probability Theory: Independence, Interchangeability, Martingales. 2nd edition, Springer, New York.

[9] Dvoretzky, A. (1972) Asymptotic normality for sums of dependent variables. Proc. Sixth Berkeley Symp. Math. Statist. Probability, University of California Press, 513-555.

[10] Hall, P. and C. C. Heyde (1980) Martingale Limit Theory and Its Application. Academic Press, New York.

[11] Hamilton, J. D. (1994) Time Series Analysis. Princeton, N.J. : Princeton University Press.

[12] Kim, T.-H., S. Pfaffenzeller, T. Rayher, P. Newbold (2003) Testing for linear trend with applications to relative commodity prices, J. of Time Series Analysis, 24, 539-551.

[13] Lütkepohl, H. (1991) Introduction to multiple time series analysis. New York, Springer-Verlag.

[14] 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.

[15] Mynbaev, K. T. (2006a) Asymptotic properties of OLS estimates in autoregressions with bounded or slowly growing deterministic trends, Communications in Statistics, 35, 499-520.

[16] Mynbaev, K. T. (2006b) L_p-approximable sequences of matrices and systems of linear regressions with deterministic regressors and dependent errors (unpublished)

[17] 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.

[18] Nielsen, B. (2005) Strong consistency results for Least Squares estimators in general vector autoregressions with deterministic terms. Econometric Theory, 21, 534-561.

[19] Pötscher, B.M., Prucha, I.R. (1991) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part I: Consistency and approximation concepts. Econometric Reviews 10, 125-216.

[20] 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.

[21] Tanaka, K. (1996) Time Series Analysis: Nonstationary and Noninvertible Distribution Theory. New York: Wiley.