ISSN 2077–9879

Eurasian Mathematical Journal

2018, Volume 9, Number 1

Founded in 2010 by
the L.N. Gumilyov Eurasian National University

in cooperation with
the M.V. Lomonosov Moscow State University
the Peoples' Friendship University of Russia
the University of Padua

Supported by the ISAAC
(International Society for Analysis, its Applications and Computation)
and
by the Kazakhstan Mathematical Society

Published by

the L.N. Gumilyov Eurasian National University
Astana, Kazakhstan

EURASIAN MATHEMATICAL JOURNAL

Aims and Scope

The Eurasian Mathematical Journal (EMJ) publishes carefully selected original research papers in all areas of mathematics written by mathematicians, principally from Europe and Asia. However papers by mathematicians from other continents are also welcome.

From time to time the EMJ publishes survey papers.

The EMJ publishes 4 issues in a year.

The language of the paper must be English only.

The contents of EMJ are indexed in Scopus, Web of Science (ESCI), Mathematical Reviews, MathSciNet, Zentralblatt Math (ZMATH), Referativnyi Zhurnal – Matematika, Math-Net.Ru.

The EMJ is included in the list of journals recommended by the Committee for Control of Education and Science (Ministry of Education and Science of the Republic of Kazakhstan) and in the list of journals recommended by the Higher Attestation Commission (Ministry of Education and Science of the Russian Federation).

Information for the Authors

Submission. Manuscripts should be written in LaTeX and should be submitted electronically in DVI, PostScript or PDF format to the EMJ Editorial Office via e-mail (eurasianmj@yandex.kz).

When the paper is accepted, the authors will be asked to send the tex-file of the paper to the Editorial Office.

The author who submitted an article for publication will be considered as a corresponding author. Authors may nominate a member of the Editorial Board whom they consider appropriate for the article. However, assignment to that particular editor is not guaranteed.

Copyright. When the paper is accepted, the copyright is automatically transferred to the EMJ. Manuscripts are accepted for review on the understanding that the same work has not been already published (except in the form of an abstract), that it is not under consideration for publication elsewhere, and that it has been approved by all authors.

Title page. The title page should start with the title of the paper and authors' names (no degrees). It should contain the Keywords (no more than 10), the Subject Classification (AMS Mathematics Subject Classification (2010) with primary (and secondary) subject classification codes), and the Abstract (no more than 150 words with minimal use of mathematical symbols).

Figures. Figures should be prepared in a digital form which is suitable for direct reproduction.

References. Bibliographical references should be listed alphabetically at the end of the article. The authors should consult the Mathematical Reviews for the standard abbreviations of journals' names.

Authors' data. The authors' affiliations, addresses and e-mail addresses should be placed after the References.

Proofs. The authors will receive proofs only once. The late return of proofs may result in the paper being published in a later issue.

Offprints. The authors will receive offprints in electronic form.

Publication Ethics and Publication Malpractice

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the EMJ implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The EMJ follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct ($http : //publicationethics.org/files/u2/New_Code.pdf$). To verify originality, your article may be checked by the originality detection service CrossCheck http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the EMJ.

The Editorial Board of the EMJ will monitor and safeguard publishing ethics.

The procedure of reviewing a manuscript, established by the Editorial Board of the Eurasian Mathematical Journal

1. **Reviewing procedure**

1.1. All research papers received by the Eurasian Mathematical Journal (EMJ) are subject to mandatory reviewing.

1.2. The Managing Editor of the journal determines whether a paper fits to the scope of the EMJ and satisfies the rules of writing papers for the EMJ, and directs it for a preliminary review to one of the Editors-in-chief who checks the scientific content of the manuscript and assigns a specialist for reviewing the manuscript.

1.3. Reviewers of manuscripts are selected from highly qualified scientists and specialists of the L.N. Gumilyov Eurasian National University (doctors of sciences, professors), other universities of the Republic of Kazakhstan and foreign countries. An author of a paper cannot be its reviewer.

1.4. Duration of reviewing in each case is determined by the Managing Editor aiming at creating conditions for the most rapid publication of the paper.

1.5. Reviewing is confidential. Information about a reviewer is anonymous to the authors and is available only for the Editorial Board and the Control Committee in the Field of Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan (CCFES). The author has the right to read the text of the review.

1.6. If required, the review is sent to the author by e-mail.

1.7. A positive review is not a sufficient basis for publication of the paper.

1.8. If a reviewer overall approves the paper, but has observations, the review is confidentially sent to the author. A revised version of the paper in which the comments of the reviewer are taken into account is sent to the same reviewer for additional reviewing.

1.9. In the case of a negative review the text of the review is confidentially sent to the author.

1.10. If the author sends a well reasoned response to the comments of the reviewer, the paper should be considered by a commission, consisting of three members of the Editorial Board.

1.11. The final decision on publication of the paper is made by the Editorial Board and is recorded in the minutes of the meeting of the Editorial Board.

1.12. After the paper is accepted for publication by the Editorial Board the Managing Editor informs the author about this and about the date of publication.

1.13. Originals reviews are stored in the Editorial Office for three years from the date of publication and are provided on request of the CCFES.

1.14. No fee for reviewing papers will be charged.

2. **Requirements for the content of a review**

2.1. In the title of a review there should be indicated the author(s) and the title of a paper.

2.2. A review should include a qualified analysis of the material of a paper, objective assessment and reasoned recommendations.

2.3. A review should cover the following topics:

- compliance of the paper with the scope of the EMJ;

- compliance of the title of the paper to its content;

- compliance of the paper to the rules of writing papers for the EMJ (abstract, key words and phrases, bibliography etc.);

- a general description and assessment of the content of the paper (subject, focus, actuality of the topic, importance and actuality of the obtained results, possible applications);

- content of the paper (the originality of the material, survey of previously published studies on the topic of the paper, erroneous statements (if any), controversial issues (if any), and so on);

- exposition of the paper (clarity, conciseness, completeness of proofs, completeness of bibliographic references, typographical quality of the text);

- possibility of reducing the volume of the paper, without harming the content and understanding of the presented scientific results;

- description of positive aspects of the paper, as well as of drawbacks, recommendations for corrections and complements to the text.

2.4. The final part of the review should contain an overall opinion of a reviewer on the paper and a clear recommendation on whether the paper can be published in the Eurasian Mathematical Journal, should be sent back to the author for revision or cannot be published.

Web-page

The web-page of EMJ is www.emj.enu.kz. One can enter the web-page by typing Eurasian Mathematical Journal in any search engine (Google, Yandex, etc.). The archive of the web-page contains all papers published in EMJ (free access).

Subscription

For Institutions

- US$ 200 (or equivalent) for one volume (4 issues)
- US$ 60 (or equivalent) for one issue

For Individuals

- US$ 160 (or equivalent) for one volume (4 issues)
- US$ 50 (or equivalent) for one issue.

The price includes handling and postage.

The Subscription Form for subscribers can be obtained by e-mail:

eurasianmj@yandex.kz

The Eurasian Mathematical Journal (EMJ)
The Editorial Office
The L.N. Gumilyov Eurasian National University
Building no. 3
Room 306a
Tel.: +7-7172-709500 extension 33312
13 Kazhymukan St
010008 Astana
Kazakhstan

EURASIAN MATHEMATICAL JOURNAL
ISSN 2077-9879
Volume 9, Number 1 (2018), 40 – 68

LEAST SQUARES ESTIMATOR ASYMPTOTICS FOR VECTOR AUTOREGRESSIONS WITH DETERMINISTIC REGRESSORS

K.T. Mynbaev

Communicated by S. Sagitov

Key words: time-series regression, asymptotic distribution, OLS estimator, polynomial trend, deterministic regressor.

AMS Mathematics Subject Classification: 46N30, 97K80.

Abstract. We consider a mixed vector autoregressive model with deterministic exogenous regressors and an autoregressive matrix that has characteristic roots inside the unit circle. The errors are $(2+\epsilon)$-integrable martingale differences with heterogeneous second-order conditional moments. The behavior of the ordinary least squares (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 diagonal normalizer.

1 Introduction

We consider the asymptotic distribution of the OLS estimator of matrix parameters A and B in the vector autoregression

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

where y_t, x_t, and e_t are random vectors. The x_t are assumed to be exogenous (determined outside the system) and known, the y_t are observed and the unobserved errors e_t are martingale differences. For the general theory of vector autoregressions one can consult [16], [14], and [9]. Models with deterministic x_t are of special interest in some applications. In particular, the OLS asymptotics for autoregressions with polynomial trends has been a long-standing issue (see [14, chapter 16] and our Section 2.9).

Our purpose is to develop an asymptotic theory of vector autoregressions (1.1) with assumptions on the deterministic regressors general enough to include various special cases (e.g., polynomial and logarithmic trends), with possible discontinuities arising in the theory of structural breaks. The framework is based on the L_p-approximability theory [17] previously applied to a scalar autoregressive model [18]

$$y_t = \alpha x_t + \beta y_{t-1} + e_t,\ t = 1, ..., n, \tag{1.2}$$

(α and β are real parameters). We call (1.2) a *basic model*. The results presented here are partially new even in this special case and show that vector autoregressions require quite different techniques than spatial models [21], [22] and static models with slowly varying regressors [20]. Because of space limitations, in this paper we consider only the stable case (when the characteristic roots of B lie inside the unit circle $|\lambda| < 1$). [28] and [24] give an idea of the state of affairs in the unstable case.

So far, the paper [3] has been the most advanced in the stable case. They allow for both stochastic and deterministic exogenous regressors. When only deterministic regressors are involved, our result has a wider area of applicability, under less stringent assumptions. Anderson and Kunitomo impose, among others, three *infinite* series of conditions, and removing them has been one of our goals. A detailed comparison of the two results, along with some methodological remarks, will be provided in the end of Section 2.

To investigate a model with polynomial trends, [27] have suggested a linear transformation. However, that transformation uses unknown coefficients and therefore is not feasible. Our asymptotic result in case of polynomial trends is negative in the sense to be described later. Moreover, in Section 2.9 we show that the derivation of the asymptotic distribution of the OLS estimator for an autoregression with a linear trend given in [14, Section 16.3] contains an error.

There have been other suggestions to model deterministic regressors. One approach is appropriate for studying consistency of the OLS estimator; see [24] for the details and history. Another has been proposed in [6] in the context of nonlinear models. Finally, Phillips [25] has employed properties of slowly varying functions [7] to model asymptotically collinear regressors. Mynbaev [20] has shown that all sequences of weights arising in the Phillips approach are L_2-approximable.

To explain the main results we need some notation. By putting equations (1.1) side by side we can write them in a matrix form

$$Y_n = AX_n + BY_n^- + \mathcal{E}_n \tag{1.3}$$

where

$$Y_n = (y_1, ..., y_n),\ X_n = (x_1, ..., x_n),\ Y_n^- = (y_0, ..., y_{n-1}),\ \mathcal{E}_n = (e_1, ..., e_n). \tag{1.4}$$

Let $s(M)$ denote the size of a matrix M (a pair of its dimensions). We suppose that $s(Y_n) = s(\mathcal{E}_n) = s \times n$, $s(X_n) = r \times n$, $s(A) = s \times r$, $s(B) = s \times s$. Denoting $\Gamma = (A, B),\ Z_n = \begin{pmatrix} X_n \\ Y_n^- \end{pmatrix}$ we write (1.3) as

$$Y_n = \Gamma Z_n + \mathcal{E}_n. \tag{1.5}$$

The formula for the OLS estimator of Γ is [16]

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

A basic fact about OLS estimators is that they should be centered and normalized to obtain convergence in distribution. Let D_n be some nonsingular, diagonal (possibly stochastic) matrix, called a *normalizer*. Then (1.5) and (1.6) imply

$$\left(\widehat{\Gamma}_n - \Gamma\right) D_n = \mathcal{E}_n Z_n' D_n^{-1} (D_n^{-1} Z_n Z_n' D_n^{-1})^{-1}. \tag{1.7}$$

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 a normal vector, and

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

Then, if $\det Q \neq 0$, convergence in distribution of $\left(\widehat{\Gamma}_n - \Gamma\right) D_n$ follows trivially from (1.7) and the conventional scheme.

Qualitatively, the contribution of this paper consists of two statements.

Positive statement. A nonstochastic normalizer D_n, which takes into account the relative rates of growth of the exogenous and endogenous regressors, is chosen in such a way that both the N-factor and D-factor converge. The convergence is proved in the class of regressors x_t that become L_2-approximable upon normalization. This class includes constants, polynomial and logarithmic trends and slowly varying regressors. Not surprisingly, all sets of regressors fall into one of two categories: either $\det Q \neq 0$ or $\det Q = 0$. The condition $\det Q \neq 0$ is completely characterized in terms of the growth rates of the regressors. In case $\det Q \neq 0$, naturally, there is asymptotic normality of the OLS estimator. The heterogeneity of the errors may result in a degenerate OLS asymptotics even when Q is nonsingular.

Negative statement. In the asymptotically collinear case (when $\det Q = 0$ with our normalizer) we prove that there is no diagonal, possibly stochastic, normalizer that would render the conventional scheme feasible. Thus, in terms of the diagonal normalizer choice, our positive statement is final.

The negative answer means that the conventional scheme has to be modified to deal with the asymptotically collinear case. We think that this can be done following the ideas of [25]. The corresponding result would be very interesting because Phillips considers only static models. In addition to the negative statement, in Section 2.9 we prove that using non-diagonal normalizers does not work for the basic model (1.2). A full study of the issue is beyond the scope of this paper.

The main results are stated and the ideas are explained in Section 2. Section 2 is concluded with examples. All proofs are given in Section 3.

2 Main results

2.1 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, trace and vectorization we use can be found in [16]. $\det A$ is also denoted $|A|$. $\|x\|_2$ denotes the Euclidean norm of a vector x.

The choice of the normalizer is of principal importance. If the lag is absent from the basic model, $y_t = \alpha x_t + e_t$, the best normalizer is known to be $D_n = \|x\|_2$, where $x = (x_1, ..., x_n)'$, see [1, Theorem 2.6.1] and [2, Theorem 3.5.4]. We call $\|x\|_2$ the *growth rate* of the regressor x_t. When, on the other hand, there is no exogenous regressor, $y_t = \beta y_{t-1} + e_t$, the normalizer is $\sqrt{n}$ [1, Theorem 5.5.6]. The question is: how interaction of x_t and y_{t-1} in the basic model with nonzero α, β is to be reflected in the normalizer? The answer has been given in [18]. Let us call $\sqrt{n}$ the *borderline rate*. If the growth rate of the regressor x_t is equal to or lower than the borderline rate, then $\hat{\beta} - \beta$ must be normalized by the borderline rate. If $\|x\|_2$ exceeds $\sqrt{n}$, then $\hat{\beta} - \beta$ must be normalized by the growth rate of x_t. $\hat{\alpha} - \alpha$ must always be normalized by the latter. This verbal description translates to

$$D_n \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} = \begin{pmatrix} \|x\|_2 & 0 \\ 0 & \max\{\|x\|_2, \sqrt{n}\} \end{pmatrix} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix}. \tag{2.1}$$

In Section 2.9 we show that for the basic model this choice cannot be improved even in the class of non-diagonal normalizers.

In the vector case (1.3) we use as a normalizer for X_n the matrix

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

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. Generalizing upon (2.1) we choose $\Delta_n I_s$ as a normalizer for Y_n^-, where

$$\Delta_n = \max\{d_{n1}, ..., d_{nr}, \sqrt{n}\} \tag{2.2}$$

and I_s is the identity matrix of size $s \times s$. Implicit in this choice is the presumption that all lags of y should converge at the same rate. Denoting

$$H_n = d_n^{-1} X_n, \ D_n = \begin{pmatrix} d_n & 0 \\ 0 & \Delta_n I_s \end{pmatrix}_{(s+r)\times(s+r)} \tag{2.3}$$

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

$$D_n^{-1} Z_n = \begin{pmatrix} H_n \\ \frac{1}{\Delta_n} Y_n^- \end{pmatrix}. \tag{2.4}$$

2.2 Representation of the N- and D-factors

It is easy to obtain from (1.1) by induction

$$y_t = \sum_{s=1}^{t} B^{t-s}(Ax_s + e_s) + B^t y_0, \ t = 1, ..., n. \tag{2.5}$$

Let F be a matrix of size $s \times n$ with columns denoted $F_t, \ t = 1, ..., n.$ With the matrix B we associate an operator P_B acting on F according to

$$(P_B F)_t = \sum_{s=1}^{t-1} B^{t-1-s} F_s, \ t = 1, ..., n.$$

Here and in the sequel we put $\sum_{j=m}^{n} ... = 0$ if $n < m$. This definition and (2.5) imply

$$Y_n^- = M_n + \rho_n$$

where

$$M_n = P_B(AX_n + \mathcal{E}_n) = P_B(Ad_n H_n + \mathcal{E}_n), \ \rho_n = (y_0, By_0, ..., B^{n-1} y_0) \tag{2.6}$$

are the *main part* and *residual*, respectively. From (2.4) we see that the N-factor equals

$$\mathcal{E}_n Z_n' D_n^{-1} = \left(\mathcal{E}_n H_n', \frac{1}{\Delta_n} \mathcal{E}_n M_n'\right) + \left(0, \frac{1}{\Delta_n} \mathcal{E}_n \rho_n'\right) \tag{2.7}$$

and the D-factor is

$$\begin{aligned} D_n^{-1} Z_n Z_n' D_n^{-1} \ &= \ \begin{pmatrix} H_n H_n' & \frac{1}{\Delta_n} H_n M_n' \\ \frac{1}{\Delta_n} M_n H_n' & \frac{1}{\Delta_n^2} M_n M_n' \end{pmatrix} \\ &+ \begin{pmatrix} 0 & \frac{1}{\Delta_n} H_n \rho_n' \\ \frac{1}{\Delta_n} \rho_n H_n' & \frac{1}{\Delta_n^2}(M_n \rho_n' + \rho_n M_n' + \rho_n \rho_n') \end{pmatrix}. \end{aligned} \tag{2.8}$$

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

2.3 L_p-approximability

The main idea behind L_p-approximability is to approximate sequences of deterministic vectors $\{w_n : n = 1, 2, ...\}$ with functions of a continuous argument. For example, for a given sequence $\{w_n\}$, where w_n has n coordinates for each n, one can postulate existence of a continuous function W on $[0, 1]$ such that $\max_{t=1,...,n} |w_{nt} - W(t/n)| \to 0$, see a similar assumption in [28, Equation (5.141)]. To allow approximation by discontinuous and unbounded functions, it is better to use L_p spaces.

Let $p \in [1,\infty]$, $\|W\|_p = \left(\int_0^1 |W(x)|^p \, dx\right)^{1/p}$ if $p < \infty$ and $\|W\|_\infty = \operatorname{ess\,sup}_{x\in(0,1)} |W(x)|$. Denote by L_p the space of measurable functions on $(0,1)$ with a finite norm $\|W\|_p < \infty$. Let us partition the interval $[0,1)$ into subintervals $i_t = [(t-1)/n, t/n)$, $t = 1, ..., n$. Since, in general, elements of L_p are not continuous, instead of values $W(t/n)$ we need to use averages over subintervals $n\int_{i_t} W(x)dx$. Those averages are scaled by $n^{-1/p}$ for convenience. In this way we obtain a *discretization operator* $\delta_{np} : L_p \to \mathbb{R}^n$ defined by

$$(\delta_{np}W)_t = n^{1-1/p} \int_{i_t} W(x)dx, \ t = 1, ..., n.$$

Let l_p be a discrete analog of L_p with the norm $\|w\|_p = \left(\sum_{t\in T} |w_t|^p\right)^{1/p}$, $p < \infty$, and $\|w\|_\infty = \max_{t\in T} |w_t|$, $p = \infty$. The set of indices T depends on the context. In particular, we use $\mathbb{R}^n_p$ (the set of n-dimensional vectors) and the set $\mathbb{M}_p$ of matrices of all sizes.

Definition 1. We say that a sequence $\{z_n\}$, where $z_n \in \mathbb{R}^n$ for each n, is *L_p-approximable* if there exists a function $z^c \in L_p$ such that $\|z_n - \delta_{np} z^c\|_p \to 0, \ n \to \infty$. In this case we also say that $\{z_n\}$ is *L_p-close to* z^c. The superscript c is used to emphasize that z^c is considered a continuous proxy for z_n.

This notion is designed for modeling deterministic regressors in linear models and should be distinguished from L_p-approximability introduced in [26] for approximating stochastic processes by other, less complex, ones in nonlinear models. The toolkit that accompanies our L_p-approximability facilitates calculation of various limits which would be hard to evaluate otherwise.

Example 1. For a polynomial trend $x_t = t^k$, where $k \geq 0$ is an integer, the norm $||x||_2$ is of order $n^{k+1/2}$ [14, Equation (16.1.13)]. Unless x_t is a constant, $||x||_2$ grows faster than $\sqrt{n}$. The normalized trend $h = x/||x||_2$ is L_2-close to $h^c(x) = (2k+1)^{1/2}x^k$ [23].

To handle vector autoregressions, 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

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

We always assume that such a function has values of the same size. By definition, the discretization operator is applied to matrices element-wise. 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 $||F_n - \delta_{np}F^c; l_p(\tau_n, \mathbb{M}_p)|| \to 0$, $n \to \infty$. If this is true we also say that $\{F_n\}$ is *L_p-close* to F^c. Obviously, uniform boundedness of norms

$$\sup_n ||F_n; l_p(\tau_n, \mathbb{M}_p)|| < \infty \tag{2.9}$$

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, $t = 1, ..., n$.

2.4 Assumptions

Here we list and discuss the assumptions used in our positive statement.

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

This condition implies existence of $c > 0$ and $\lambda \in (0,1)$ such that $||B^k|| \le c\lambda^k,\ k = 0, 1, ...,$ see, for example, [1, Lemma 5.5.1]. For this reason all series containing powers of B will converge.

Assumption 2. (On normalized regressors) The sequence $\{H_n\}$ (see (2.3)) is L_2-close to some vector $H^c \in L_2$.

Assumption 2 is satisfied for polynomial trends and is not satisfied for exponential trends [23]. Normalized sequences arising from slowly varying functions also satisfy Assumption 2 [20].

By $\xrightarrow{d}$ and dlim ($\xrightarrow{p}$ and plim) we denote convergence and limit in distribution (in probability, respectively). $I(A)$ denotes the indicator of a set A.

Assumption 3. (On errors) For the error matrices $\mathcal{E}_n$ we assume a slightly more general structure than (1.4): $\mathcal{E}_n = (e_{n1}, ..., e_{nn})$ where the columns e_{nt} may depend on n and satisfy the following conditions:

(i) For each n, the columns e_{nt} are martingale differences (m.d.s) with respect to nested σ-fields $\mathcal{F}_{n0} \subset \mathcal{F}_{n1} \subset ... \subset \mathcal{F}_{nn}$, that is, e_{nt} is $\mathcal{F}_{nt}$-measurable and $\mathrm{E}(e_{nt}|\mathcal{F}_{n,t-1}) = 0$.

(ii) $\sup_{n,t} E||e_{nt}||_2^p < \infty$ for some $p > 2$ and conditional expectations $\Sigma_{nt} = \mathrm{E}(e_{nt}e'_{nt}|\mathcal{F}_{n,t-1})$ 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) $\operatorname{plim}_{K\to\infty} \sup_{n,t} \mathrm{E}(||e_{nt}||_2^2 I(||e_{nt}||_2 > K)|\mathcal{F}_{n,t-1}) = 0$ and the σ-fields are nested over n: $\mathcal{F}_{nt} \subset \mathcal{F}_{n+1,t}$ 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

$$\mathrm{E}(e_{ns}e'_{nt}|\mathcal{F}_{n,\max\{s,t\}-1}) = \begin{cases} 0, & s \neq t, \\ \Sigma_{nt}, & s = t. \end{cases} \tag{2.10}$$

Normally this equation will be used in conjunction with the law of iterated expectations, without explicitly mentioning it. One of conditions in [3] is

$$\frac{1}{n}\sum_{t=1}^{n} \Sigma_{nt} \xrightarrow{p} \Sigma.$$

In this equation, the information about heterogeneity contained in Σ_{nt} is forgotten in the limit matrix Σ. Assumption 3(iii) and Theorem 3.1 allow us to prove

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

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

Example 2. Let $e_1,\ e_2,\ \ldots$ be i.i.d. variables satisfying $\mathrm{E}e_t = 0$ and $\sigma^2 = \mathrm{E}e_t^2 < \infty$. Take any sequence $\{f_n\}$ of vectors $f_n \in \mathbb{R}^n$ such that $\{f_n\}$ is L_∞-close to some $f \in C[0,1]$ (one can take $f_n = (f(1/n), f(2/n), \ldots, f(1))'$, for example, see [17, Theorem 3.3(b)]), and put $e_{nt} = f_{nt}e_t$. Let $\mathcal{F}_t = \sigma(e_j : j \le t)$ be the least σ-field such that $e_1, \ldots, e_t$ are $\mathcal{F}_t$-measurable. Then e_{nt} is $\mathcal{F}_t$-measurable, $\mathrm{E}(e_{nt}|\mathcal{F}_{t-1}) = 0$ by independence, $\Sigma_{nt} = \mathrm{E}(e_{nt}^2|\mathcal{F}_{t-1}) = f_{nt}^2 \mathrm{E}e_t^2 = \sigma^2 f_{nt}^2 \le c$. It is easy to see that $\{\Sigma_n\}$ is L_∞-close to $\sigma^2 f^2$. Thus, in the limit Σ_n vanishes where f^2 vanishes.

Assumption 4. (Stabilization of relative growth rates of regressors) The limits (see (2.2))

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

exist.

By looking at κ_Q we shall be able to distinguish the cases $|Q| = 0$ from $|Q| \neq 0$. Denoting

$$b_n = \frac{1}{\Delta_n} d_n, b = \mathrm{diag}[\kappa_1, \ldots, \kappa_r],$$

under Assumption 4 one has $b = \lim b_n$. Since b accounts for the balance between the growth rates of regressors (including lags), it is natural to name it a *balancer*. The matrix $J = (I-B)^{-1}Ab$ will be called a *jack*, because it is a "subordinate part of a machine, rendering convenient service".

Assumption 5. (On the initial value) $\mathrm{E}||y_0||_2^2 < \infty$.

2.5 Convergence of the D-factor

Denote

$$G = \int_0^1 H^c(H^c)'dx,\ \Xi(x) = \sum_{s=0}^{\infty} B^s \Sigma^c(x) B'^s$$

G is the Gram matrix of the system $(H^c)_1, \ldots, (H^c)_r \in L_2$. The condition $|G| \neq 0$ means linear independence of this system (or asymptotic linear independence of the rows of H) and implies positivity of $|G|$ [12, Chapter IX, § 5]. In all statements below $|G| \neq 0$ will be assumed or implied by other conditions.

It is easy to see that the matrix $\Xi(x)$ solves the equation $\Xi(x) - B\Xi(x)B' = \Sigma^c(x)$, see [1, Chapter 5, Exercise 27]. Since the solutions of similar equations $\Xi - B\Xi = \Sigma^c$ and $\Xi - \Xi B' = \Sigma^c$ are given by a *left resolvent* $(I-B)^{-1}\Sigma^c$ and *right resolvent* $\Sigma^c(I-B')^{-1}$, it is natural to call $\Xi(x)$ an *enveloping resolvent*. Let

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

L_p-lim denotes the limit in the norm $(\mathrm{E}\,\|X\|_2^p)^{1/p}$.

Theorem 2.1. *If Assumptions 1-5 hold, then*

(i) the D-factor converges in $L_1(\Omega)$: L_1-$\lim D_n^{-1} Z_n Z_n' D_n^{-1} = Q$.

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

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

Note that when $|G| \neq 0$, by the determinant of a partitioned matrix rule

$$|Q| = |G| \left| JGJ' + \kappa_Q^2 \int_0^1 \Xi(x)dx - JGG^{-1}GJ' \right| = \kappa_Q^{2s}|G| \left| \int_0^1 \Xi(x)dx \right|. \tag{2.11}$$

This equation shows that $|Q|$ does not depend on A.

2.6 Convergence of the N-factor

Put

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

so that $Q = \int_0^1 \Omega_1(x)dx$. The asymptotic behavior of the N-factor involves more complex integrals of Ω_1.

Theorem 2.2. *Let Assumptions 1-5 hold. Then*

(i) the N-factor converges in distribution

$$\operatorname{vec}(\mathcal{E}_n Z_n' D_n^{-1}) \xrightarrow{d} N\left(0, \int_0^1 \Omega_1(x) \otimes \Sigma^c(x)dx\right). \tag{2.12}$$

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

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

An integral like $\int_0^1 \Omega_1(x) \otimes \Sigma^c(x)dx$ can be termed an *error-weighed integral* of Ω_1. In the last proposition we see three such integrals. Any of them may degenerate, depending on the behavior of Σ^c (see the example in the next section).

Corollary 2.1. *If Assumptions 1-5 hold and $\kappa_Q = 0$, then, in addition to convergence* (2.12), *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 $\operatorname{vec} U \sim N\left(0, \int_0^1 [H^c(H^c)'] \otimes \Sigma^c dx\right)$ and V is proportional to U, $V = UJ'$.*

2.7 Positive statement

Equations $\operatorname{vec}(AB) = (B' \otimes I)\operatorname{vec}A$ and (1.7) lead to

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

which, in combination with Theorems 2.1 and 2.2, immediately gives the following result:

Theorem 2.3. *(Convergence of the OLS estimator, case $\kappa_Q > 0$) Let Assumptions 1-5 hold. If $|Q| \neq 0$, then*

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

Example 3. Here we show that the condition $|Q| \neq 0$ does not guarantee nondegeneracy of the limit distribution. Let there be only two regressors. Take functions H_1^c and H_2^c and a matrix Σ^c with nonoverlapping 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 2.2 that the N-factor converges to a degenerate normal vector. If we take the regressors to be of form $\{\delta_{n2} H_1^c\}$, $\{\delta_{n2} H_2^c\}$, then the norms of the rows of X_n will have finite limits $\|H_1^c\|_2$, $\|H_2^c\|_2$ and Δ_n will equal $\sqrt{n}$ for all large n, leading to $\kappa_Q = 1$. Then using Theorem 2.1 it is straightforward to show that $|Q| \neq 0$ and Theorem 2.3 is applicable.

2.8 Negative statement

The system of vector equations (1.1) with just one lag of the dependent variable encompasses a variety of cases we do not cover in the negative result. For example, [16, Section 10.5.1] has a system in which y_t includes lagged exogenous regressors. The additional assumption we need will be modeled on the system of scalar equations

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

which can be written in form (1.1) with matrices

$$A = \begin{pmatrix} \alpha_1 & ... & \alpha_r \\ 0 & ... & 0 \\ ... & ... & ... \\ 0 & ... & 0 \end{pmatrix}_{s\times r}, \ B = \begin{pmatrix} \beta_1 & ... & \beta_{s-1} & \beta_s \\ 1 & ... & 0 & 0 \\ ... & ... & ... & ... \\ 0 & ... & 1 & 0 \end{pmatrix}_{s\times s} \tag{2.13}$$

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.

We say that model (1.1) is a *proper mixed autoregression* if all diagonal elements of JGJ' are positive for any nonzero balancer b. The name is explained by the fact that this condition is fulfilled in case (2.13) with all the alphas different from zero (Lemma 3.7).

Theorem 2.4. *(Inapplicability of the conventional scheme, case $\kappa_Q = 0$) Suppose, Assumptions 1-5 are satisfied, $|G| \neq 0$, $\kappa_Q = 0$ and the model is a proper mixed autoregression. Then by Theorem 2.1 the D-factor converges in mean to Q with $|Q| = 0$, so that the conventional scheme does not work. Moreover, there is no diagonal (possibly stochastic) normalizer $\widetilde{D}_n$ for which the D-factor would converge in probability to a nonsingular nonstochastic matrix:*

$$\text{plim} \widetilde{D}_n^{-1} Z_n Z_n' \widetilde{D}_n^{-1} = \widetilde{Q}, \ |\widetilde{Q}| \neq 0. \tag{2.14}$$

To emphasize the strength of this result, we state it in a different form.

Corollary 2.2. *Under the conditions of the last theorem, our normalizer is unique up to an asymptotically constant diagonal factor. If the conventional scheme works with some diagonal $\widetilde{D}_n$, then it will work with our D_n too.*

Besides, in the next section we show that, under reasonable conditions, for the basic model the negative statement extends to non-diagonal $\widetilde{D}_n$.

2.9 Examples and discussion

We give examples for both the positive and negative statements. The way heterogeneous errors are implemented here is different from that in [3]. Besides, Anderson and Kunitomo expend a lot of effort to relax the integrability requirement on the errors. To facilitate comparison and concentrate on essential differences, we assume homogeneous errors and replace their conditional uniform integrability condition on e_t by our p-summability condition. Thus, in Examples 4, 5 and 6 we suppose that $|\beta| < 1$ and $\{e_t\}$ is an i.i.d. sequence with $\mathrm{E}e_t = 0$, $\sigma^2 = \mathrm{E}e_t^2$, $\sup \mathrm{E}|e_t|^p < \infty$ for some $p > 2$.

Example 4. Regarding the elements of the basic model suppose that

(a) The normalized regressor $h = x/\|x\|_2$, where $x = (x_1, ..., x_n)'$, is L_2-close to a function $h^c \in L_2$.

(b) The limits $\kappa_1 = \lim_{n\to\infty} \|x\|_2/\Delta_n$, $\kappa_Q = \lim_{n\to\infty} \sqrt{n}/\Delta_n$, where $\Delta_n = \max\{\|x\|_2, \sqrt{n}\}$, exist.

Put

$$Q = \begin{pmatrix} 1 & \rho \\ \rho & \rho^2 + \frac{\sigma^2\kappa_Q^2}{(1-\beta)^2} \end{pmatrix}, \text{ where } \rho = \frac{\alpha\kappa_1}{1-\beta}.$$

If κ_Q is positive, then $|Q| \neq 0$ and

$$\begin{pmatrix} \|x\|_2(\hat\alpha - \alpha) \\ \Delta_n(\hat\beta - \beta) \end{pmatrix} \xrightarrow{d} N(0, \sigma^2 Q^{-1}).$$

This statement follows from Theorem 2.3 and has actually been obtained in [18] under a stronger errors integrability condition.

[3] for the same basic model assume:

(i) Letting $\mathcal{F}_t = \sigma(e_j : j \le t)$ suppose that x_t is $\mathcal{F}_{t-1}$-measurable.

(ii) $\frac{1}{n}\max_{t=1,...,n} x_t^2 \xrightarrow{p} 0$, as $n \to \infty$.

(iii) $\frac{1}{n}\sum_{t=1}^n x_t^2 \xrightarrow{p} M$, as $n \to \infty$, where M is a constant.

(iv) $\frac{1}{n}\sum\limits_{t=1}^{n-h} x_{t+h}x_t \xrightarrow{p} M_h = M_{-h}$, as $n \to \infty$, for all $h = 0, 1, 2, ...$, where M_h are constants and $M_0 > 0$.

(v) $\frac{1}{n}\sum\limits_{t=1}^{n-h} x_{t+h}e_t \xrightarrow{p} 0$, as $n \to \infty$, for all $h = 1, 2, ...$

(vi) $\frac{1}{n}\sum\limits_{t=\max\{r,s\}+2}^{n} e_{t-1-r}e_{t-1-s} \xrightarrow{p} \delta_{rs}\sigma^2 = \begin{cases} \sigma^2, & r = s \\ 0, & r \neq s \end{cases}$, as $n \to \infty$, for all $r, s = 0, 1, 2, ...$

Define $L = \sum_{s=0}^\infty \beta^s \alpha M_{-(s+1)}$. Then the limit $Q = \operatorname{plim}\frac{1}{n}\sum_{t=1}^n y_{t-1}^2$ exists, $\det\begin{pmatrix} Q & L \\ L & M_0 \end{pmatrix} \neq 0$ and

$$\sqrt{n}\begin{pmatrix} \hat\alpha - \alpha \\ \hat\beta - \beta \end{pmatrix} \xrightarrow{d} N\left(0, \sigma^2\begin{pmatrix} Q & L \\ L & M_0 \end{pmatrix}^{-1}\right).$$

Remark 1. Since Anderson and Kunitomo focus on stochastic regressors, no wonder they need to impose many more conditions. Note that each of (iv), (v) and (vi) consists of an infinite series of assumptions.

Remark 2. Unfortunately, in case of a deterministic x_t, the only obvious simplification is that (i) becomes trivial and in (ii), (iii) and (iv) $\xrightarrow{p}$ can be replaced by $\to$. Condition (vi) has nothing to do with the regressors and does not simplify. In case of i.i.d. errors it follows from the weak law of large numbers but otherwise can be burdensome. Conditions (ii) and (iii) exclude quickly growing regressors (in case of polynomials, starting from the linear trend). There is clearly a need for a dedicated result for deterministic regressors.

Remark 3. In their general Theorem 3, Anderson and Kunitomo put exogenous and endogenous regressors into one pile, and then in their Theorems 4, 5 and 6 specify the general result to autoregression and mixed autoregression by sorting out the regressors. A drastic reduction in the number of assumptions in our result is due to L_p-approximability and the fact that, from the very beginning, we keep separately the exogenous regressors from the lags of the dependent variable. The method here also improves upon [18]: the weak law of large numbers for mixingales [5] and theorem on transforms of martingales [8] are not used, while the errors integrability requirement is lower and heterogeneous errors are allowed.

Example 5. For the model $y_t = \alpha_1 + \alpha_2 t + \beta_1 y_{t-1} + e_t$ with $\alpha_1\alpha_2 \neq 0$, the pair of regressors $\{1, t\}$ by Example 1 is L_2-close (upon normalization) to a pair of functions $\{1, \sqrt{3}t\}$, which is linearly independent in L_2. d_{n1} is of order $\sqrt{n}$, d_{n2} and Δ_n are of order $n^{3/2}$, $\kappa_1 = 0$, $\kappa_2 = 1$, $\kappa_Q = 0$ and by the negative result the conventional scheme does not work with our or any other diagonal normalizer.

Next we discuss the issue of non-diagonal normalization. [1] worked with diagonal normalizers. [3] in their Theorem 3 introduced a non-diagonal normalizer D_n. They have fixed the properties of D_n axiomatically, without indicating how it can be constructed from the data. Therefore their Theorem 3 cannot be applied in practice. No attempt to justify the introduction of non-diagonal normalizers has been made. In their specification of that theorem to a mixed autoregression, they have reverted back to the traditional $\sqrt{n}$, thus forfeiting the opportunity to capture different growth rates of the regressors (which was not their goal). One may conjecture that when $\kappa_Q = 0$ and, hence, $|Q| = 0$, using non-diagonal normalizers may save the situation. In the following example we present evidence that this is not the case for the basic model.

Example 6. Consider the basic model under the assumptions (a), (b) of Example 4. Suppose $\kappa_Q = 0$ and $\widetilde{D}_n$ is a non-diagonal matrix. By [12, Chapter IX, §14, Theorem 9] there exist a symmetric matrix S_n, $S_n = S_n'$, and an orthogonal matrix O_n, $O_n'O_n = I$, such that $\widetilde{D}_n = S_n O_n$. This equation implies $\widetilde{D}_n' = O_n' S_n$ and

$$\widetilde{Q}_n \equiv \widetilde{D}_n^{-1} Z_n Z_n' \widetilde{D}_n'^{-1} = O_n^{-1} S_n^{-1} Z_n Z_n' S_n^{-1} O_n.$$

Since $O_n^{\pm 1}$ have determinants 1 or -1, premultiplication by O_n^{-1} and postmultiplication by O_n cannot change the asymptotic behavior of $|S_n^{-1} Z_n Z_n' S_n^{-1}|$, so we can safely assume that $\widetilde{D}_n$ is symmetric. (This part of the argument holds for the general model (1.3).)

By [12, Chapter IX, §13, Equation (119)] there exist a diagonal matrix $\overline{D}_n = \mathrm{diag}[\bar{d}_{n1}, \bar{d}_{n2}]$ and an orthogonal matrix O_n such that $\widetilde{D}_n = O_n \overline{D}_n O_n^{-1}$. From this equation and the definition of $\widetilde{Q}_n$ and Q_n we get

$$\begin{aligned} \widetilde{Q}_n &= O_n \overline{D}_n^{-1} O_n^{-1} Z_n Z_n' O_n \overline{D}_n^{-1} O_n^{-1} \\ &= O_n \overline{D}_n^{-1} O_n^{-1} D_n D_n^{-1} Z_n Z_n' D_n^{-1} D_n O_n \overline{D}_n^{-1} O_n^{-1} \\ &= O_n \overline{D}_n^{-1} O_n^{-1} D_n Q_n D_n O_n \overline{D}_n^{-1} O_n^{-1}. \end{aligned} \tag{2.15}$$

Note that

$$\kappa_Q = \lim_{n\to\infty} \frac{\sqrt{n}}{\max\{d_{n1}, \sqrt{n}\}} = 0$$

means that $\sqrt{n} = o(d_{n1})$ and $\Delta_n = d_{n1}$. Hence,

$$D_n = \begin{pmatrix} d_{n1} & 0 \\ 0 & \Delta_n \end{pmatrix} = d_{n1} I \text{ for all large } n$$

and D_n commutes with any second-order matrix. Therefore (2.15) can be rearranged as

$$O_n^{-1}\widetilde{Q}_nO_n = \overline{D}_n^{-1}D_n(O_n^{-1}Q_nO_n)D_n\overline{D}_n^{-1}.$$

It is reasonable to assume that

(c) the limit $O = \operatorname{plim} O_n$ exists.

Then (2.14) implies $O_n^{-1}\widetilde{Q}_nO_n \xrightarrow{p} O^{-1}\widetilde{Q}O$ and an argument similar to Step 2 of the proof of Theorem 2.4 shows that all diagonal elements of $O^{-1}\widetilde{Q}O$ are positive.

Now we want to see when both diagonal elements of $C = O^{-1}QO$ are positive. Since $\kappa_1 = 1$, the jack equals $J = \alpha\kappa_1/(1-\beta) = \alpha/(1-\beta)$ and by Theorem 2.1

$$Q = \begin{pmatrix} 1 & \rho \\ \rho & \rho^2 \end{pmatrix}, \text{ where } \rho = \frac{\alpha}{1-\beta}.$$

A second-order orthogonal matrix O can be represented as

$$O = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}, \ -\pi < \phi \le \pi.$$

Simple algebra gives the following expressions for the diagonal elements of C:

$$c_{11} = (\cos\phi - \rho\sin\phi)^2, \ c_{22} = (\sin\phi + \rho\cos\phi)^2.$$

c_{11} is zero where $\cot\phi = \rho$ and c_{22} is zero where $\tan\phi = -\rho$. Thus, for any ϕ such that

$$\cot\phi \ne \rho, \ \tan\phi \ne -\rho \tag{2.16}$$

both c_{11}, c_{22} are positive.

The positivity of the diagonal elements of $O^{-1}\widetilde{Q}O$ and $O^{-1}QO$ leads to existence of positive limits $\operatorname{plim}(\bar{d}_{ni}/d_{ni})^2$ and to a contradiction, as in Steps 3 and 4 of the proof of Theorem 2.4. This contradiction arises because of the assumption (2.14) with a non-diagonal $\widetilde{Q}$. Hence, we have to restrict our attention to diagonal normalizers in which case our normalizer is unique up to an asymptotically constant factor. Now we make one more assumption:

(d) The normalizer is a continuous function of the model parameters.

Under this condition uniqueness of our diagonal normalizer (which does not depend on the model parameters) extends to the values of ϕ excluded in (2.16).

The conclusion is that the benefits of non-diagonal normalization, if they exist, are rather an exception than a rule. They may exist when one or more of the conditions imposed in this example are violated.

Example 7. [14, Section 16.3] derives the asymptotic distribution of the OLS estimator for $y_t = \alpha + \delta t + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + e_t$ following the suggestion by [27] to detrend the series. Specifically, the argument is based on the idea that $y_t - \alpha - \delta t$ should be covariance-stationary (see [14], p.464, second paragraph from the bottom). This idea neglects the fact that not only the term $\alpha + \delta t$ but also the lags of y contribute to the long-run trend of y. Here, by deriving the long-run trend, we show that $y_t - \alpha - \delta t$ is not covariance-stationary.

For simplicity, take $p = 1$, $y_0 = 0$ and denote

$$\phi = \phi_1, \ \alpha^* = \frac{\alpha}{1-\phi}, \ \beta^* = \frac{\beta}{1-\phi}, \text{ where } |\phi| < 1.$$

By induction, it is easy to derive the representation

$$y_t = \sum_{k=0}^{t-1}[\alpha + (t-k)\beta + e_{t-k}]\phi^k, \ k \ge 1.$$

With the notation
$$\alpha_t = \alpha \sum_{k=0}^{t-1} \phi^k, \ \beta_t = \beta \sum_{k=0}^{t-1} \left(1 - \frac{k}{t}\right) \phi^k$$
it takes the form
$$y_t = \alpha_t + \beta_t t + \sum_{k=0}^{t-1} \phi^k e_{t-k}.$$
Obviously, $\alpha_t \to \alpha^*$ as $t \to \infty$. For a given $\varepsilon > 0$ choose t_0 such that $|\phi|^{t_0} < \varepsilon$. Let t_1 be so large that $t_0/t_1 \leq \varepsilon$. Since $0 \leq 1 - k/t \leq 1$ for $0 \leq k < t$, we have for all $t \geq t_1$
$$\begin{aligned}
|\beta_t - \beta^*| &= |\beta| \left| \left(\sum_{k=0}^{t_0} + \sum_{k=t_0+1}^{t-1} \right) \left(1 - \frac{k}{t}\right) \phi^k - \left(\sum_{k=0}^{t_0} + \sum_{k>t_0} \right) \phi^k \right| \\
&= |\beta| \left| - \sum_{k=0}^{t_0} \frac{k}{t} \phi^k + \sum_{k=t_0+1}^{t-1} \left(1 - \frac{k}{t}\right) \phi^k - \sum_{k>t_0} \phi^k \right| \\
&\leq |\beta| \left(\sum_{k=0}^{t_0} \frac{t_0}{t} |\phi^k| + \sum_{k=t_0+1}^{t-1} |\phi^k| + \sum_{k\geq t_0} |\phi^k| \right) \\
&\leq |\beta| \left(\frac{\varepsilon}{1 - |\phi|} + \frac{2|\phi|^{t_0}}{1 - |\phi|} \right) \leq \frac{3\varepsilon|\beta|}{1 - |\phi|}.
\end{aligned}$$
This proves that $\beta_t \to \beta^*$.

Of course, one can detrend y_t by $\alpha^* + \beta^* t$ but then the detrended series will not be truly stationary. The realization of the idea requires a finer technique than just stationarity.

3 Appendix

3.1 Operators arising in the theory of autoregressive models

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:
$$\begin{aligned}
(P_AF)_t &= \sum_{s=1}^{t-1} A^{t-1-s} F_s, \ (Q_AF)_t = \sum_{s=t+1}^{n} F_s A^{s-1-t}, \\
(R_{A,B}F)_t &= \sum_{s=1}^{t-1} A^{t-1-s} F_s B^{t-1-s}, \ t = 1, ..., n,
\end{aligned}$$
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 decreasing powers of A one can think of increasing powers 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. The next theorem is taken from [19, Chapter 8].

Theorem 3.1. *(i) If $\{X_n\}$ is L_p-approximable and $p < \infty$, then* $\lim_{n\to\infty} \max_{1\leq t\leq n} ||X_{nt}||_p = 0$.

(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 operator norms 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* $\{P_B X_n\}$ *is* L_p*-close to the left resolvent* $(I - B)^{-1} X^c$, $\{Q_B X_n\}$ *is* L_p*-close to the right resolvent* $X^c(I - B)^{-1}$ *and* $\{R_{B,B'} X_n\}$ *is* L_p*-close to the enveloping resolvent* $\sum_{s=0}^{\infty} B^s X^c B'^s$.

(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* $\{X_n Y_n\}$ *is* L_p*-close to* $X^c Y^c$. *In particular, if* $\{A_n\}$ *is a sequence of matrices converging to* A, *then* $\{A_n X_n\}$ *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.

3.2 An integral version of the Cramér-Wold device

To reduce the problem of convergence in distribution of a sequence of stochastic matrices $\{W_n\}$ to the one-dimensional case we prove an extension of a well-known device [1, Theorem 7.7.7].

Lemma 3.1. *Convergence in distribution*

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

where Ω_1, Ω_2 *are symmetric matrices with square-integrable components, takes place if and only if for any constant matrix* C

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

Proof. Using

$$\operatorname{tr}(ABC) = (\operatorname{vec} B')'(I \otimes C)\operatorname{vec} A \tag{3.1}$$

we get

$$\operatorname{tr}(W_n C) = c' \operatorname{vec} W_n \tag{3.2}$$

where $c = \operatorname{vec}(C')$. From (3.1) and

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

we see that

$$\begin{aligned} \int_0^1 \operatorname{tr}[(C'\Omega_1) C \Omega_2] dx &= \int_0^1 c'(I \otimes \Omega_2)\operatorname{vec}(C'\Omega_1) dx \qquad (3.3) \\ = c' \int_0^1 (I \otimes \Omega_2)(\Omega_1' \otimes I) dx c &= c' \int_0^1 \Omega_1(x) \otimes \Omega_2(x) dx c. \end{aligned}$$

(3.2), (3.3) and the Cramér-Wold theorem prove the lemma. ∎

3.3 Operations with uniformly integrable functions

In the next lemma ν and μ are arbitrary sets of indices and, as before, matrices in a sequence are of the same size. "u.i." will mean "uniformly integrable".

Lemma 3.2. *(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., variables $||Y_m||_2^q$, $m \in \mu$, 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 from [10, Theorem 12.9]: $\{X_n\}$ is u.i. if and only if $\sup_n \mathrm{E}|X_n| < \infty$ and for any $\varepsilon > 0$ there is a $\delta > 0$ such that for all events A of probability $P(A) < \delta$ one has $\sup_n \mathrm{E}|X_n| I(A) < \varepsilon$.

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

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

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

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

(iii) follows from

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

Let us prove (iv). If all elements of $X_n'X_n$ are u.i., then such 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 $\mathrm{E}||X_n||_2^2 I(A) < \varepsilon$ for all A satisfying $P(A) < \delta$. Then for the (i,j)th element of $X_n'X_n$ we have

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

which is what we want. □

3.4 A martingale weak law and central limit theorem

For proving convergence in mean the following Chow-Davidson theorem is useful, see [10, Theorem 19.7].

Theorem 3.2. *If $\{X_{nt}, \mathcal{F}_{nt}\}$ is an m.d. array, positive constants c_{nt} satisfy $\sup_n \sum_{t=1}^n c_{nt} < \infty$ and $\lim_{n\to\infty} \sum_{t=1}^n c_{nt}^2 = 0$ and variables X_{nt}/c_{nt} are uniformly integrable, then $\mathrm{E}|\sum_{t=1}^n X_{nt}| \to 0$.*

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

Following [3], among different versions of martingale central limit theorems we choose the format suggested by [11], 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 main results in Section 2). Anderson and Kunitomo do not do such analysis.

Theorem 3.3. *(Dvoretzky CLT) If $\{X_{nt}, \mathcal{F}_{nt}\}$ is an m.d. array, σ_{nt}^2 denotes* $\mathrm{E}(X_{nt}^2 | \mathcal{F}_{n,t-1})$,

$$\operatorname{plim} \sum_{t=1}^{n} \sigma_{nt}^2 = \sigma^2, \tag{3.4}$$

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

$$\operatorname{plim} \sum_{t=1}^{n} \mathrm{E}(X_{nt}^2 I(|X_{nt}| > \varepsilon) | \mathcal{F}_{n,t-1}) = 0, \tag{3.5}$$

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 [13] for details.

3.5 Convergence in mean of two auxiliary vector sequences

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

$$V_n = \frac{1}{n} \sum_{t=1}^{n} X_{nt} (P_B \mathcal{E}_n)_t (P_B \mathcal{E}_n)_t' = \frac{1}{n} \sum_{t=1}^{n} X_{nt} \sum_{k,l=1}^{t-1} B^{t-1-k} e_{nk} e_{nl}' B'^{t-1-l} \tag{3.6}$$

(here $\{X_n\}$ is some deterministic sequence and not the X_n from (1.4)). They will be used to control second-order conditional moments in the proof of our vector central limit theorem (Theorem 2.2).

Lemma 3.3. *Let Assumptions 1 and 3 hold.*

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

$$L_2\text{-}\lim U_n = 0. \tag{3.7}$$

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

$$L_1\text{-}\lim V_n = \lim \mathrm{E} V_n = \int_0^1 X^c(x) \Xi(x) dx. \tag{3.8}$$

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

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

By Theorem 3.1(iii) and Assumption 3(iii) $||R_{B,B'}\Sigma_n||_\infty \le c$, so

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

which proves (3.7).

(b) By orthogonality (2.10)

$$\mathrm{E}V_n = \frac{1}{n}\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1} B^{t-1-k}\Sigma_{nk}B'^{t-1-k} = \sum_{t=1}^n \frac{1}{\sqrt{n}}X_{nt}\left(R_{B,B'}\frac{1}{\sqrt{n}}\Sigma_n\right)_t. \tag{3.9}$$

By Theorem 3.1(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 3.1(iii) $\left\{R_{B,B'}\frac{1}{\sqrt{n}}\Sigma_n\right\}$ is L_2-close to $\Xi(x)$. Thus the second equation in (3.8) follows from (3.9) and Theorem 3.1(ii).

Before proving the other part of (3.8), we need to reveal the m.d. structure of the difference $V_n - \mathrm{E}V_n$. From (3.6) and (3.9) we have

$$\begin{aligned} V_n - \mathrm{E}V_n &= \frac{1}{n}\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1} B^{t-1-k}(e_{nk}e'_{nk} - \Sigma_{nk})B'^{t-1-k} \\ &+\frac{1}{n}\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1}\sum_{l=1}^{k-1}[B^{t-1-k}e_{nk}e'_{nl}B'^{t-1-l} \\ &+B^{t-1-l}e_{nl}e'_{nk}B'^{t-1-k}]. \end{aligned} \tag{3.10}$$

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 (3.10) we get

$$\begin{aligned} &\sum_{t=1}^n X_{nt}\sum_{k=1}^{t-1} B^{t-1-k}(e_{nk}e'_{nk} - \Sigma_{nk})B'^{t-1-k} \\ &= \sum_{k=1}^{n-1}\sum_{t=k+1}^n X_{nt}B^{t-1-k}(e_{nk}e'_{nk} - \Sigma_{nk})B'^{t-1-k}. \end{aligned} \tag{3.11}$$

A part of the second sum in (3.10) can be rearranged as follows:

$$\begin{aligned} & \sum_{t=1}^{n} X_{nt} \sum_{k=1}^{t-1}\sum_{l=1}^{k-1} B^{t-1-k} e_{nk} e'_{nl} B'^{t-1-l} \qquad (3.12) \\ = & \sum_{k=1}^{n-1}\sum_{t=k+1}^{n} X_{nt} B^{t-1-k} e_{nk} \left(\sum_{l=1}^{k-1} e'_{nl} B'^{k-1-l} \right) B'^{t-k} \\ = & \sum_{k=1}^{n-1}\sum_{t=k+1}^{n} X_{nt} B^{t-1-k} e_{nk} (P_B \mathcal{E}_n)'_k B'^{t-k}. \end{aligned}$$

Similarly,

$$\begin{aligned} & \sum_{t=1}^{n} X_{nt} \sum_{k=1}^{t-1} \left(\sum_{l=1}^{k-1} B^{t-1-l} e_{nl} \right) e'_{nk} B'^{t-1-k} \qquad (3.13) \\ = & \sum_{k=1}^{n-1}\sum_{t=k+1}^{n} X_{nt} B^{t-k} \left(\sum_{l=1}^{k-1} B^{k-1-l} e_{nl} \right) e'_{nk} B'^{t-1-k} \\ = & \sum_{k=1}^{n-1}\sum_{t=k+1}^{n} X_{nt} B^{t-k} (P_B \mathcal{E}_n)_k e'_{nk} B'^{t-1-k}. \end{aligned}$$

Equations (3.10)-(3.13) are summarized in

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

where

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

Since $(P_B\mathcal{E}_n)_k$ is $\mathcal{F}_{n,k-1}$-measurable, $\{Y_{nk}\}$ is clearly a vector m.d. array. The numbers $c_{nt} = 1/n$, $t = 1, ..., n$, satisfy conditions of Theorem 3.2.

By Assumption 3(ii) the family $\{||e_{nt}||_2^2\}$ is u.i., so by the equivalent characterization [10, Theorem 12.9]

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

Hence, by Lemma 3.2(ii) the family $\{e_{nk}e'_{nl}\}$ is u.i. Next we apply Assumption 1 and Lemma 3.2(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, invoking also Assumption 3(iii), we see that the family consisting of those products and differences $e_{nk}e'_{nk} - \Sigma_{nk}$ is u.i. Finally, the variables Y_{nk}/c_{nt} are u.i. by Lemma 3.2(iii), because $||X_{nt}||_\infty \le c < \infty$ and

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

Thus Theorem 3.2 yields $\mathrm{E}||V_n - \mathrm{E}V_n||_2 \to 0$ which completes the proof. $\square$

3.6 Negligibility of the terms containing the residual

The next lemma establishes the standard fact that the influence of the initial value in (2.7) and (2.8) is asymptotically negligible.

Lemma 3.4. *If Assumptions 1 through 5 hold, then*

$$\operatorname{dlim} \mathcal{E}_n Z_n' D_n^{-1} = \operatorname{dlim}\left(\mathcal{E}_n H_n', \frac{1}{\Delta_n}\mathcal{E}_n M_n'\right), \tag{3.14}$$

$$L_1\text{-}\lim D_n^{-1} Z_n Z_n' D_n^{-1} = L_1\text{-}\lim \begin{pmatrix} H_n H_n' & \frac{1}{\Delta_n} H_n M_n' \\ \frac{1}{\Delta_n} M_n H_n' & \frac{1}{\Delta_n^2} M_n M_n' \end{pmatrix}, \tag{3.15}$$

assuming that the limits on the right exist.

Proof. (3.14) follows from (2.7), Assumptions 1, 3 and 5 and the bound

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

$\{H_n\}$ satisfies a condition of type (2.9), so

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

Obviously,

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

By Assumptions 2, 4 and Theorem 3.1(v) $\{Ab_nH_n\}$ is L_2-close to AbH^c. Assumption 1 and Theorem 3.1(iii) therefore imply

$$P_B A b_n H_n \text{ is } L_2\text{-close to } (I-B)^{-1}AbH^c = JH^c. \tag{3.16}$$

Combine this fact with (2.6) to get

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

Now (3.15) follows from (2.8) and the last three bounds. $\square$

3.7 "Scalarization" of the problem of convergence of the N-factor

The last lemma and (2.7) explain why we are interested in studying the vector

$$W_n = \left(\mathcal{E}_n H_n', \frac{1}{\Delta_n}\mathcal{E}_n M_n'\right). \tag{3.17}$$

Proving convergence of W_n will be a long journey. Partitioning C conformably, $C' = (C_1', C_2')$, and utilizing (3.17) we get

$$\begin{aligned}
\operatorname{tr}(W_n C) &= \operatorname{tr}\left(\mathcal{E}_n H_n' C_1 + \frac{1}{\Delta_n}\mathcal{E}_n M_n' C_2\right) \\
&= \operatorname{tr}\left[\sum_{t=1}^n e_{nt} H_{nt}' C_1 + \frac{1}{\Delta_n}\sum_{t=1}^n e_{nt}(P_B A d_n H_n + P_B \mathcal{E}_n)_t' C_2\right] \\
&= \sum_{t=1}^n \left[H_{nt}' C_1 + (P_B A b_n H_n)_t' C_2\right] e_{nt} + \frac{1}{\Delta_n}\sum_{t=1}^n (P_B \mathcal{E}_n)_t' C_2 e_{nt}.
\end{aligned}$$

Hence, denoting

$$G_{nt} = C_1' H_{nt} + C_2'(P_B A b_n H_n)_t, \; S_{nt} = G_{nt}' e_{nt}, \tag{3.18}$$

$$T_{nt} = \frac{1}{\Delta_n}(P_B \mathcal{E}_n)_t' C_2 e_{nt} = \frac{1}{\Delta_n}\sum_{s=1}^{t-1} e_{ns}' B'^{t-1-s} C_2 e_{nt} \tag{3.19}$$

we have the decomposition

$$\operatorname{tr}(W_n C) = \sum_{t=1}^n (S_{nt} + T_{nt}). \tag{3.20}$$

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

3.8 Convergence of conditional second moments

Denote

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

Lemma 3.5. *Under Assumptions 1-4*

$$\lim \sum_{t=1}^n \mathrm{E}(S_{nt}^2 | \mathcal{F}_{n,t-1}) = \operatorname{tr}\int_0^1 C'\Omega_0(x) C \Sigma^c(x) dx, \tag{3.21}$$

$$L_2\text{-}\lim \sum_{t=1}^n \mathrm{E}(S_{nt} T_{nt} | \mathcal{F}_{n,t-1}) = 0, \tag{3.22}$$

$$L_1\text{-}\lim \sum_{t=1}^n \mathrm{E}(T_{nt}^2 | \mathcal{F}_{n,t-1}) = \kappa_Q^2 \operatorname{tr}\int_0^1 C_2'\Xi(x) C_2 \Sigma^c(x) dx. \tag{3.23}$$

Proof. From (2.10) and (3.18) we see that

$$\sum_{t=1}^n \mathrm{E}(S_{nt}^2 | \mathcal{F}_{n,t-1}) = \sum_{t=1}^n G_{nt}' \mathrm{E}(e_{nt} e_{nt}' | \mathcal{F}_{n,t-1}) G_{nt} = \sum_{t=1}^n G_{nt}' \Sigma_{nt} G_{nt}. \tag{3.24}$$

Using (3.16), by Theorem 3.1, items (iv) and (v), we get

$$K_n \equiv C_1' H_n + C_2' P_B A b_n H_n \text{ is } L_2\text{-close to } K^c \equiv (C_1' + C_2' J)H^c. \tag{3.25}$$

Since $\{\Sigma_n\}$ is L_∞-close to Σ^c, by Theorem 3.1(ii)

$$\sum_{t=1}^n K_{nt}' \Sigma_{nt} K_{nt} \to \int_0^1 (K^c)' \Sigma^c K^c dx. \tag{3.26}$$

Note that

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

so

$$\int_0^1 (K^c)'\Sigma^c K^c dx = \operatorname{tr} \int_0^1 K^c (K^c)' \Sigma^c dx = \operatorname{tr} \int_0^1 C'\Omega_0 C \Sigma^c dx. \tag{3.27}$$

Now (3.21) follows from (3.24), (3.26) and (3.27).

Using definitions (3.18) and (3.19) rearrange

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

This type of variable appeared in Lemma 3.3(a) with $X_{nt} = C_2 \Sigma_{nt} K_{nt}$. By (3.25), Assumption 3(iii) and Theorem 3.1(v) $\{X_n\}$ is L_2-close to $C_2 \Sigma^c K^c$, so by Lemma 3.3(a) (3.22) is true.

Since $(P_B \mathcal{E}_n)_t$ is $\mathcal{F}_{n,t-1}$-measurable, (3.19) implies

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

Here $X_n = C_2 \Sigma_n C_2'$ is L_∞-close to $C_2 \Sigma^c C_2'$ by Assumption 3(iii) and Theorem 3.1(v). Therefore (3.23) follows from Lemma 3.3(b) and Assumption 4. □

3.9 Convergence of $\operatorname{tr}(W_n C)$

Denote

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

Lemma 3.6. *If Assumptions 1 through 4 hold, then for any constant matrix C*

$$\operatorname{tr}(W_n C) \xrightarrow{d} N(0, \sigma^2). \tag{3.28}$$

Proof. We are going to apply Theorem 3.3. According to (3.20) we need to consider $X_{nt} = S_{nt} + T_{nt}$. By Lemma 3.5 we have a stronger statement than (3.4):

$$L_1\text{- }\lim \sum_{t=1}^{n} \sigma_{nt}^2 = \operatorname{tr} \int_0^1 [C'\Omega_0 C\Sigma^c + \kappa_Q^2 C_2' \Xi(x) C_2 \Sigma^c] dx = \sigma^2.$$

The proof of (3.5) 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 $\mathcal{F}_{n,t-1}$-measurable and by Lemma 3.3(b) and Assumption 4

$$\begin{aligned} L_1\text{- }\lim \sum_{t=1}^{n} \phi_{nt}^2 &= L_1\text{- }\lim \frac{n}{\Delta_n^2} \operatorname{tr} \frac{1}{n} \sum_{t=1}^{n} (P_B\mathcal{E}_n)_t (P_B\mathcal{E}_n)_t' \\ &= \kappa_Q^2 \operatorname{tr} \int_0^1 \Xi(x) dx. \end{aligned} \tag{3.29}$$

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

$$\mathrm{E}I(\phi_{nt} > \delta) \le \frac{1}{\delta \Delta_n} \mathrm{E} \left\| \sum_{k=1}^{t-1} B^{t-1-k} e_{nk} \right\|_2 \le \frac{c_1}{\delta \Delta_n}. \tag{3.30}$$

By the Minkowski inequality and Assumption 3(ii)

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

With $p_1 = p/2,\ q_1 = p_1/(p_1 - 1)$ from the last two bounds we get

$$\begin{aligned} \mathrm{E}\phi_{nt}^2 I(\phi_{nt} > \delta) &\le (\mathrm{E}I(\phi_{nt} > \delta))^{1/q_1} (\mathrm{E}|\phi_{nt}|^{2p_1})^{1/p_1} \\ &\le \frac{c_3}{(\delta\Delta_n)^{1/q_1} \Delta_n^2}. \end{aligned} \tag{3.31}$$

Since $\{K_n\}$ is L_2-approximable (see (3.25)), by (2.9) and Theorem 3.1(i) there exists $n_0 = n_0(\delta)$ such that

$$\sup_{n\ge1} ||K_n; l_2(\tau_n, \mathbb{M}_2)|| < \infty,\ \sup_{n\ge1} \max_{1\le t\le n} ||K_{nt}||_2 \le \delta. \tag{3.32}$$

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

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

This together with

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

allows us to proceed with proving (3.5):

$$\begin{aligned} &\sum_{t=1}^{n} \mathrm{E}(X_{nt}^2 I(|X_{nt}| > \varepsilon)|\mathcal{F}_{n,t-1}) \\ &\le c \sum_{t=1}^{n} (||K_{nt}||_2^2 + \phi_{nt}^2) \mathrm{E}\left(||e_{nt}||_2^2 I\left(||e_{nt}||_2 > \frac{\varepsilon}{2\delta c} \right) |\mathcal{F}_{n,t-1} \right) \\ &\quad + c \sum_{t=1}^{n} (||K_{nt}||_2^2 + \phi_{nt}^2) I(\phi_{nt} > \delta) \mathrm{E}\left(||e_{nt}||_2^2 |\mathcal{F}_{n,t-1} \right). \end{aligned} \tag{3.33}$$

By (3.29) and (3.32)

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

which in combination with Assumption 3(iv) leads to

$$\begin{aligned} &\sum_{t=1}^{n}(||K_{nt}||_2^2+\phi_{nt}^2)\mathrm{E}\left(||e_{nt}||_2^2 I\left(||e_{nt}||_2>\frac{\varepsilon}{2\delta c}\right)|\mathcal{F}_{n,t-1}\right) \qquad (3.34)\\ \le\ &\alpha_n \sup_{n,t}\mathrm{E}\left(||e_{nt}||_2^2 I\left(||e_{nt}||_2>\frac{\varepsilon}{2\delta c}\right)|\mathcal{F}_{n,t-1}\right)\xrightarrow{p}0,\ \delta\to 0. \end{aligned}$$

Further, application of (3.30), (3.31) and (3.32) results in

$$\begin{aligned} &\mathrm{E}\sum_{t=1}^{n}(||K_{nt}||_2^2+\phi_{nt}^2)I(\phi_{nt}>\delta)\mathrm{E}\left(||e_{nt}||_2^2|\mathcal{F}_{n,t-1}\right) \qquad (3.35)\\ =\ &\sum_{t=1}^{n}||K_{nt}||_2^2\mathrm{E}I(\phi_{nt}>\delta)\mathrm{tr}\Sigma_{nt}+\sum_{t=1}^{n}\mathrm{E}\phi_{nt}^2 I(\phi_{nt}>\delta)\mathrm{tr}\Sigma_{nt}\\ \le\ &\frac{c_1}{\delta\Delta_n}+\frac{c_2 n}{(\delta\Delta_n)^{1/q_1}\Delta_n^2}\to 0,\ n\to\infty, \end{aligned}$$

for any $\delta>0$, because $q_1<\infty$. The left side of (3.34) can be made small uniformly in n by choosing a small δ. For the selected δ, the left side of (3.35) can be made small by taking n sufficiently large. Then (3.33), (3.34) and (3.35) prove (3.5). By Theorem 3.3, (3.28) follows. □

3.10 Proofs of Theorems 2.1 and 2.2

Proof of Theorem 2.1. (i) We consider the blocks of the matrix at the right of (3.15) one by one. By Assumption 2 and Theorem 3.1(ii)

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

Denote $F_n=P_BAb_nH_n$. From (2.6), (3.16), Theorem 3.1(ii) and Lemma 3.3(a)

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

The block in the lower right corner of (3.15) equals

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

Here by (3.16) and Lemma 3.3(a)

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

so by Lemma 3.3(b) and Assumption 4

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

The proof is complete.

(ii) Suppose $|Q| \neq 0$. If $|G| = 0$, then some row of G is a linear combination of the 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 the 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. Equation (2.11) implies (a) and (c). Sufficiency of (a), (b) and (c) also follows from (2.11). □

Proof of Theorem 2.2. (i) Lemma 3.4 reduces convergence of the N-factor to that of W_n. By Lemma 3.1 W_n converges if $\mathrm{tr}(W_n C)$ converges for any C. This last convergence has been established in Lemma 3.6. Lemma 3.1 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 2.1. □

Proof of Corollary 2.1. Convergence of $\mathrm{vec}\, U_n$ and $\mathrm{vec}\, V_n$ is a consequence of $\mathrm{vec}\,(\mathcal{E}_n Z_n' D_n^{-1}) = \begin{pmatrix} \mathrm{vec}\, U_n \\ \mathrm{vec}\, V_n \end{pmatrix}$ and (2.12). Denoting $\mathcal{G} = \int_0^1 [H^c(H^c)'] \otimes \Sigma^c dx$, we can write the variance matrix in (2.12) 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 $\mathrm{vec}\, U = (J \otimes I)\mathrm{vec}\, U$, so that $\begin{pmatrix} \mathrm{vec}\, U \\ \mathrm{vec}\, V \end{pmatrix}$ has the same variance. Since a normal vector is uniquely defined by its mean and variance, this proves the corollary. □

3.11 Example of a proper mixed autoregression

Lemma 3.7. *If A and B are defined by* (2.13) *with $\alpha_1 ... \alpha_r \neq 0$ and $|G| \neq 0$, then model* (1.1) *is a proper mixed autoregression.*

Proof. By definition, we need to fix b with at least one nonzero diagonal element. Since none of the alphas is zero, the vector $\zeta = (\alpha_1 \kappa_1, ..., \alpha_r \kappa_r)'$ is not null. Let $e_i = (0, ..., 0, 1, 0, ..., 0)'$ denote the ith unit vector in $\mathbb{R}^s$. Then

$$Ab = \begin{pmatrix} \alpha_1 & ... & \alpha_r \\ ... & ... & ... \\ 0 & ... & 0 \end{pmatrix} \begin{pmatrix} \kappa_1 & & 0 \\ & \ddots & \\ 0 & & \kappa_r \end{pmatrix} = \begin{pmatrix} \zeta' \\ 0 \end{pmatrix}$$

and

$$AbG(Ab)' = \begin{pmatrix} \zeta' \\ 0 \end{pmatrix} G \begin{pmatrix} \zeta & 0 \end{pmatrix} = \zeta' G \zeta e_1 e_1'$$

where $\zeta' G \zeta > 0$ because G is positive definite. Hence,

$$JGJ' = \zeta' G \zeta (I - B)^{-1} e_1 e_1' (I - B')^{-1}. \tag{3.36}$$

Let us prove that

$$e_i' B^k = e_{i-1}' B^{k-1} \text{ for all } k \geq 1,\ 2 \leq i \leq s. \tag{3.37}$$

To apply induction in k, consider $k=1$ first. $e_i'B$ is the ith row of B and $e_{i-1}'I$ is the ith row of I, so (3.37) is true. Now suppose it holds for some $k\geq 1$. Denoting b_{lm}^k (b_{lm}^{k-1}) the elements of B^k (B^{k-1}, respectively), by the induction assumption we have

$$(b_{i1}^k,...,b_{is}^k)=(b_{i-1,1}^{k-1},...,b_{i-1,s}^{k-1}) \text{ for all } i.$$

This leads to

$$\begin{aligned} e_i'B^{k+1} &= e_i'\begin{pmatrix} b_{11}^k & ... & b_{1s}^k \\ ... & ... & ... \\ b_{s1}^k & ... & b_{ss}^k \end{pmatrix}\begin{pmatrix} \beta_1 & \beta_2 & ... & \beta_s \\ 1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{pmatrix} \\ &= (b_{i1}^k\beta_1+b_{i2}^k, b_{i1}^k\beta_2+b_{i3}^k,...,b_{i1}^k\beta_{s-1}+b_{is}^k, b_{i1}^k\beta_s) \\ &= (b_{i-1,1}^{k-1}\beta_1+b_{i-1,2}^{k-1},...,b_{i-1,1}^{k-1}\beta_s)=e_{i-1}'B^k. \end{aligned}$$

The proof of (3.37) is complete.

Repetitive application of (3.37) gives

$$\begin{aligned} e_i'(I-B)^{-1}e_1 &= \left(e_i'\sum_{k=0}^{i-2}B^k+e_i'\sum_{k=i-1}^{\infty}B^k\right)e_1 \\ &= \left(\sum_{k=0}^{i-2}e_{i-k}'B^{k-k}+\sum_{k=i-1}^{\infty}e_{i-(i-1)}'B^{k-(i-1)}\right)e_1 \\ &= \left(\sum_{k=0}^{i-2}e_{i-k}'I+e_1'\sum_{k=0}^{\infty}B^k\right)e_1=e_1'(I-B)^{-1}e_1. \end{aligned}$$

This equation entails equality of diagonal elements of $(I-B)^{-1}e_1e_1'(I-B')^{-1}$:

$$e_i'(I-B)^{-1}e_1e_1'(I-B')^{-1}e_i=e_1'(I-B)^{-1}e_1e_1'(I-B')^{-1}e_1,\ i=1,...,s.$$

Combining this with (3.36) we see that all diagonal elements of JGJ' are equal to $\zeta'G\zeta$. (In fact, a slight modification of the last step shows that all elements of JGJ' are the same). $\square$

3.12 Proof of Theorem 2.4 and Corollary 2.2

Proof of Theorem 2.4. **Step 1.** Let us prove that all diagonal elements of Q are positive. To avoid notational clutter, denote

$$K_n=D_n^{-1}Z_n=\begin{pmatrix} k_{n1} \\ ... \\ k_{n,r+s}\end{pmatrix},$$

so that by Theorem 2.1

$$Q_n=D_n^{-1}Z_nZ_n'D_n^{-1}=K_nK_n'=(k_{ni}k_{nj}')_{i,j=1}^{r+s}\xrightarrow{p}Q.$$

The first r diagonal elements of Q are unities because by construction $k_{ni}=h_n^{(i)}$ are the rows of H_n and $\left\|h_n^{(i)}\right\|_2=1$ for all n. Since the system is assumed to be a proper mixed autoregression, the last s diagonal elements of Q are positive if the balancer is not null. Suppose $b=0$, that is

$$\lim_{n\to\infty}\frac{d_{ni}}{\max\{d_{n1},...,d_{nr},\sqrt{n}\}}=0,\ i=1,...,r.$$

From these equations for the function $f_n = \max\{d_{n1}/\sqrt{n}, ..., d_{nr}/\sqrt{n}\}$ we obtain $\lim f_n / \max\{f_n, 1\} = 0$. Hence, $f_n \to 0$ and

$$\kappa_Q = \lim_{n\to\infty} \frac{\sqrt{n}}{\max\{d_{n1}, ..., d_{nr}, \sqrt{n}\}} = \lim \frac{1}{\max\{f_n, 1\}} = 1,$$

which contradicts the assumption $\kappa_Q = 0$.

Step 2. All objects in the parallel world (with the alternative normalizer $\widetilde{D}_n$) will be capped with a tilde. Thus, there are diagonal elements $\tilde{d}_{ni}$ of $\widetilde{D}_n$, rows $\tilde{k}_{ni}$ of $\widetilde{K}_n$ and elements $\tilde{q}_{ij}$ of $\widetilde{Q}$. Here we show that all diagonal elements of $\widetilde{Q}$ are positive.

Suppose $\tilde{q}_{ii} = 0$ for some i. $\widetilde{Q}_n \xrightarrow{p} \widetilde{Q}$ implies $\tilde{k}_{ni}\tilde{k}'_{nj} \xrightarrow{p} \tilde{q}_{ij}$ for all i, j. By the Cauchy-Schwartz inequality for any j

$$|\tilde{k}_{ni}\tilde{k}'_{nj}| \le \left\|\tilde{k}_{ni}\right\|_2 \left\|\tilde{k}_{nj}\right\|_2 = (\tilde{k}_{ni}\tilde{k}'_{ni}\tilde{k}_{nj}\tilde{k}'_{nj})^{1/2}.$$

By taking limits we get $|\tilde{q}_{ij}| \le (\tilde{q}_{ii}\tilde{q}_{jj})^{1/2}$ for any j. Thus, the whole ith row in $\widetilde{Q}$ is null and $|\widetilde{Q}| = 0$, contradicting (2.14). Hence, our assumption is wrong and $\tilde{q}_{ii}$ is positive (it cannot be negative because $\tilde{q}_{ii} = \text{plim} \left\|\tilde{k}_{ni}\right\|_2^2$).

Step 3. Denote $r_{ni} = (\tilde{d}_{ni}/d_{ni})^2$, $i = 1, ..., r+s$, where $d_{n,r+i} = \Delta_n$ for $i = 1, ..., s$. Here we prove that the limits $\text{plim} r_{ni} = r_i$ exist.

Letting $C_n = D_n^{-1}\widetilde{D}_n$ we note the relationship between K_n and $\widetilde{K}_n$:

$$K_n = D_n^{-1}\widetilde{D}_n\widetilde{D}_n^{-1}Z_n = C_n\widetilde{K}_n \tag{3.38}$$

which implies $k_{ni} = (\tilde{d}_{ni}/d_{ni})\tilde{k}_{ni}$. Recalling that the diagonal elements of both Q and $\widetilde{Q}$ are not zero, we see that the limits

$$\text{plim} r_{ni} = \text{plim}\frac{k_{ni}k'_{ni}}{\tilde{k}_{ni}\tilde{k}'_{ni}} = \frac{q_{ii}}{\tilde{q}_{ii}} \tag{3.39}$$

exist and are positive

Step 4. (3.38) implies $Q_n = K_nK'_n = C_n\widetilde{K}_n\widetilde{K}'_nC_n$. This equation and the assumptions $|Q| = 0$ and (2.14) give

$$\prod_{i=1}^{r+s}\left(\frac{\tilde{d}_{ni}}{d_{ni}}\right)^2 = |C_n|^2 = \frac{|Q_n|}{|\widetilde{Q}_n|} \xrightarrow{p} 0.$$

This obviously contradicts (3.39). Thus (2.14) is impossible. □

Proof of Corollary 2.2. Step 2 of the proof of Theorem 2.4, applied to Q, shows that if $|Q| \ne 0$, then all diagonal elements of Q are positive. In case $|Q| = 0$ the same conclusion is true by Step 1. Thus, regardless of the value of $|Q|$, the limits r_i are positive and the limits $\lambda_i = \text{plim}\tilde{d}_{ni}/d_{ni}$ are nonzero. We obtain the uniqueness statement $\widetilde{D}_n = D_nD_n^{-1}\widetilde{D}_n = D_nC_n$ with $\text{plim}C_n = \text{diag}[\lambda_1, ..., \lambda_{r+s}] \equiv \Lambda$.

Suppose the conventional scheme works with $\widetilde{D}_n$, that is the limits

$$\text{dlim}\mathcal{E}_nZ'_n\widetilde{D}_n^{-1},\ \widetilde{Q} = \text{plim}\widetilde{D}_n^{-1}Z_nZ'_n\widetilde{D}_n^{-1}$$

exist and $|\widetilde{Q}| \ne 0$. Then the limits

$$\text{dlim}\mathcal{E}_nZ'_nD_n^{-1} = \text{dlim}\mathcal{E}_nZ'_n\widetilde{D}_n^{-1}\Lambda,\ Q = \text{plim}D_n^{-1}Z_nZ'_nD_n^{-1} = \Lambda\widetilde{Q}\Lambda$$

also exist and $|Q| \ne 0$. □

Acknowledgments

The author thanks the anonymous referee for valuable suggestions. This work was supported by project no. 4084-GF4 of the Ministry of Education and Science of the Republic of Kazakhstan.

References

[1] T.W. Anderson, *The Statistical Analysis of Time Series*, Wiley & Sons, 1971.

[2] T. Amemiya, *Advanced Econometrics*, Oxford, Blackwell, 1985.

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

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

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

[6] D.W.K. Andrews, C.J. McDermott, *Nonlinear econometric models with deterministically trending variables,* Rev. Econ. Stud. 62 (1995), 343–360.

[7] N.H. Bingham, C.M. Goldie, J.L. Teugels, *Regular Variation,* Cambridge University Press, 1987.

[8] D.L. Burkholder, *Distribution function inequalities for martingales,* Ann. Prob. 1 (1973), 14–42.

[9] W. Charemza, D.F. Deadman, *New Directions in Econometric Practice: General to Specific Modelling, Cointegration, and Vector Autoregression*, E. Elgar, 1992.

[10] J. Davidson, *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press, 1994.

[11] A. Dvoretzky, *Asymptotic normality for sums of dependent variables,* Proceedings of the Sixth Berkeley Symposium in Mathematical Statistics and Probability, University of California Press, 1972, 513–555.

[12] F.R. Gantmacher, *Matrizentheorie*, VEB Deutscher Verlag der Wissenschaften, 1986.

[13] P. Hall, C.C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, 1980.

[14] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.

[15] T.-H. Kim, S. Pfaffenzeller, T. Rayher, P. Newbold, *Testing for linear trend with applications to relative commodity prices,* J. Time Ser. Anal. 24 (2003), 539–551.

[16] H. Lütkepohl, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, 1991.

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

[18] K.T. Mynbaev, *Asymptotic properties of OLS estimates in autoregressions with bounded or slowly growing deterministic trends,* Commun. Stat. Theory. 35 (2006), 499–520.

[19] K.T. Mynbaev, *Short-memory linear processes and econometric applications*, Wiley & Sons, 2011.

[20] K.T. Mynbaev, *Central limit theorems for weighted sums of linear processes: L_p-approximability versus Brownian motion,* Economet. Theory, 25 (2009), 748-763.

[21] K.T. Mynbaev, A. Ullah, *Asymptotic distribution of the OLS estimator for a purely autoregressive spatial model,* J. Multivariate Anal. 99 (2008), 245–277.

[22] K.T. Mynbaev, *Asymptotic distribution of the OLS estimator for a mixed regressive, spatial autoregressive model,* J. Multivariate Anal. 101 (2010), 733–748.

[23] K.T. Mynbaev, I. Castelar, *The Strengths and Weaknesses of L_2-approximable Regressors. Two Essays on Econometrics*, Fortaleza: Expressão Gráfica. 1, 2001. http://mpra.ub.uni-muenchen.de/9056/

[24] B. Nielsen, *Strong consistency results for Least Squares estimators in general vector autoregressions with deterministic terms,* Economet. Theory. 21 (2005), 534–561.

[25] P.C.B. Phillips, *Regression with slowly varying regressors and nonlinear trends,* Economet. Theory. 23 (2007), 557-614.

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

[27] C.A. Sims, J.H. Stock, M.W. Watson, *Inference in linear time series models with some unit roots,* Econometrica. 58 (1990), 113–144.

[28] K. Tanaka, *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, Wiley & Sons, 1996.

Kairat T. Mynbaev
International School of Economics
Kazakh-British Technical University
Tolebi 59, Room 419,
050035 Almaty, Kazakhstan
E-mail: kairat_mynbayev@yahoo.com

Received: 06.10.2016
Revised version: 01.02.2017