Asymptotic properties of least squares statistics in general vector autoregressive models.

Bent Nielsen

Department of Economics, University of Oxford
& Nuffield College, Oxford OX1 1NF, UK
E-mail: bent.nielsen@nuf.ox.ac.uk
http://www.nuff.ox.ac.uk/users/nielsen

12 July 2001

Summary: A vector autoregression with deterministic terms and with no restrictions to its characteristic roots is considered. Strong consistency results and also some weak convergence results are given for a number of least squares statistics. These statistics are related to the denominator matrix of the least squares estimator as well as the least squares estimator itself. Applications of these results to the statistical analysis of non-stationary economic time-series are briefly discussed.

Keywords: *Asymptotic normality, Cointegration, Least squares, Martingales, Sample correlations, Strong consistency, Vector autoregressive model, Weak consistency.*

1 Introduction

In the asymptotic analysis of autoregressive time series it is usually assumed that the process has no explosive roots. This is somewhat in contrast with the statistical analysis which is performed without that restriction. Often the assumption of non-explosiveness will not be necessary and consequently the autoregressive model can be used for analysing economic data exhibiting explosive growth. Examples of this are the strong consistency results for least squares estimators proved by Rubin (1950) and Lai and Wei (1985). While Lai and Wei focussed on the least squares estimator this paper is oriented towards sample correlations which have a more natural normalisation than the least squares estimator. The results are to a large extent derived using methods presented by Lai and Wei (1982a,b, 1983a,b, 1985) and find applications in autoregression problems such as lag determination, see Pötscher (1989) and Nielsen (2001b), unit root testing (Nielsen 2001a) and cointegration analysis (Nielsen, 2000).

The model in this paper is a p-dimensional time series, $X_{1-k}, \ldots, X_0, \ldots, X_T$ satisfying a kth order vector autoregressive equation

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_t + \varepsilon_t, \qquad t = 1, \ldots, T. \tag{1.1}$$

Here the component D_t is a vector of deterministic terms such as a constant, a linear trend, or seasonal dummies, while ε_t is an innovation term. For convenience let $\mathbf{X}_t = (X'_{t-1}, \ldots, X'_{t-k})'$ denote the stacked vector of lags of the process.

The aim of the paper is to describe, in an almost sure sense, the order of magnitude of two statistics arising in regression analysis. The first of these statistics is the denominator of the least squares estimator

$$\sum_{t=1}^{T} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_t \end{pmatrix} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_t \end{pmatrix}', \tag{1.2}$$

for which the smallest and largest eigenvalues will be discussed. To get a more detailed understanding of (1.2) the process $\mathbf{X}_t$ is decomposed into processes U_t, V_t, W_t with characteristic roots whose absolute values are smaller than, equal to, or larger than one. The sample correlations between these components vanish asymptotically and the rate of convergence is discussed. These results for the denominator matrix will facilitate a discussion of the normalised least squares estimator which is the second statistic of interest,

$$\left\{\sum_{t=1}^{T} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_t \end{pmatrix} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_t \end{pmatrix}'\right\}^{-1/2} \sum_{t=1}^{T} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_t \end{pmatrix} \varepsilon'_t. \tag{1.3}$$

The two sets of results generalise those of Lai and Wai (1983b, 1985) in that results concerning the order of magnitude of sample covariances are sharpened and since deterministic terms are included. An immediate consequence of the two results is that the sum of squared residuals in the regression given by (1.1) is asymptotically equivalent to $\sum_{t=1}^{T} \varepsilon_t \varepsilon'_t$ whereas more involved applications are given by Nielsen (2000, 2001a,b).

When discussing the order of magnitude of the above statistics the arguments of Lai and Wei can be followed to a large extent. Following their precedence it is assumed that the sequence of innovations (ε_t) is a martingale difference sequence with respect to an increasing sequence of σ-fields $(\mathcal{F}_t)$, that is ε_t is $\mathcal{F}_t$-measurable with $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ $a.s.$ It will often be required that the innovations satisfy the two conditions

$$\sup_t \mathsf{E}\left(\|\varepsilon_t\|^{2+\gamma} |\mathcal{F}_{t-1}\right) \overset{a.s.}{<} \infty \qquad \text{for some } \gamma > 2, \tag{1.4}$$

$$\liminf_{T\to\infty} \lambda_{\min} \mathsf{E}\left(\varepsilon_t \varepsilon'_t|\mathcal{F}_{t-1}\right) \overset{a.s.}{>} 0, \tag{1.5}$$

where $\lambda_{\min}$ denotes the smallest eigenvalue of a symmetric matrix. Lai and Wei (1983a) refer to (1.4), (1.5) as the local Marcinkiewicz-Zygmund conditions.

The deterministic process will be assumed to satisfy the assumption

$$D_t = \mathbf{D} D_{t-1}, \qquad |\text{eigen}\,(\mathbf{D})| = 1, \qquad \text{rank}\,(D_1, \ldots, D_{\dim \mathbf{D}}) = \dim \mathbf{D}, \tag{1.6}$$

which is inspired by Johansen (2000).

The outline of the paper is as follows.

§2: Here the process $\mathbf{X}_t$ is decomposed into the three processes U_t, V_t, W_t.

§3: The deterministic process D_t and in particular the order of magnitude for (1.2) for the purely deterministic case is discussed.

§4: The order of magnitude of the process $\mathbf{X}_t$ is discussed.

§5: The sample correlation of U_t and D_t is proved to vanish asymptotically.

§6: The order of magnitude of the largest eigenvalue of (1.2) is found.

§7: The non-explosive components and in particular the order of magnitude of the smallest eigenvalue of (1.2) for this case is analysed.

§8: The order of magnitude of the sample correlations of U_t, V_t, W_t, D_t and of the smallest eigenvalue of (1.2) is discussed.

§9: The least squares statistic (1.3) is discussed

§10: Some Central Limit Theorems which can be used in the analysis of (1.3) are presented.

§11: The process V_t with characteristic roots of length one can be decomposed further into processes with unit roots at different frequencies. The sample correlations between these are discussed. These results only hold weakly.

The Sections 5, 8, 9 also contain some examples showing how the results can be applied in the analysis of non-stationary time series.

The following notation is used throughout the paper: For a matrix α let $\alpha^{\otimes 2} = \alpha\alpha'$ and let $\|\alpha\|$ be the Euclidean norm. When α is symmetric then $\lambda_{\min}(\alpha)$ and $\lambda_{\max}(\alpha)$ denote the smallest and the largest eigenvalue respectively. While $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1})$ is a conditional expectation the notation $(Y_t|Z_t)$ denotes the residual of the least squares regression of Y_t on Z_t. The abbreviations *a.s.* and P are used for properties holding almost surely and in probability, respectively.

2 Decompositions of the process

The process X_t satisfying (1.1) is decomposed into stationary and explosive components as well as components with roots at various locations on the unit circle. In later sections it will be established that the sample correlations of these components are asymptotically negligible.

As a first step write the process on companion form

$$\begin{pmatrix} \mathbf{X}_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \boldsymbol{\mu} \\ 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_t \\ 0 \end{pmatrix}, \qquad (2.1)$$

where

$$\mathbf{B} = \left\{ \begin{array}{cc} A_1 \cdots A_{k-1} & A_k \\ I_{p(k-1)} & 0 \end{array} \right\}, \qquad \boldsymbol{\iota} = \left\{ \begin{array}{c} I_p \\ 0_{(k-1)p \times p} \end{array} \right\}, \qquad \boldsymbol{\mu} = \boldsymbol{\iota}\mu\mathbf{D}, \qquad \mathbf{e}_t = \boldsymbol{\iota}\varepsilon_t.$$

The eigenvalues of $\mathbf{B}$ are the characteristic roots of the process X_t.

The decomposition can now be introduced using a similarity transformation. Following Herstein (1975, p. 308) there exists a regular, real matrix M which transforms $\mathbf{B}$ into a real, rational canonical form. In particular, M can be chosen so $M\mathbf{B}M^{-1} = \text{diag}\,(\mathbf{U}, \mathbf{V}, \mathbf{W})$ is a block diagonal matrix where the absolute values of the eigenvalues of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are smaller than one, equal to one and at least one, respectively. Correspondingly, define the processes

$$M\mathbf{X}_t = \begin{pmatrix} U_t \\ V_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 & 0 & \mu_U \\ 0 & \mathbf{V} & 0 & \mu_V \\ 0 & 0 & \mathbf{W} & \mu_W \end{pmatrix} \begin{pmatrix} U_{t-1} \\ V_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{U,t} \\ e_{V,t} \\ e_{W,t} \end{pmatrix}.$$

The process V_t which has roots on the unit circle can be transformed further, into components $V_{1,t}, \ldots V_{l,t}$ so $V_{j,t}$ has characteristic roots at $\exp(i\theta_j)$ and $\exp(-i\theta_j)$ where $\theta_1, \ldots, \theta_l$ are distinct and $0 \leq \theta_j \leq \pi$.

The deterministic term D_t has roots on the unit circle and no other roots. The Lemma 2.1 given below shows that one further similarity transformation leads to the representations

$$U_t = \tilde{U}_t + \tilde{\mu}_U D_t \quad \text{where} \quad \tilde{U}_t = \mathbf{U}\tilde{U}_{t-1} + e_{U,t}, \qquad (2.2)$$

$$W_t = \tilde{W}_t + \tilde{\mu}_W D_t \quad \text{where} \quad \tilde{W}_t = \mathbf{W}\tilde{W}_{t-1} + e_{W,t}, \qquad (2.3)$$

for some $\tilde{\mu}_U$ and $\tilde{\mu}_W$. An example of this is when $D_t = 1$ in which case $\tilde{\mu}_U = (I - \mathbf{U})^{-1}\mu$. It is an immediate consequence of the representations (2.2), (2.3) that

$$\left(\tilde{U}_t | D_t\right) = (U_t | D_t), \qquad \left(\tilde{W}_t | D_t\right) = (W_t | D_t).$$

The relation between V_t and the process $\tilde{V}_t = \mathbf{V}\tilde{V}_{t-1} + e_{V,t}$ is more complicated since $\mathbf{D}$ and $\mathbf{V}$ have common eigenvalues and the process $\tilde{V}_t$ will therefore not be discussed.

The representations (2.2), (2.3) follow immediately from the next Lemma since the matrices $\mathbf{U}$, $\mathbf{W}$, $\mathbf{D}$ have no common eigenvalues.

Lemma 2.1 *Let $A \in \mathbf{R}^{a\times a}$, $B \in \mathbf{R}^{b\times b}$ be square matrices while $C \in \mathbf{R}^{a\times b}$ is rectangular. Then the equation $AD - DB = C$ has a unique solution $D \in \mathbf{R}^{a\times b}$ if and only if the matrices A and B have no common eigenvalues.*

Proof of Lemma 2.1. The equation can be rewritten using the vec-operator as

$$(A' \otimes I_b - I_a \otimes B)\,\mathrm{vec}\,D = \mathrm{vec}\,C,$$

see Magnus and Neudecker (1999, Chapter 2). A unique solution can therefore be found when the matrix $A' \otimes I_b - I_a \otimes B$ has full rank or, in other words, when it has no zero eigenvalues.

Two properties of Kronecker products are needed. First, the two matrices $A' \otimes I_b$ and $I_a \otimes B$ commute and hence they are simultaneously unitarily similar to triangular matrices (Mirsky, 1961, Theorem 10.6.5). Secondly, a Kronecker product $F \otimes G$ has eigenvalues of the form $f_i g_j$ where f, g are the eigenvalues of F, G respectively, see Magnus and Neudecker (1999, Theorem 2.3.1). As a consequence the matrix $A' \otimes I_b - I_a \otimes B$ is unitarily similar to a triangular matrix with diagonal elements given by $a_i - b_j$ where a_i and b_j are eigenvalues of A and B respectively. Hence, if A and B have no common eigenvalues then $A' \otimes I_b - I_a \otimes B$ has full rank.

Now suppose A and B have a common eigenvalue λ and let x and y be associated eigenvectors so $A'x = \lambda x$ and $By = \lambda y$. Since $(A' \otimes I_b - I_a \otimes B)(x \otimes y) = 0$ the vector $x \otimes y$ is an eigenvector for $A' \otimes I_b - I_a \otimes B$ associated with the eigenvalue 0. ■

3 Limiting results for the deterministic component

In the following some limit results will be given for the order of magnitude of the deterministic process D_t and the sum of squares $\sum_{t=1}^{T} D_t^{\otimes 2}$. The formulation of the results is inspired by Lai and Wei (1983b) as will become evident in §7.

The results in this section exploit a similarity transformation as in §2. Because of Assumption (1.6) there exists a regular, real matrix M so $MD_t = (D_{1,t}', \ldots, D_{n,t}')'$ and $M\mathbf{D}M^{-1}$ has the form

$$\mathrm{diag}\left(\mathbf{D}_1, \ldots, \mathbf{D}_n\right), \tag{3.1}$$

where the blocks have different eigenvalues and each block, $\mathbf{D}_j$, is a real Jordan matrix

$$\mathbf{D}_j = \begin{pmatrix} \Lambda & I & & \\ & \ddots & \ddots & \\ & & \ddots & I \\ & & & \Lambda \end{pmatrix}, \tag{3.2}$$

where (Λ, I) is one of the pairs

$$(1,1)\,, \qquad (-1,1)\,, \qquad \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}. \tag{3.3}$$

The multiplicity of the eigenvalues of each block, $\mathbf{D}_j$, is denoted δ_j and is given by $\dim \mathbf{D}_j / \dim \Lambda_j$. The maximum multiplicity of any eigenvalue of $\mathbf{D}$ is given by

$$\delta = \max_{1 \le j \le n} \delta_j = \max \{ \text{multiplicity}\,(\lambda) : \ \lambda \text{ is an eigenvalue of } \mathbf{D} \}\,. \tag{3.4}$$

The first result concerns the order of D_t.

Theorem 3.1 *Suppose Assumption 1.6 is satisfied. Then* $\|D_T\| = \mathrm{O}(T^\delta)$.

The proof will be formulated in terms of a Lemma which will be used once again in Section 4.

Lemma 3.2 *Let λ be a complex number with $|\lambda| = 1$ and let $\mathbf{D}_r, \mathbf{D}_s$ have dimension r and s respectively and be complex Jordan matrices of the form* (3.2) *with* $(\Lambda, I) = (\lambda, 1)$. *Further, let $\boldsymbol{\mu}$ be an $(r \times s)$ matrix. Then*

$$\left\| \sum_{t=1}^{T} \mathbf{D}_r^{T-t} \boldsymbol{\mu} \mathbf{D}_s^t \right\| = \mathrm{O}\left(T^{r+s-1} \right).$$

Proof of Lemma 3.2. The (j,m)-th element of $\mathbf{D}_r^t$ is an upper triangular band matrix with elements

$$\left(\mathbf{D}_r^t \right)_{j,m} = \binom{t}{m-j} \lambda^{m-j}$$

with the convention that the binomial coefficient (a,b) is 1 if $a = b = 0$ and 0 if either $a < b$ or $b < 0$. Therefore the (j,m)-th element of $\sum_{t=1}^{T} \mathbf{D}_r^{T-t} \boldsymbol{\mu} \mathbf{D}_s^t$ is

$$\left(\sum_{t=1}^{T} \mathbf{D}_r^{T-t} \boldsymbol{\mu} \mathbf{D}_s^t \right)_{l,m} = \sum_{t=1}^{T} \sum_{k=j}^{p} \binom{T-t}{k-j} \lambda^{T-t-k+j} \sum_{l=1}^{m} \boldsymbol{\mu}_{kl} \binom{t}{m-l} \lambda^{t-m+l}.$$

Taking absolute values, using $|\lambda| = 1$ and properties of the binomial coefficient shows

$$\begin{aligned}\left|\left(\sum_{t=1}^{T}\mathbf{D}_r^{T-t}\boldsymbol{\mu}\mathbf{D}_s^{t}\right)_{l,m}\right| &= \mathrm{O}\left\{\max_{k,l}|\boldsymbol{\mu}_{kl}|\sum_{t=1}^{T}\sum_{k=j}^{p}\binom{T-t}{p-j}\sum_{l=1}^{m}\binom{t}{m-1}|\lambda|^{T-k+j-m+l}\right\} \\ &= \mathrm{O}\left\{\sum_{t=1}^{T}(T-t)^{p-j}\,t^{m-1}\right\} = \mathrm{O}\left(T^{p-j+m}\right) = \mathrm{O}\left(T^{p+q-1}\right).\end{aligned}$$

Finally, use that the matrix of interest has a finite number of elements. ■

Proof of Theorem 3.1. Assume without loss of generality that $\mathbf{D}$ is of Jordan form (3.2). By the triangle inequality, $\|D_T\| \leq \sum_{j=1}^{n}\|D_{j,T}\|$ where each term $\|D_{j,T}\| = \mathrm{O}(T^{\dim D_j})$ according to Lemma 3.2. ■

A second result concerns the order of magnitude of the sum of squares of D_t.

Theorem 3.3 *Suppose Assumptions 1.6 is satisfied. Then there exists constants* $c_1, c_2 \in \mathbf{R}_+$ *so*

$$\lambda_{\min}\left(T^{-1}\sum_{t=1}^{T}D_t^{\otimes 2}\right) \to c_1, \qquad \lambda_{\max}\left(T^{1-2\delta}\sum_{t=1}^{T}D_t^{\otimes 2}\right) \to c_1,$$

and

$$\max_{t\leq T} D_t'\left(\sum_{s=1}^{T}D_s^{\otimes 2}\right)^{-1}D_t = \mathrm{O}\left(T^{-1}\right).$$

Two Lemmas are needed to prove this result, of which the first is essentially Theorem 3.3 formulated for the case where $\mathbf{D}$ only has eigenvalues at one frequency.

Lemma 3.4 *Suppose Assumption 1.6 is satisfied and that* $D_{j,t} = \mathbf{D}_j D_{j,t}$ *where* $\mathbf{D}_j$ *is a real Jordan block of the form* (3.2)*. Define the normalisation matrix*

$$N_{j,T} = T^{1/2}\mathrm{diag}\left(T^{-\dim \mathbf{D}_j}, \ldots, T^{-1}\right) \otimes I,$$

where I *is defined in* (3.3)*. Then it holds that* $\sum_{t=1}^{T}(N_{j,T}D_{j,t})^{\otimes 2}$ *converges to a positive definite matrix, and*

$$\max_{t\leq T} D_{j,t}'\left(\sum_{s=1}^{T}D_{j,s}^{\otimes 2}\right)^{-1}D_{j,t} = \mathrm{O}\left(T^{-1}\right).$$

Proof of Lemma 3.4. Only the most complicated case, where $\mathbf{D}_j$ has non-real eigenvalues is considered. In that case $I = I_2$ is the bivariate identity matrix and

$$\Lambda = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}\left\{\begin{array}{cc} \exp(i\theta) & 0 \\ 0 & \exp(-i\theta) \end{array}\right\}\begin{pmatrix} 1 & -1 \\ -i & 1 \end{pmatrix}. \qquad (3.5)$$

The matrix $\mathbf{D}_j^t$ is the upper block matrix with elements

$$\left(\mathbf{D}_j^t\right)_{m,n} = \binom{t}{n-m}\Lambda^{t-n+m}$$

as in the proof of Lemma 3.2. The binomial coefficient satisfies

$$\frac{1}{T^{n-m}}\binom{t}{n-m} = \frac{1}{(n-m)!}\left(\frac{t}{T}\right)^{n-m} + \mathrm{o}(1)$$

for large T where the error is uniform in t. Since $\|\Lambda\| = 1$ it follows that

$$\max_{t\le T} N_{j,T}D_{j,t} = \mathrm{O}\left(T^{-1/2}\right),$$

and the desired results follows by proving that $\sum_{s=1}^{T}\left(N_{j,T}D_{j,t}\right)^{\otimes 2}$ is convergent with a positive definite limit.

Now, let b, a denote the last and second last element of $D_{j,0}$, satisfying $a^2 + b^2 > 0$ by Assumption (1.6). The elements of the block matrix are therefore of the form

$$\begin{aligned} &\left\{(\dim\mathbf{D}_j - n)!\,(\dim\mathbf{D}_j - m)!\sum_{s=1}^{T}\left(N_{j,T}D_{j,t}\right)^{\otimes 2}\right\}_{m,n} \\ = \; & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\dim\mathbf{D}_j-n-m}\Lambda^{t-\dim\mathbf{D}_j+n}\begin{pmatrix} a \\ b \end{pmatrix}^{\otimes 2}\left(\Lambda^{t-\dim\mathbf{D}_j+m}\right)' + \mathrm{o}(1). \end{aligned}$$

Decomposing Λ as in (3.5) this expression becomes

$$\frac{1}{4}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} M_T \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}, \qquad (3.6)$$

where

$$M_T = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\dim\mathbf{D}_j-n-m}\left\{\begin{array}{cc} (a-ib)^2\exp(2it\theta) & -i(a+b)^2 \\ -i(a+b)^2 & (b-ia)^2\exp(-2it\theta) \end{array}\right\}.$$

The order of magnitude of this matrix can be judged by noting that

$$\sum_{t=1}^{T} t^k\exp(2it\theta) = \left\{\left(\frac{1}{2i}\right)\left(\frac{\partial}{\partial\theta}\right)\right\}^k\sum_{t=1}^{T}\exp(2it\theta) = \mathrm{O}\left(T^k\right) = \mathrm{o}\left(T^{k+1}\right), \qquad (3.7)$$

and it holds that

$$M_T = (a+b)^2 \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{2\dim \mathbf{D}_j - n - m} \right\} \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} + \mathrm{o}(1). \tag{3.8}$$

Inserting the expression (3.8) for M_T in (3.6) and noting that

$$\frac{1}{2} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

shows that

$$\lim_{T\to\infty} \sum_{t=1}^{T} (N_{j,T} D_{j,t})^{\otimes 2} = \left\{ \frac{(a+b)^2}{2} I_2 \right\} \otimes \left[\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left\{ \begin{matrix} \frac{(t/T)^{\dim \mathbf{D}_j - 1}}{(\dim \mathbf{D}_j - 1)!} \\ \vdots \\ \frac{(t/T)^0}{0!} \end{matrix} \right\}^{\otimes 2} \right].$$

The latter limit is positive definite since polynomials of increasing order constitute a basis for functions on $\mathbf{N}$. ■

A second Lemma is concerned with sample correlations of deterministic components with different roots.

Lemma 3.5 *Suppose Assumptions 1.6 is satisfied. Then*

$$\left(\sum_{t=1}^{T} D_{q,t}^{\otimes 2} \right)^{-1/2} \left(\sum_{t=1}^{T} D_{q,t} D_{r,t}' \right) \left(\sum_{t=1}^{T} D_{r,t}^{\otimes 2} \right)^{-1/2} = \mathrm{O}\left(\frac{1}{T} \right).$$

Proof of Lemma 3.5. Only the most complicated case where $\mathbf{D}_q$ and $\mathbf{D}_r$ both have non-real eigenvalues is considered. It suffices to argue that

$$\sum_{t=1}^{T} N_{q,T} D_{q,t} D_{r,t}' N_{r,T} = \mathrm{O}\left(\frac{1}{T} \right),$$

where $N_{q,T}$, $N_{r,T}$ are given in Lemma 3.4. The elements of this block matrix are

$$\begin{aligned} & \left\{ (\dim \mathbf{D}_q - n)!\, (\dim \mathbf{D}_r - m)! \sum_{t=1}^{T} N_{q,T} D_{q,t} D_{r,t}' N_{r,T} \right\}_{m,n} \\ = \; & \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{\dim \mathbf{D}_q + \dim \mathbf{D}_r - n - m} \Lambda^{t - \dim \mathbf{D}_q + n} \begin{pmatrix} a_q \\ b_q \end{pmatrix} \begin{pmatrix} a_r \\ b_r \end{pmatrix}' \left(\Lambda^{t - \dim \mathbf{D}_r + m} \right)' \{1 + \mathrm{o}(1)\} \\ = \; & \frac{1}{4} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{\dim \mathbf{D}_q + \dim \mathbf{D}_r - n - m} L_T \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}, \end{aligned}$$

where

$$L_T = \left[\begin{array}{c} a_q \exp\left\{i\theta_q \left(t - \dim \mathbf{D}_q + m\right)\right\} \\ b_q \exp\left\{-i\theta_q \left(t - \dim \mathbf{D}_q + m\right)\right\} \end{array} \right] \left[\begin{array}{c} a_r \exp\left\{i\theta_r \left(t - \dim \mathbf{D}_r + n\right)\right\} \\ b_r \exp\left\{-i\theta_r \left(t - \dim \mathbf{D}_r + n\right)\right\} \end{array} \right]'.$$

The desired result follows as in the proof of Lemma 3.4 using (3.7). ■

Proof of Theorem 3.3. Decompose D_t as indicated above. The components of D_t are asymptotically uncorrelated by Lemma 3.5 and the result then follows by Lemma 3.4. ■

4 The order of magnitude of the process

In the following the order of magnitude of the process $\mathbf{X}_t$ will be investigated. This is a generalisation of Lai and Wei (1985, Theorem 1) where the case without deterministic components is considered. The result will be formulated in terms of the multiplicity of the largest eigenvalue of $\mathbf{B}$, that is

$$\rho = \max\left\{\text{multiplicity}\left(\lambda\right): \ \lambda \text{ is an eigenvalue of } \mathbf{B} \text{ and } \left|\lambda\right| = \max\left|\text{eigen}\left(\mathbf{B}\right)\right|\right\}. \tag{4.1}$$

Theorem 4.1 *Suppose Assumptions 1.4, 1.6 are satisfied. Then, for* $\xi < \gamma / \left(2+\gamma\right)$,

$$\left\|\mathbf{X}_T\right\| \overset{a.s.}{=} \begin{cases} \mathrm{o}\{T^{(1-\xi)/2}\} + \mathrm{O}(T^{\delta}) & \textit{if } \max\left|\text{eigen}\left(\mathbf{B}\right)\right| < 1, \\ \mathrm{O}\{(T^{2\rho-1}\log\log T)^{1/2}\} + \mathrm{O}(T^{\rho+\delta-1}) & \textit{if } \max\left|\text{eigen}\left(\mathbf{B}\right)\right| = 1, \\ \mathrm{O}(T^{\rho-1}\max\left|\text{eigen}\left(\mathbf{B}\right)\right|^T) & \textit{if } \max\left|\text{eigen}\left(\mathbf{B}\right)\right| > 1. \end{cases}$$

Proof of Theorem 4.1. If $\max\left|\text{eigen}\left(\mathbf{B}\right)\right| < 1$ then by (2.2) and the triangle inequality

$$\left\|\mathbf{X}_T\right\| = \left\|U_T\right\| = \left\|\tilde{U}_T + \tilde{\mu}_U D_T\right\| = \mathrm{O}\left(\left\|\tilde{U}_T\right\|\right) + \mathrm{O}\left(\left\|D_T\right\|\right).$$

Use that $||\tilde{U}_T|| = \mathrm{o}\{T^{(1-\xi)/2}\}$ *a.s.* as shown by Lai and Wei (1985, Theorem 1,*i*) together with Theorem 3.1.

When $\max\left|\text{eigen}\left(\mathbf{B}\right)\right| = 1$ the triangle inequality shows $\left\|\mathbf{X}_T\right\| \le \left\|V_T\right\| + \left\|U_T\right\|$. Thus it suffices to consider $\left\|V_T\right\|$. A similarity transformation of $(V_t', D_t')'$ results in components of the type

$$\left(\begin{array}{c} V_{j,t} \\ D_{j,t} \end{array}\right) = \left(\begin{array}{cc} \mathbf{D}_{\dim V_j} & \tilde{\mu}_{V_j} \\ 0 & \mathbf{D}_{\dim D_j} \end{array}\right)\left(\begin{array}{c} V_{j,t-1} \\ D_{j,t-1} \end{array}\right) + \left(\begin{array}{c} \mathbf{e}_{V_j,t} \\ 0 \end{array}\right),$$

where $\mathbf{D}_{\dim V_j}$ and $\mathbf{D}_{\dim D_j}$ are Jordan matrices with a common eigenvalue. It suffices to show that $V_{j,T}$ is of the right order for each j. Now,

$$V_{j,T} = \sum_{t=1}^{T} \mathbf{D}_{\dim V_j}^{T-t} \mathbf{e}_{V_j,t} + \sum_{t=1}^{T} \mathbf{D}_{\dim V_j}^{T-t} \tilde{\mu}_{V_j} \mathbf{D}_{\dim D_j}^{t-1} D_{j,0},$$

where the first term is $\mathrm{O}\{(T^{2\rho-1}\log\log T)^{1/2}\}$ *a.s.* according to Lai and Wei while the second term is $\mathrm{O}(T^{\rho+\delta-1})$ by Lemma 3.2.

Suppose $\max|\mathrm{eigen}\,(\mathbf{B})| > 1$. By the triangle inequality and (2.3)

$$\|\mathbf{X}_T\| = \mathrm{O}\left(\left\|\tilde{W}_T\right\|\right) + \mathrm{O}\left(\|D_T\|\right) + \mathrm{O}\left(\|V_T\|\right) + \mathrm{O}\left(\|U_T\|\right).$$

The first term is $\mathrm{O}(T^{\rho-1}\max|\mathrm{eigen}\,(\mathbf{B})|^T)$ according to Lai and Wei (1985, Theorem 1,*iii*) while the other terms are of smaller order. ■

In the above proof of Theorem 4.1 the following generalised version of the Marcinkiewicz-Zygmund Theorem was implicitely used.

Theorem 4.2 (*Lai and Wei, 1983a, Corollaries 3,4*)
Suppose Assumptions 1.4, 1.5 are satisfied. Then for any sequence of matrices A_t

$$\sum_{t=1}^{T} \|A_t\| < \infty \qquad \Leftrightarrow \qquad \sum_{t=1}^{T} A_t\varepsilon_t \text{ converges a.s.}$$

If this holds, and $a'A_t \neq 0$ *for any vector* a *and infinitely many* t *then* $\mathsf{P}(\sum_{t=1}^{T} A_t\varepsilon_t = Y) = 0$ *for any variable* Y *that is* $\mathcal{F}_t$ *measurable for some* t.

This result yields a more precise statement about the order of magnitude of the explosive component.

Corollary 4.3 *Suppose Assumption 1.4, 1.5, 1.6 are satisfied. Then*

(*i*) $\mathbf{W}^{-T}W_T$ *converges a.s. to* $W = W_0 + \sum_{t=1}^{\infty}\mathbf{W}^{-t}e_{W,t}$ *satisfying* $\mathsf{P}(a'W = 0) = 0$ *for any* a.

(*ii*) $\sum_{t=1}^{T}||\mathbf{W}^{-T}W_t|| \to \sum_{t=1}^{\infty}\|\mathbf{W}^{-t}W\|$ *a.s.*

Proof of Corollary 4.3. (*i*) The decomposition (2.3) and Lemma 3.2 show that $W_T = \tilde{W}_T + \mathrm{o}(\mathbf{W}^T)$. The result then follows from Theorem 4.2, see also Lai and Wei (1985, Lemma 2).

(*ii*) Rewrite $\sum_{t=1}^{T}||\mathbf{W}^{-T}W_t|| = \sum_{t=1}^{T}||\mathbf{W}^{-(T-t)}\mathbf{W}^{-t}W_t||$ and use (*i*) and that $||\mathbf{W}^{-T}||$ is exponentially decreasing. ■

5 Correlation between stationary and deterministic component

One major difference between the results presented here and the work of Lai and Wei (1985) is that deterministic terms are included in the model. Before turning to the question of how big the denominator matrix can be in Section 6 it is convenient to consider the asymptotic order of magnitude of correlations between the zero mean process with roots smaller than one, $\tilde{U}_t$, and the deterministic component, D_t.

As a first step towards discussing the sample correlation of $\tilde{U}_t$ and D_t a result of Lai and Wei concerning the matrix $T^{-1}\sum_{t=1}^{T}\tilde{U}_t^{\otimes 2}$ is stated. The result gives conditions for relative compactness of a sequence of such matrices. Recalling that the relative compactness of a sequence is the property that the limit points fall in a compact set, this enables a discussion of the order of magnitude of the sequence under weak assumptions. In particular, a condition is given ensuring that the limit points are bounded away from zero.

Theorem 5.1 (*Lai and Wei, 1985, Theorem 2, Lemma 3, Example 3*).
Suppose Assumption 1.4 is satisfied. Then it holds with probability one that the matrix sequences

$$\left(\frac{1}{T}\sum_{t=1}^{T} e_{U,t}^{\otimes 2} : T \geq 1\right), \quad \left\{\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left(e_{U,t}^{\otimes 2}\middle|\mathcal{F}_{t-1}\right) : T \geq 1\right\}, \quad \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{U}_t^{\otimes 2} : T \geq 1\right)$$

are relatively compact and it holds, also with probability one, that

$$\begin{aligned} & \mathbf{E} \text{ is a limit point of } \left(\frac{1}{T}\sum_{t=1}^{T} e_{U,t}^{\otimes 2} : T \geq 1\right) \\ \Leftrightarrow\ & \mathbf{E} \text{ is a limit point of } \left\{\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left(e_{U,t}^{\otimes 2}\middle|\mathcal{F}_{t-1}\right) : T \geq 1\right\} \\ \Rightarrow\ & \sum_{t=0}^{\infty}\mathbf{U}^t\mathbf{E}\left(\mathbf{U}^t\right)' \text{ is a limit point of } \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{U}_t^{\otimes 2} : T \geq 1\right), \end{aligned}$$

If in addition Assumption 1.5 *is satisfied the above limit points are positive definite.*

The result for the sample correlation of $\tilde{U}_t$ and D_t can now be stated and proved.

Theorem 5.2 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then, for all* $\eta > 0$

$$\left(\sum_{t=1}^{T}\tilde{U}_t^{\otimes 2}\right)^{-1/2}\left(\sum_{t=1}^{T}\tilde{U}_t D_t'\right)\left(\sum_{t=1}^{T}D_t^{\otimes 2}\right)^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{\eta-1/2}\right).$$

Proof of Theorem 5.2. The norm of the matrix of interest is bounded by

$$\left\| \sum_{t=1}^{T} \tilde{U}_t^{\otimes 2} \right\|^{-1/2} \sum_{t=1}^{T} \left\| \tilde{U}_t \right\| \left\{ \max_{t \leq T} D_t' \left(\sum_{s=1}^{T} D_s^{\otimes 2} \right)^{-1} D_t \right\}^{1/2}.$$

The first term is $\mathrm{O}(T^{-1/2})$ by Theorem 5.1. In the second term, the process $Y_t = \sum_{s=1}^{t} \tilde{U}_s$ is of order $\mathrm{O}\{(T \log\log T)^{1/2}\}$ according to Theorem 4.1 because it satisfies an autoregression with $\dim U_t$ roots at one and $\dim U_t$ roots equal to those of $\mathbf{U}$. The third term is of order $\mathrm{O}(T^{-1/2})$ according to Theorem 3.3. ■

One immediate consequence of these result is the following.

Example 5.3 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then the Theorems 5.1, 5.2 and equation* (2.2) *imply that the sequence of matrices*

$$\frac{1}{T} \sum_{t=1}^{T} (U_t | D_t)^{\otimes 2} = \frac{1}{T} \sum_{t=1}^{T} \left(\tilde{U}_t | D_t \right)^{\otimes 2} \overset{a.s.}{=} \frac{1}{T} \sum_{t=1}^{T} \tilde{U}_t^{\otimes 2} + \mathrm{o}\left(T^{\eta - 1} \right)$$

is relatively compact with positive definite limit points. Moreover, this series will be convergent almost surely if $T^{-1} \sum_{t=1}^{T} \varepsilon_t^{\otimes 2}$ *is convergent. In particular, if* $U_t = \mu D_t + \varepsilon_t$ *then the sum of squared residuals from a least squares regression is asymptotically equivalent to the sum of squared innovations.*

6 The largest eigenvalue of the denominator matrix

The order of magnitude of the largest eigenvalue of the denominator matrix is described in the following generalisation of Lai and Wei (1985, Corollary 1).

Theorem 6.1 *Suppose Assumptions 1.4, 1.6 are satisfied. Then*

$$\lambda_{\max} \left\{ \sum_{t=1}^{T} \begin{pmatrix} \mathbf{X}_t \\ D_t \end{pmatrix}^{\otimes 2} \right\} \overset{a.s.}{=} \begin{cases} \mathrm{O}(T) + \mathrm{O}(T^{2\delta - 1}) & \text{if } \max |\mathrm{eigen}(\mathbf{B})| < 1, \\ \mathrm{O}(T^{2\rho} \log\log T) + \mathrm{O}(T^{2\delta + 2\rho - 1}) & \text{if } \max |\mathrm{eigen}(\mathbf{B})| = 1, \\ \mathrm{O}(T^{2\rho - 2} \max |\mathrm{eigen}(\mathbf{B})|^{2T}) & \text{if } \max |\mathrm{eigen}(\mathbf{B})| > 1. \end{cases}$$

Proof of Theorem 6.1. If $\max|\mathrm{eigen}(\mathbf{B})| < 1$ then $\mathbf{X}_t = \tilde{U}_t + \mu_u D_t$ by (2.2). By Theorem 5.2 the sample correlation of $\tilde{U}_t$ and D_t vanishes asymptotically. The result then follows since $\lambda_{\max}(\sum_{t=1}^{T} \tilde{U}_t^{\otimes 2}) = \mathrm{O}(T)$ *a.s.* by Lai and Wei (1985, Corollary 1) and since $\lambda_{\max}(\sum_{t=1}^{T} D_t^{\otimes 2}) = \mathrm{O}(T^{2\delta - 1})$ by Theorem 3.3.

If $\max|\mathrm{eigen}(\mathbf{B})| \geq 1$ the result follows directly from Theorem 4.1. ■

For the explosive part of the process the following generalisation of Lai and Wei (1985, Corollary 2) can be established.

Corollary 6.2 *Suppose Assumption 1.4, 1.5, 1.6 are satisfied. Recall the definition of* $W = \lim_{T\to\infty} \mathbf{W}^{-T} W_T$ *in Corollary 4.3. Then*

$$\mathbf{W}^{-T} \sum_{t=1}^{T} W_t^{\otimes 2} (\mathbf{W}^{-T})' \overset{a.s.}{\to} \mathbf{F}_W = \sum_{t=1}^{\infty} (\mathbf{W}^{-t} W)^{\otimes 2},$$

where $\mathbf{F}_W$ *is positive definite a.s., hence*

$$\begin{aligned} \lim T^{-1} \log \lambda_{\min} \left(\sum_{t=1}^{T} W_t^{\otimes 2} \right) &\overset{a.s.}{=} 2 \log \min \left| \mathrm{eigen}\left(\mathbf{W}\right) \right|, \\ \lim T^{-1} \log \lambda_{\max} \left(\sum_{t=1}^{T} W_t^{\otimes 2} \right) &\overset{a.s.}{=} 2 \log \max \left| \mathrm{eigen}\left(\mathbf{W}\right) \right|. \end{aligned}$$

Proof of Corollary 6.2. The result is a consequence of Corollary 4.3,*i*. The details are given by Lai and Wei (1983b, Theorem 2). ■

While Theorem 6.1 gives a bound for the sum of squares of the process the following result gives a bound for sum of higher order powers of the stationary component. The statement is less precise than that of Theorem 6.1 but the result is occasionally useful and will be used in the proof of the Central Limit Theorem 10.2

Theorem 6.3 *Suppose Assumption* (1.4) *is satisfied. Then, for all* $\eta > 0$ *and* $\zeta < \gamma$

$$\frac{1}{T^{1+\eta}} \sum_{t=1}^{T} \left\| \tilde{U}_t \right\|^{2+\zeta} \overset{a.s.}{\to} 0.$$

Proof of Theorem 6.3. For convenience define $e_{U,0} = \tilde{U}_0$. Using Hölder's inequality it follows that

$$\left\| \tilde{U}_t \right\|^{2+\zeta} = \left\| \sum_{j=0}^{t} \mathbf{U}^{t-j} e_{U,j} \right\|^{2+\zeta} \leq \left\{ \sum_{j=0}^{t} \left\| \mathbf{U}^{(t-j)/2} \right\|^{(2+\zeta)/(1+\zeta)} \right\}^{1+\zeta} \sum_{j=0}^{t} \left\| \mathbf{U}^{(t-j)/2} e_{U,j} \right\|^{2+\zeta}.$$

Summation over t then gives the following bound

$$\begin{aligned} \sum_{t=1}^{T} \left\| \tilde{U}_t \right\|^{2+\zeta} &\leq \left\{ \sum_{j=0}^{\infty} \left\| \mathbf{U}^{j/2} \right\|^{(2+\zeta)/(1+\zeta)} \right\}^{1+\zeta} \sum_{t=1}^{T} \sum_{j=0}^{t} \left\| \mathbf{U}^{(t-j)/2} e_{U,j} \right\|^{2+\zeta} \\ &= \left\{ \sum_{j=0}^{\infty} \left\| \mathbf{U}^{j/2} \right\|^{(2+\zeta)/(1+\zeta)} \right\}^{1+\zeta} \sum_{j=0}^{T} \sum_{t=\max(0,1-j)}^{T-j} \left\| \mathbf{U}^{t} \right\|^{1+\zeta/2} \left\| e_{U,j} \right\|^{2+\zeta} \\ &\leq \left\{ \sum_{j=0}^{\infty} \left\| \mathbf{U}^{j/2} \right\|^{(2+\zeta)/(1+\zeta)} \right\}^{1+\zeta} \sum_{t=0}^{\infty} \left\| \mathbf{U}^{t} \right\|^{1+\zeta/2} \sum_{j=0}^{T} \left\| e_{U,j} \right\|^{2+\zeta}. \end{aligned}$$

The first two sums are finite because the eigenvalues of $\mathbf{U}$ are smaller than one in absolute value. As for the third term decompose it as

$$\sum_{j=0}^{T}\|e_{U,j}\|^{2+\zeta}=\sum_{j=0}^{T}\left\{\|e_{U,j}\|^{2+\zeta}-\mathsf{E}\left(\|e_{U,j}\|^{2+\zeta}\Big|\mathcal{F}_{t-1}\right)\right\}+\sum_{j=0}^{T}\mathsf{E}\left(\|e_{U,j}\|^{2+\zeta}\Big|\mathcal{F}_{t-1}\right).$$

The latter term is of order $\mathrm{O}(T)=\mathrm{o}(T^{1+\eta})$ by Assumption (1.4). The first term is a martingale. Normalised by $T^{1+\eta}$ it converges to zero $a.s.$ on the set

$$\left[\sum_{j=0}^{T}j^{-(1+\eta)}\mathsf{E}\left\{\left|\|e_{U,j}\|^{2+\zeta}-\mathsf{E}\left(\|e_{U,j}\|^{2+\zeta}\Big|\mathcal{F}_{t-1}\right)\right|\Big|\mathcal{F}_{t-1}\right\}<\infty\right],$$

see Hall and Heyde (1980, Theorem 2.18). Minkowski's inequality shows that this sum is finite if the sum $\sum_{j=0}^{T}j^{-(1+\eta)}\mathsf{E}(\|e_{U,j}\|^{2+\gamma}|\mathcal{F}_{t-1})$ is finite. Assumption (1.4) ensures this is the case. ■

7 Limiting results for the non-explosive components

Two results are given for the non-explosive case, that is when $\max|\mathrm{eigen}(\mathbf{B})|\leq 1$. The first concerns the smallest eigenvalue and thus a lower bound for the matrix $\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}$ which is related to the denominator matrix (1.2). The second result is of a more technical nature and will be used in the subsequent Section. It gives the order of magnitude of the scalar $\max_{t\leq T}S_t'\mathbf{S}_T^{-1}S_t$ where

$$S_t=\begin{pmatrix}\mathbf{X}_t\\ D_t\end{pmatrix},\qquad \mathbf{S}_T=\sum_{s=1}^{T}\begin{pmatrix}\mathbf{X}_t\\ D_t\end{pmatrix}^{\otimes 2}. \tag{7.1}$$

The first result gives a lower bound for $\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}$ and is related to Theorem 3 of Lai and Wei (1985).

Theorem 7.1 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied and* $\max|\mathrm{eigen}(\mathbf{B})|\leq 1$*. Then*

$$\liminf_{T\to\infty}\lambda_{\min}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t^{\otimes 2}\right)\geq\liminf_{T\to\infty}\lambda_{\min}\left\{\frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}\right\}\overset{a.s.}{>}0.$$

In other words, the rate of convergence of the denominator matrix is at least T which is the rate of convergence for sums of squares of stationary processes. This result is useful in various ways as will be demonstrated in Section 8. While it also holds that $\liminf_{T\to\infty}\lambda_{\min}(T^{-1}\sum_{t=1}^{T}D_t^{\otimes 2})>0$, see Theorem 3.3, it is interesting to note that these results are not sufficient to ensure that $\liminf_{T\to\infty}\lambda_{\min}(T^{-1}\sum_{t=1}^{T}S_t^{\otimes 2})>0$

a.s., and such a result has not been proved, although a version holding in probability is given in §11.

The proof of Theorem 7.1 is based on the following generalisation of Lemma 6,*iii* of Lai and Wei (1983b). A slightly stronger version of this result is presented towards the end of this section.

Lemma 7.2 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied and* $\max|\text{eigen}(\mathbf{B})| \leq 1$. *Then, for* $j \geq 1$,

$$\left(\sum_{t=1}^{T} \varepsilon_{t+j} S_t'\right) \mathbf{S}_T^{-1} \left(\sum_{t=1}^{T} S_t \varepsilon_{t+j}'\right) \overset{a.s.}{=} \mathrm{O}(\log T).$$

Proof of Lemma 7.2. Lai and Wei (1982a, Lemma 1,*iii*) show that the object of interest is $\mathrm{O}\{\log \lambda_{\max}(\mathbf{S}_T)\}$ on the set $\{\lambda_{\max}(\mathbf{S}_T) \to \infty\}$. While Theorem 6.1 show that $\log \lambda_{\max}(\mathbf{S}_T) = \mathrm{O}(\log T)$ *a.s.* Theorem 3.3 shows that $\lambda_{\max}(\mathbf{S}_T) \geq \lambda_{\max}(\sum_{t=1}^{T} D_t^{\otimes 2}) \to \infty$. ■

Proof of Theorem 7.1. The idea of the proof is taken from Lai and Wei (1982b). Adding an additional regressor reduces the sum of squared residuals so

$$\sum_{t=1}^{T} (\mathbf{X}_t | D_t)^{\otimes 2} \geq \sum_{t=1}^{T} (\mathbf{X}_t | D_t, \mathbf{X}_{t-\dim \mathbf{X}})^{\otimes 2}.$$

The model equation (2.1) shows

$$\begin{aligned} \mathbf{X}_t &= \sum_{j=0}^{\dim \mathbf{X}-\mathbf{1}} \mathbf{B}^j (\mathbf{e}_{t-j} + \boldsymbol{\mu} D_{t-1-j}) + \mathbf{B}^{\dim \mathbf{X}} \mathbf{X}_{t-\dim \mathbf{X}} \\ &= \sum_{j=0}^{\dim \mathbf{X}-\mathbf{1}} \mathbf{B}^j \mathbf{e}_{t-j} + \sum_{j=0}^{\dim \mathbf{X}-\mathbf{1}} \mathbf{B}^j \boldsymbol{\mu} \mathbf{D}^{-1-j} D_t + \mathbf{B}^{\dim \mathbf{X}} \mathbf{X}_{t-\dim \mathbf{X}}, \end{aligned}$$

and consequently

$$\sum_{t=1}^{T} (\mathbf{X}_t | D_t, \mathbf{X}_{t-\dim \mathbf{X}})^{\otimes 2} = \sum_{t=1}^{T} \left(\sum_{j=0}^{\dim \mathbf{X}-\mathbf{1}} \mathbf{B}^j \mathbf{e}_{t-j} \middle| D_t, \mathbf{X}_{t-\dim \mathbf{X}} \right)^{\otimes 2}.$$

The Lemma 7.2 implies that the regressor is negligible so

$$\frac{1}{T} \sum_{t=1}^{T} (\mathbf{X}_t | D_t, \mathbf{X}_{t-\dim \mathbf{X}})^{\otimes 2} \overset{a.s.}{=} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=0}^{\dim \mathbf{X}-\mathbf{1}} \mathbf{B}^j \mathbf{e}_{t-j} \right)^{\otimes 2} \{1 + \mathrm{o}(T)\},$$

which has positive definite limit points, see Lai and Wei (1985, equation 3.19). ■

The second and more technical result is an extended version of Theorem 3.3. Recall the definition of the stacked process S_t and the sum of squares $\mathbf{S}_T$ in (7.1).

Lemma 7.3 *Suppose Assumptions 1.6, 1.4, 1.5 are satisfied and* $\max|\text{eigen}(\mathbf{B})| \leq 1$. *Then, for all* $\xi < \gamma/(2+\gamma)$,

$$\max_{t\leq T} S_t'\mathbf{S}_T^{-1}S_t = \mathrm{o}\left(T^{-\xi/4}\right).$$

The Lemma 7.3 is proved by Lai and Wei (1983b, Theorem 4) for the special case where X is a univariate autoregression without deterministic components and the order of magnitude is shown to be $\mathrm{o}(1)$. The outline of their proof can also be used to prove this more general and stronger result.

The first step in the argument is an algebraic result strengthening Lemma 3 of Lai and Wei (1983b).

Lemma 7.4 *Let* $\{a_t\}$ *be a sequence of non-negative numbers such that*

$$\sum_{t=1}^{T} a_t = \mathrm{o}\left(T^{\delta}\right) \tag{7.2}$$

for all $\delta > 0$ *and there exists* $C > 0$ *and* $\kappa > 0$ *such that*

$$a_{t+1} \leq a_t + CT^{-\kappa} \qquad \text{for all large } t. \tag{7.3}$$

Then

$$a_t = \mathrm{o}\left(T^{\delta-\rho}\right), \qquad \text{for all } \rho = \min(1, \kappa/2).$$

Proof of Lemma 7.4. Condition (7.3) implies that for every $0 < \rho < 1$

$$\min_{T>t\geq T-T^{\rho}} a_t \geq a_T - 2CT^{\rho-\kappa} \qquad \text{for all large } T.$$

In particular, choosing $0 < \rho < \min(1, \kappa/2)$, it is seen that

$$\sum_{t=1}^{T} a_t \geq \sum_{t=T-T^{\rho}1}^{T} a_t \geq T^{\rho}\left(a_T - 2CT^{\rho-\kappa}\right) \geq T^{\rho}a_T - 2C \qquad \text{for all large } T.$$

Combining this with (7.2) it follows that

$$a_T \leq T^{-\rho}\left(\sum_{t=1}^{T} a_t + 2C\right) = \mathrm{o}\left(T^{\delta-\rho}\right).$$

■

The second step is to generalise and strengthen Lemma $6i, ii$ of Lai and Wei (1983b).

Lemma 7.5 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied and* $\max|\text{eigen}(\mathbf{B})| \leq 1$. *Let* $T_0 = \inf\{t\colon \text{rank}\mathbf{S}_t = \dim S_t\}$ *with the convention* $\inf\emptyset = \infty$. *Then*
(i) $T_0 \overset{a.s.}{<} \infty$,
(ii) $S_t'\mathbf{S}_t^{-1}S_t \leq 1$ *for* $t \geq T_0$,
(iii) $\sum_{t=T_0+1}^{T} S_t'\mathbf{S}_t^{-1}S_t \overset{a.s.}{=} \mathrm{O}(\log T)$.

Proof of Lemma 7.5. (i) decompose $\mathbf{S}_T$ as

$$\mathbf{S}_T = \begin{pmatrix} I_{\dim\mathbf{X}} & A \\ 0 & I_{\dim\mathbf{D}} \end{pmatrix} \begin{Bmatrix} \sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2} & 0 \\ 0 & \sum_{t=1}^{T} D_t^{\otimes 2} \end{Bmatrix} \begin{pmatrix} I_{\dim\mathbf{X}} & 0 \\ A' & I_{\dim\mathbf{D}} \end{pmatrix}.$$

where $A = \sum_{t=1}^{T}\mathbf{X}_t D_t'(\sum_{t=1}^{T} D_t^{\otimes 2})^{-1}$. It follows that $\mathbf{S}_t$ is regular if the second matrix is regular, but this follows by Theorems 3.3, 7.1.

(ii) is an immediate consequence of Lai and Wei, 1982, Lemma 2,i.

(iii) note that $\mathbf{S}_{t-1} = \mathbf{S}_t - S_tS_t'$ and therefore

$$S_t'\mathbf{S}_t^{-1}S_t = 1 - \det\mathbf{S}_{t-1}/\det\mathbf{S}_t \leq -\log\left(\det\mathbf{S}_{t-1}/\det\mathbf{S}_t\right). \tag{7.4}$$

Using that $-\log(a/b) = \log b - \log a$ this implies

$$\sum_{t=T_0+1}^{T} S_t'\mathbf{S}_t^{-1}S_t \leq \log\det\mathbf{S}_T - \log\det\mathbf{S}_{T_0}.$$

Since $\det\mathbf{S}_{T_0} > 0$ by construction, it suffices to argue that $\det\mathbf{S}_T < \lambda_{\max}(\mathbf{S}_T) = \mathrm{O}(T^{\varsigma})$ $a.s.$ for some $\zeta > 0$ and that $\liminf_{T\to\infty}\det\mathbf{S}_T > 1$ $a.s.$ The first property follows from Theorem 7.1 whereas the second follows by noting that $\det\mathbf{S}_T = \det\{\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}\}\det(\sum_{t=1}^{T} D_t^{\otimes 2})$ and then using Theorems 3.3, 7.1. ∎

The third step is to generalise Lemma 7 of Lai and Wei (1983b). The proof is omitted since it is identical to their proof except for a reference to their Lemma 6 which is replaced by the above Lemma 7.5.

Lemma 7.6 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied and* $\max|\text{eigen}(\mathbf{B})| \leq 1$. *Then, for all* $\xi < \gamma/(2+\gamma)$,

$$S_{T+1}\mathbf{S}_{T+1}^{-1}S_{T+1} \leq S_T\mathbf{S}_T^{-1}S_T + \mathrm{o}\left(T^{-\xi/2}\right).$$

The Lemma 7.3 can now be proved.

Proof of Lemma 7.3. Let T_0 be defined as in Lemma 7.5 and note

$$\max_{1\leq t\leq T} S_t'\mathbf{S}_T^{-1}S_t \leq \max_{1\leq t<T_0} S_t'\mathbf{S}_T^{-1}S_t + \max_{T_0\leq t\leq T} S_t'\mathbf{S}_T^{-1}S_t.$$

The first term is of order $\mathrm{O}(T^{-1})$ since $\lambda_{\max}(\mathbf{S}_T^{-1}) = \mathrm{O}(T^{-1})$ by Theorem 7.1 while $\max_{t<T_1}\|S_t\|$ is finite. The second term can in the first instance be bounded by

$$\max_{T_0\le t\le T} S_t'\mathbf{S}_T^{-1}S_t \le \max_{T_0\le t\le T} S_t'\mathbf{S}_t^{-1}S_t,$$

since the increment $\mathbf{S}_t^{-1} - \mathbf{S}_{t+1}^{-1}$ is positive semidefinite, see Lai and Wei (1982a, Equation 1.4b). The Lemma 7.4 now shows that $S_t'\mathbf{S}_t^{-1}S_t \overset{a.s.}{=} \mathrm{o}(T^{-\xi/4})$ since the conditions (7.2), (7.3) are satisfied by the Lemmas 7.5,*iii* and 7.6 respectively. ■

As a final result in this Section the Lemma 7.2 is strengthened slightly

Theorem 7.7 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied,* $\dim \mathbf{D} = 0$, *and* $\max|\mathrm{eigen}(\mathbf{B})| \le 1$. *Then, for* $j \ge 1$,

$$\left(\sum_{t=1}^{T}\varepsilon_{t+j}S_t'\right)\mathbf{S}_T^{-1}\left(\sum_{t=1}^{T}S_t\varepsilon_{t+j}'\right) \overset{a.s.}{=} \mathrm{O}(\log\log T).$$

For the proof of Theorem 7.7 a result by Wei (1985) is used.

Lemma 7.8 (*Wei 1985, Lemma 2*)
Suppose Assumption 1.4 is satisfied. Let (x_t) *be a sequence of random variables adapted to* $(\mathcal{F}_t)$ *and let* $s_T^2 = \sum_{t=1}^T x_t^2$ *and assume* $x_T^2 = \mathrm{o}(s_T^{2-\eta})$ *a.s. for some* $\eta > 0$. *Then*

$$\sum_{t=1}^{T} x_{t-1}\varepsilon_t \overset{a.s.}{=} \mathrm{O}\left\{s_T(\log\log s_T)^{1/2}\right\}.$$

Proof of Theorem 7.7. Using a similarity transformation $\mathbf{S}_T$ can be diagonalised and therefore without loss of generality it can be assumed that $u_t = S_t$ is univariate. By Lemma 7.3 it holds that $s_T^{-2}x_T^2 = \mathrm{o}(T^{-\xi/4})$. Theorem 7.1 then shows that $s_T^{-2}x_T^2 = \mathrm{o}(s_T^{-\eta})$. The desired result then follows from Lemma 7.8. ■

8 Sample correlations and lower bound for denominator matrix

It has already been established in Section 5 that the sample correlation of $\tilde{U}_t$ and D_t vanishes asymptotically. In the following the remaining sample correlations of pairs of the processes $\tilde{U}_t, V_t, W_t, D_t$ are studied in a series of theorems. Subsequently these are used to give a lower bound for the denominator matrix $\sum_{t=1}^T(\mathbf{X}_t|D_t)^{\otimes 2}$ without the condition $|\mathrm{eigen}(\mathbf{B})| \le 1$, and it is shown how that result can be applied to cointegration analysis.

Theorem 8.1 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then*

$$\left(\sum_{t=1}^{T} W_t^{\otimes 2}\right)^{-1/2} \left(\sum_{t=1}^{T} W_t D_t'\right) \left(\sum_{t=1}^{T} D_t^{\otimes 2}\right)^{-1/2} \overset{a.s.}{=} \mathrm{O}\left(T^{-1/2}\right).$$

Proof of Theorem 8.1. The norm of the matrix of interest is bounded by

$$\left\|\mathbf{W}^{-T} \sum_{t=1}^{T} W_t^{\otimes 2} \left(\mathbf{W}^{-T}\right)'\right\|^{-1/2} \left\|\mathbf{W}^{-T} \sum_{t=1}^{T} W_t\right\| \left\{\max_{t \leq T} D_t' \left(\sum_{s=1}^{T} D_s^{\otimes 2}\right)^{-1} D_t\right\}^{1/2}.$$

The first and the second terms are convergent according to Corollaries 4.3, 6.2 whereas the third term is of order $\mathrm{O}(T^{-1/2})$ according to Theorem 3.3. ■

Theorem 8.2 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$,

$$\left(\sum_{t=1}^{T} W_t^{\otimes 2}\right)^{-1/2} \left(\sum_{t=1}^{T} W_t \tilde{U}_t'\right) \left(\sum_{t=1}^{T} \tilde{U}_t^{\otimes 2}\right)^{-1/2} \overset{a.s.}{=} \mathrm{O}\left(T^{-\xi/2}\right).$$

Proof of Theorem 8.2. The norm of the matrix of interest is bounded by

$$\left\|\mathbf{W}^{-T} \sum_{t=1}^{T} W_t^{\otimes 2} \left(\mathbf{W}^{-T}\right)'\right\|^{-1/2} \sum_{t=1}^{T} \left\|\mathbf{W}^{-T} W_t\right\| \left(\max_{t \leq T} \left\|\tilde{U}_t\right\|\right) \left(\sum_{t=1}^{T} \tilde{U}_t^{\otimes 2}\right)^{-1/2}.$$

The first two terms are convergent according to the Corollaries 4.3, 6.2. Theorem 4.1 shows that $T^{-(1-\xi)/2}\tilde{U}_t \to 0$ *a.s.* and therefore the third term is $\mathrm{o}\{T^{-(1-\xi)/2}\}$. Finally, the fourth term is $\mathrm{O}(T^{-1/2})$ by Example 5.3. ■

Theorem 8.3 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$,

$$\left(\sum_{t=1}^{T} \tilde{U}_t^{\otimes 2}\right)^{-1/2} \left\{\sum_{t=1}^{T} \tilde{U}_t \left(V_t|D_t\right)'\right\} \left\{\sum_{t=1}^{T} \left(V_t|D_t\right)^{\otimes 2}\right\}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/8}\right),$$

$$\left(\sum_{t=1}^{T} W_t^{\otimes 2}\right)^{-1/2} \left\{\sum_{t=1}^{T} W_t \left(V_t|D_t\right)'\right\} \left\{\sum_{t=1}^{T} \left(V_t|D_t\right)^{\otimes 2}\right\}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/8}\right).$$

Proof of Theorem 8.3. The norm of the first expression is bounded by

$$\left[\max_{t\leq T}(V_t|D_t)'\left\{\sum_{t=1}^{T}(V_t|D_t)^{\otimes 2}\right\}^{-1}(V_t|D_t)\right]^{1/2}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{U}_t\right\|\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{U}_t^{\otimes 2}\right)^{-1/2}.$$

The first term is of order $\mathrm{o}(T^{-\xi/8})$ according to Lemma 7.3. In the second term the process $Y_t=\sum_{s=1}^{t}\tilde{U}_s$ is of order $\mathrm{O}\{(T\log\log T)^{1/2}\}$ because it satisfies a second order autoregression with $\dim U_t$ roots at one and $\dim U_t$ roots equal to those of $\mathbf{U}$. The third term is finite *a.s.* according to Example 5.3.

Correspondingly the second expression of interest is bounded by

$$\left[\max_{t\leq T}(V_t|D_t)'\left\{\sum_{t=1}^{T}(V_t|D_t)^{\otimes 2}\right\}^{-1}(V_t|D_t)\right]^{1/2}\left\|\sum_{t=1}^{T}\mathbf{W}^{-T}W_t\right\|\left\{\mathbf{W}^{-T}\sum_{t=1}^{T}W_t^{\otimes 2}(\mathbf{W}')^{-T}\right\}^{-1/2}.$$

The first term is of order $\mathrm{o}(T^{-\xi/8})$ according to Lemma 7.3. The two terms involving W are convergent and finite as found in Corollary 6.2. ■

By combining these results nearly all pairs of the processes $\tilde{U}_t, V_t, W_t, D_t$ have been considered. The only exception is the sample correlation of V_t and D_t which will not be negligible when these processes have common characteristic roots. The Table 1 summarises the results given in Theorem 5.2, 8.1, 8.2, 8.3.

	$\tilde{U}_t$	$(V_t\|D_t)$	W_t	D_t
$\tilde{U}_t$	1	$\mathrm{o}(T^{-\xi/8})$	$\mathrm{O}(T^{-\xi/2})$	$\mathrm{o}(T^{\eta-1/2})$
$(V_t\|D_t)$	$\mathrm{o}(T^{-\xi/8})$	1	$\mathrm{o}(T^{-\xi/8})$	0
W_t	$\mathrm{O}(T^{-\xi/2})$	$\mathrm{o}(T^{-\xi/8})$	1	$\mathrm{O}(T^{-1/2})$
D_t	$\mathrm{o}(T^{\eta-1/2})$	0	$\mathrm{O}(T^{-1/2})$	1

Table 1: Order of magnitude of pairwise sample correlations, where $\eta>0$ and $\xi<\gamma/(2+\gamma)$.

As an alternative to Table 1 the Table 2 presents the sample correlations of the residuals $(\tilde{U}_t|D_t)$, $(V_t|D_t)$, $(W_t|D_t)$. To derive these result let $\mathrm{C}(x,y)$ and $\mathrm{C}(x,y|z)$ denote the sample correlations of processes x_t and y_t and of the residuals $(x_t|z_t)$ and $(y_t|z_t)$, respectively, and use the formulas

$$\begin{aligned}\mathrm{C}(x_t,y_t|z_t) &= \left\{I_{\dim x}+\mathrm{C}(x,z)^{\otimes 2}\right\}^{-1/2}\mathrm{C}\{x,(y|z)\}\\ &= \left\{I_{\dim x}+\mathrm{C}(x,z)^{\otimes 2}\right\}^{-1/2}\{\mathrm{C}(x,y)-\mathrm{C}(x,z)\,\mathrm{C}(z,y)\}\left\{I_{\dim y}+\mathrm{C}(y,z)^{\otimes 2}\right\}^{-1/2}.\end{aligned}$$

	$(\tilde{U}_t\vert D_t)$	$(V_t\vert D_t)$	$(W_t\vert D_t)$
$(\tilde{U}_t\vert D_t)$	1	$\mathrm{o}(T^{-\xi/8})$	$\mathrm{O}(T^{-\xi/2})$
$(V_t\vert D_t)$	$\mathrm{o}(T^{-\xi/8})$	1	$\mathrm{o}(T^{-\xi/8})$
$(W_t\vert D_t)$	$\mathrm{O}(T^{-\xi/2})$	$\mathrm{o}(T^{-\xi/8})$	1

Table 2: Order of magnitude of pairwise sample correlations, where $\xi < \gamma/(2+\gamma)$.

As an second alternative to Table 1 the Table 3 presents the sample correlations of the processes $\tilde{U}_t$, V_t, W_t. To derive these note that the formula for partitioned inversion implies that

$$\mathrm{C}\left\{x, \begin{pmatrix} y \\ z \end{pmatrix}\right\} = \mathrm{o}\left(T^{-a}\right) \quad \Leftrightarrow \quad \mathrm{C}\left\{x, (y|z)\right\} = \mathrm{o}\left(T^{-a}\right) \quad \text{and} \quad \mathrm{C}\left(x, z\right) = \mathrm{o}\left(T^{-a}\right).$$

	$\tilde{U}_t$	V_t	W_t
$\tilde{U}_t$	1	$o(T^{-\xi/8})$	$o(T^{-\xi/2})$
V_t	$o(T^{-\xi/8})$	1	$o(T^{-\xi/8})$
W_t	$o(T^{-\xi/2})$	$o(T^{-\xi/8})$	1

Table 3: Order of magnitude of pairwise sample correlations, where $\xi < \gamma/(2+\gamma)$.

A final manipulation of these sample correlation results is to consider concatenated processes using the formula

$$\mathrm{C}\left\{x, \begin{pmatrix} y \\ z \end{pmatrix}\right\} = \mathrm{o}\left(T^{-a}\right) \quad \Leftrightarrow \quad \mathrm{C}\left(x, y\right), \mathrm{C}\left(x, z\right) = \mathrm{o}\left(T^{-a}\right) \quad \text{and} \quad \mathrm{C}\left(y, z\right) = \mathrm{o}\left(1\right).$$

This gives the following example.

Example 8.4 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then the results in Table 2 show, with (Y_t, Z_t) defined as either of the pairs (U_t, V_t), (U_t, W_t), (V_t, W_t) and $\xi < \gamma/(2+\gamma)$,*

$$\left\{\sum_{t=1}^{T} (Y_t|D_t)^{\otimes 2}\right\}^{-1/2} \left\{\sum_{t=1}^{T} (Y_t|D_t)(Z_t|D_t)'\right\} \left\{\sum_{t=1}^{T} (Z_t|D_t)^{\otimes 2}\right\}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(1\right).$$

This result is used in the study of the asymptotic behaviour of procedures for determining the order of vector autoregressive models by Nielsen (2001*b*).

A further implication of the above results is a generalisation of Theorem 7.1 concerning a lower bound for the denominator matrix.

Corollary 8.5 *Suppose Assumption 1.4, 1.5, 1.6 are satisfied. Then*

$$\liminf_{T\to\infty}\lambda_{\min}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t^{\otimes 2}\right)\geq\liminf_{T\to\infty}\lambda_{\min}\left\{\frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}\right\}\overset{a.s.}{>}0.$$

Proof of Corollary 8.5. Let $Y_t=(U_t',V_t')'$. Then the results in Table 2 show

$$\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}\overset{a.s.}{=}\left\{\begin{array}{cc}\sum_{t=1}^{T}(Y_t|D_t)^{\otimes 2} & 0\\ 0 & \sum_{t=1}^{T}(W_t|D_t)^{\otimes 2}\end{array}\right\}\{1+\mathrm{o}(1)\}.$$

Thus normalise the matrix on the right hand side with T and apply Theorem 7.1 to the upper left block and Corollary 6.2 and Theorem 8.1 to the lower left block. ∎

For each of the components of the process $\mathbf{X}$ more precise convergence rates can be obtained in that the weak limit of $T^{-2}\sum_{t=1}^{T}\tilde{V}_t^{\otimes 2}$ is positive definite, see §11, and $\sum_{t=1}^{T}\tilde{W}_t^{\otimes 2}$ is exponentially growing, see Corollary 6.2. While these results are often difficult to exploit since the similarity transformation M and the appropriate normalisation matrix, N_T say, do not commute in general, the Corollary 8.5 can be used more easily as illustrated by the following example.

Example 8.6 *Consider the vector autoregression*

$$\Delta X_t=\Pi X_{t-1}+\sum_{j=1}^{k}\Gamma_j\Delta X_{t-j}+\varepsilon_t,\qquad t=1,\ldots,T,$$

where the innovations are independent identically $\mathsf{N}(0,\Omega)$ *distributed. The hypothesis of at most* r *cointegrating vectors can be formulated as* $H(r)$: $\operatorname{rank}\Pi\leq r$ *or equivalently* $\Pi=\alpha\beta'$ *for some* $(p\times r)$*-matrices* α,β. *Under the hypothesis the process* X_t *is non-stationary but the cointegrating relation* $\beta'X_t$ *can be given a stationary initial distribution, provided* (i) $\operatorname{rank}\Pi=r$, (ii) *the process has exactly* $p-r$ *characteristic roots at one, and* (iii) *the other characteristic roots are stationary. Consequently,*

$$S_{\beta\beta}=\frac{1}{T}\sum_{t=1}^{T}(\beta'X_{t-1}|\Delta X_{t-1},\ldots,\Delta X_{t-k})^{\otimes 2}$$

converges in probability to the corresponding long-run covariance matrix which is invertible. The convergence of $S_{\beta\beta}^{-1}$ *is used by Johansen* (1996) *in the asymptotic theory for the likelihood ratio test statistic for* $H(r)$ *against* $H(p)$, *a theory which assumes* (i)-(iii). *Now, it is an immediate consequence of Corollary 8.5 and the Lemma 8.7 below that*

$$\limsup_{T\to\infty}\lambda_{\max}\left(S_{\beta\beta}^{-1}\right)\overset{a.s.}{<}\infty,$$

as long as just (i) *is satisfied. This result is used by Nielsen* (2000) *to show that the assumptions* (ii), (iii) *are redundant for various aspects of the asymptotic analysis of cointegration tests.*

Lemma 8.7 *Consider a sequence of vectors, $(x_t', y_t')'$, $t = 1, \ldots, T$. Then*

$$\lambda_{\min}\left\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\right\} \geq \frac{1}{\dim(x)+\dim(y)}\lambda_{\min}\left\{\sum_{t=1}^{T}\begin{pmatrix} x_t \\ y_t \end{pmatrix}^{\otimes 2}\right\} \geq 0, \tag{8.1}$$

$$\lambda_{\min}\left(\sum_{t=1}^{T} y_t^{\otimes 2}\right) \geq \lambda_{\min}\left\{\sum_{t=1}^{T}\begin{pmatrix} x_t \\ y_t \end{pmatrix}^{\otimes 2}\right\} \geq 0. \tag{8.2}$$

Proof of Lemma 8.7. The inequality (8.1) holds trivially if the matrix $\sum_{t=1}^{T}\{(x_t', y_t')'\}^{\otimes 2}$ is singular. Therefore assume it is positive definite. The upper left block of its inverse is given by $\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\}^{-1}$ and therefore

$$\operatorname{tr}\left[\left\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\right\}^{-1}\right] \leq \operatorname{tr}\left[\left\{\sum_{t=1}^{T}\begin{pmatrix} x_t \\ y_t \end{pmatrix}^{\otimes 2}\right\}^{-1}\right].$$

Further, for a positive definite matrix A then $\lambda_{\min}(A) = \{\lambda_{\max}(A^{-1})\}^{-1}$ and $\lambda_{\max}(A) \leq \operatorname{tr}(A) \leq \dim(A)\lambda_{\max}(A)$ which leads to the inequality (8.1). The inequality (8.2) is a consequence of Poincaré's Separation Theorem, see Magnus and Neudecker (1999, Exercise 11.11.1). ■

9 *Least squares statistics*

The asymptotic order of the normalised least squares statistic (1.3) is stated and proved in the following. Subsequently the least squares statistic itself and the sum of squared residuals are considered.

Theorem 9.1 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then, for all $\xi < \gamma/(2+\gamma)$ and $j \geq 1$ it holds*

$$\left\{\sum_{t=1}^{T}\begin{pmatrix} \mathbf{X}_t \\ D_t \end{pmatrix}^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} \mathbf{X}_t \\ D_t \end{pmatrix}\varepsilon_{t+j}' \overset{a.s.}{=} \mathrm{o}\left\{T^{(1-\xi)/2}\right\}.$$

If $\max|\mathrm{eigen}(\mathbf{B})| \leq 1$ *then right hand side expression equals* $\mathrm{O}\{(\log T)^{1/2}\}$ *a.s.*

Pötscher (1989, Lemma A.1) states this result with $\xi = 0$ and $\dim D_t = 0$ while Nielsen (2001a) proves a univariate version holding in probability. A slightly stronger version of the result was stated as Theorem 7.7 for the case $|\mathrm{eigen}(\mathbf{B})| \leq 1$ and $\dim \mathbf{D} = 0$. The Theorem 9.1 is a special case of the following slightly more general result.

Lemma 9.2 *Suppose the innovations ε_t satisfy Assumptions 1.4, 1.5, 1.6 and that m_t is a martingale difference sequence adapted to $\mathcal{F}_t$ and satisfying Assumptions 1.4, 1.5. Then, for $j \geq 1$,*

$$\left\{\sum_{t=1}^{T}\begin{pmatrix}\mathbf{X}_t\\ D_t\end{pmatrix}^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}\mathbf{X}_t\\ D_t\end{pmatrix} m'_{t+j} \overset{a.s.}{=} \mathrm{O}\left(\max_{t\leq T}\|m_t\|\right).$$

Proof of Lemma 9.2. Let $S_t = (U'_t, V'_t, D'_t)'$. As a consequence of Theorems 8.1, 8.2 and 8.3, see also Table 1, the vector of interest equals

$$\left\{\sum_{t=1}^{T}\begin{pmatrix}W_t^{\otimes 2} & 0\\ 0 & S_t^{\otimes 2}\end{pmatrix}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}W_t\\ S_t\end{pmatrix} m'_{t+j}\left\{1+\mathrm{o}(1)\right\}.$$

Thus the explosive and non-explosive components can be considered separately.

For the explosive component note that

$$\left\|\left(\sum_{t=1}^{T}W_t^{\otimes 2}\right)^{-1/2}\sum_{t=1}^{T}W_t m'_{t+j}\right\| \leq \left\|\left(\sum_{t=1}^{T}W_t^{\otimes 2}\right)^{-1/2}\sum_{t=1}^{T}W_t\right\|\max_{t\leq T}\|m_t\|,$$

where the first term is convergent because of the Corollaries 4.3, 6.2 while the second term is of the desired order by Theorem 4.1.

For the second component the argument is essentially that of Lemma 7.2. ■

Rather than discussing the normalised least square estimator as above the literature has been more concerned with the consistency of a the least squares estimator for $A_1, \ldots, A_k$ or equivalently of $\mathbf{B}$. The issue of strong consistency was first discussed for a Gaussian first order autoregression by Rubin (1950) and later for a vector autoregression without deterministic terms by Lai and Wei (1985, Theorem 4). By combining Theorem 9.1 with Theorem 7.1 and Corollary 6.2 a generalisation to the model with deterministic terms is achieved and the rate of convergence can be addressed more precisely. This has previously been done by Duflo, Senoussi and Touati (1991, Theorem 1) in the case where the explosive roots have multiplicity one whereas their Theorem 2 falsely suggests that the least square estimator for $\mathbf{B}$ otherwise is inconsistent.

Theorem 9.3 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then, for all $\xi < \gamma/(2+\gamma)$ it holds*

$$\left(\hat{A}_1 - A_1, \ldots, \hat{A}_k - A_k\right) = \sum_{t=1}^{T}\varepsilon_t\left(\mathbf{X}_t|D_t\right)'\left\{\sum_{t=1}^{T}\left(\mathbf{X}_t|D_t\right)^{\otimes 2}\right\}^{-1} \overset{a.s.}{=} \mathrm{O}\left\{\left(T^{-1}\log T\right)^{1/2}\right\}.$$

Proof of Theorem 9.3. As in the proof of Lemma 9.2 let $S_t = (U_t', V_t', D_t')'$ and use the similarity transformation described in §2 so

$$\left(\hat{A}_1 - A_1, \ldots, \hat{A}_k - A_k\right)' = (M')^{-1} \left[\begin{array}{c} \left\{\sum_{t=1}^T (S_t|D_t)^{\otimes 2}\right\}^{-1} \sum_{t=1}^T (S_t|D_t)\, \varepsilon_t' \\ \left\{\sum_{t=1}^T (W_t|D_t)^{\otimes 2}\right\}^{-1} \sum_{t=1}^T (W_t|D_t)\, \varepsilon_t' \end{array} \right].$$

Use Theorem 9.1 together with Theorem 7.1 for the non-explosive part and together with Corollary 6.2 for the explosive part. ■

Another consequence of Theorem 9.1 is that variance of the innovations can be estimated consistently.

Example 9.4 *Suppose Assumptions 1.4, 1.5, 1.6 are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$ *it holds that the series*

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^T (X_t|\mathbf{X}_{t-1}, D_t)^{\otimes 2} = \frac{1}{T}\sum_{t=1}^T (\varepsilon_t|\mathbf{X}_{t-1}, D_t)^{\otimes 2} \overset{a.s.}{=} \frac{1}{T}\sum_{t=1}^T \varepsilon_t^{\otimes 2} + \mathrm{o}\left(T^{-\xi}\right),$$

is relatively compact with positive definite limit points. Thus, if the variance of the innovations $\mathsf{Var}(\varepsilon_t) = \Omega$ *is constant over time and* $T^{-1}\sum_{t=1}^T \varepsilon_t^{\otimes 2} \to \Omega$ *a.s. then* $\hat{\Omega}$ *is a consistent estimator for* Ω.

These two additional conditions appearing in Example 9.4 for the consistency of $\hat{\Omega}$ are for instance satisfied if the conditional variance of the innovations is constant $\mathsf{Var}(\varepsilon_t|\mathcal{F}_{t-1}) = \Omega$ *a.s.* This is demonstrated in the below Theorem.

Theorem 9.5 *Suppose Assumption 1.4 is satisfied and* $\mathsf{E}(\varepsilon_t^{\otimes 2}|\mathcal{F}_{t-1}) = \Omega$ *a.s. Then*

$$\frac{1}{T}\sum_{t=1}^T \varepsilon_t^{\otimes 2} = \Omega + \mathrm{o}\left(T^{-\zeta}\right), \qquad \text{for all } \zeta < \min\left(\xi, \frac{1}{2}\right) \text{ and } \xi < \frac{\gamma}{2+\gamma}.$$

Proof of Theorem 9.5. It suffices to consider the univariate case. The process $\sum_{s=1}^t \varepsilon_t^2 - \Omega$ is a martingale and therefore $T^{\zeta-1}(\sum_{t=1}^T \varepsilon_t^2 - \Omega) \to 0$ *a.s.* on the set

$$\left\{\sum_{t=1}^\infty t^{p(\zeta-1)} \mathsf{E}(\left\|\varepsilon_t^2 - \Omega\right\|^p \Big| \mathcal{F}_{t-1}) < \infty\right\}, \qquad \text{for } 1 \le p \le 2,$$

see Hall and Heyde (1980, Theorem 2.18). This set has probability one if $p \le 1 + \gamma/2$ and $p(\zeta - 1) < -1$ according to Assumption 1.4. These restrictions are satisfied when $\zeta < \min(\xi, 1/2)$. ■

The limiting distribution of the least squares estimator is in general complicated to describe except for the purely stationary case where a Central Limit Theorem can be formulated. Thus, consider the statistic

$$\sqrt{T}\left(\sum_{t=1}^{T}\tilde{U}_{t-1}^{\otimes 2}\right)^{-1/2}\left(\sum_{t=1}^{T}\tilde{U}_{t-1}\varepsilon_t'\right)\left(\sum_{t=1}^{T}\varepsilon_t^{\otimes 2}\right)^{-1/2}.$$

The middle term $\sum_{t=1}^{T}\tilde{U}_{t-1}\varepsilon_t'$ is a martingale and the Central Limit Theorem of Brown (1971) can therefore be applied. Slightly stronger assumptions to the innovations than (1.4), (1.5) are necessary in order to describe the correlation structure of $\tilde{U}_{t-1}\varepsilon_t'$. That is, assume the conditional variance of the innovations is constant over time

$$\mathsf{E}\left(\varepsilon_t^{\otimes 2}|\mathcal{F}_{t-1}\right)\overset{a.s.}{=}\Omega, \tag{10.1}$$

where Ω is positive definite, and therefore (1.5) is trivially satisfied. The conditions of Brown (1971, Theorem 1) are verified by the following Lemma which is proved towards the end of this section.

Lemma 10.1 *Suppose Assumptions* (1.4), (10.1) *are satisfied and* $\mathsf{E}\left\|\varepsilon_t\right\|^4<\infty$ *for some* $\eta>0$*. Define* $x_t=a'\tilde{U}_{t-1}\varepsilon_t'b$ *for arbitrary vectors* $a\in\mathbf{R}^{\dim\mathbf{U}}$, $b\in\mathbf{R}^{\dim X}$ *and let*

$$v_T^2=\sum_{t=1}^{T}\mathsf{E}\left(x_t^2|\mathcal{F}_{t-1}\right),\qquad s_T^2=\sum_{t=1}^{T}\mathsf{E}x_t^2.$$

Then it holds

$$\frac{v_T^2}{s_T^2}\overset{\mathsf{P}}{\to}1,\qquad \frac{1}{s_T^2}\sum_{t=1}^{T}\mathsf{E}\left\{x_t^2 1_{\left(x_t^2\geq\delta s_T^2\right)}\middle|\mathcal{F}_{t-1}\right\}\overset{\mathsf{P}}{\to}0. \tag{10.2}$$

Brown's Central Limit Theorem states that the condition (10.2) implies $s_T^{-1}\sum_{t=1}^{T}x_t$ converges in distribution to a standard normal. The Lemma 10.1 therefore gives the following result.

Theorem 10.2 *Suppose Assumptions* (1.4), (10.1) *are satisfied and* $\mathsf{E}\left\|\varepsilon_t\right\|^4<\infty$ *for some* $\eta>0$*. Then*

$$\sqrt{T}\left(\sum_{t=1}^{T}\tilde{U}_{t-1}^{\otimes 2}\right)^{-1/2}\left(\sum_{t=1}^{T}\tilde{U}_{t-1}\varepsilon_t'\right)\left(\sum_{t=1}^{T}\varepsilon_t^{\otimes 2}\right)^{-1/2}\overset{\mathsf{D}}{\to}\mathsf{N}\left(0,1\right).$$

In the above result the assumption of constant conditional variance (10.1) could be replaced by Brown's assumption (10.2) or alternatively the result could be based on a mixingale assumption for $\sum_{t=1}^{T}\tilde{U}_{t-1}\varepsilon_t'$, see McLeish (1977).

Proof of Lemma 10.1. It suffices to assume $\tilde{U}_t, \varepsilon_t$ are scalar processes, so $U_t = \mathbf{U}U_{t-1} + \varepsilon_t$ and $x_t = \tilde{U}_{t-1}\varepsilon_t$ with $\varepsilon_0 = \tilde{U}_0$. The assumption $\mathsf{E}\,\|\varepsilon_t\|^4 < \infty$ ensures the existence of all conditional expectations in the following.

Under Assumptions (1.4), (10.1) the Theorems 5.1, 9.1 imply that $T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \to \Omega$ and $T^{-1}\sum_{t=1}^{T}\tilde{U}_t^2 \to F_U = \sum_{t=0}^{\infty}\mathbf{U}\boldsymbol{\iota}\Omega\boldsymbol{\iota}'\mathbf{U}'$ *a.s.*

The first condition in (10.2) now follows by Assumption (10.1)

$$\begin{aligned}
\frac{1}{T}v_T^2 &= \frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left(x_t^2|\mathcal{F}_{t-1}\right) = \frac{1}{T}\sum_{t=1}^{T}\tilde{U}_{t-1}^2\mathsf{E}\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right) \overset{a.s.}{=} \frac{1}{T}\sum_{t=1}^{T}\tilde{U}_{t-1}^2\Omega \overset{a.s.}{\to} F_U\Omega, \\
\frac{1}{T}s_T^2 &= \frac{1}{T}\sum_{t=1}^{T}\mathsf{E}x_t^2 \overset{a.s.}{=} \frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\tilde{U}_{t-1}^2\Omega \overset{a.s.}{=} \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{U}^{2t}U_0 + \sum_{j=0}^{t-1}\mathbf{U}^{2j}\Omega\right)\Omega \to F_U\Omega.
\end{aligned}$$

As for the Lindeberg expression in (10.2) note that $1_{(x_t^2 \geq \delta s_T^2)} \leq |\delta s_T^2|^{-\eta}\,|x_t^2|^{\eta}$ for any $\eta > 0$ so

$$\mathsf{E}\left\{x_t^2 1_{(x_t^2\geq\delta s_T^2)}\middle|\mathcal{F}_{t-1}\right\} = \tilde{U}_t^2\mathsf{E}\left\{\varepsilon_t^2 1_{(x_t^2\geq\delta s_T^2)}\middle|\mathcal{F}_{t-1}\right\} \leq |\delta s_T^2|^{-\eta}\left(\tilde{U}_t^2\right)^{1+\eta}\mathsf{E}\left(\left|\varepsilon_t^2\right|^{1+\eta}\middle|\mathcal{F}_{t-1}\right).$$

The Lindeberg expression is consequently bounded by

$$\frac{1}{\delta|T^{-1}s_T^2|^{1+\eta}}\left(\frac{1}{T^{1+\eta}}\sum_{t=1}^{T}\left|\tilde{U}_{t-1}\right|^{2+2\eta}\right)\sup_t \mathsf{E}\left(\left|\varepsilon_t^2\right|^{1+\eta}\middle|\mathcal{F}_{t-1}\right).$$

Here the first term is convergent as proved above, the second sum converges to zero by Theorem 6.3, whereas the third term is bounded by assumption (1.4). ∎

11 *Correlations of processes with unit roots at particular frequencies*

Until now it has been argued that the sample correlations of the processes $\tilde{U}_t$, V_t, W_t are asymptotically negligible in an almost sure sense. If an argument holding in terms of weak convergence suffices a little more can be said. Using a similarity transformation as in §2 the process V_t can be decomposed into processes $V_{1,t}, \ldots, V_{l,t}$ with distinct roots on the unit circle. For the case where X is univariate Chan and Wei (1988) have shown that the sample correlations of these are asymptotically negligible and the order of convergence has been refined by Nielsen (2001a). A multivariate version will be stated here and a proof will be sketched.

To establish a common notation decompose the processes V_t and D_t into components $V_{1,t}, \ldots, V_{l,t}$ and $D_{1,t}, \ldots, D_{l,t}$, possibly of dimension zero, so $V_{j,t}$ and $D_{j,t}$ have roots at $\exp(\theta_j)$ and $\exp(-\theta_j)$ where $\theta_1, \ldots, \theta_l$ are distinct. As in the Section 10 the Assumption (10.1) that $\mathsf{E}(\varepsilon_t^{\otimes 2}|\mathcal{F}_{t-1}) = \Omega$ *a.s.* will be used. The primary function of this is to ensure that the processes $V_{1,t}, \ldots, V_{l,t}$, suitably normalised, converge jointly in distribution. A weakly asymptotic result for the sample correlations of these processes can now be formulated.

Theorem 11.1 *Suppose Assumptions 1.4, 1.6, 10.1 are satisfied. Let S_t be given by $\tilde{U}_t$, W_t and components $V_{q,t}$, $D_{q,t}$ so $\theta_q \neq \theta_r$. Then it holds for all $\eta > 0$*

$$\left(\sum_{t=1}^{T} S_t^{\otimes 2}\right)^{-1/2} \left\{\sum_{t=1}^{T} S_t \begin{pmatrix} V_{r,t} \\ D_{r,t} \end{pmatrix}'\right\} \left\{\sum_{t=1}^{T} \begin{pmatrix} V_{r,t} \\ D_{r,t} \end{pmatrix}^{\otimes 2}\right\}^{-1/2} = \mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right).$$

With this result in hand the order of magnitudes reported in Table 1 can now be strengthened using Lemma 3.5 and Theorem 5.2, 8.1, 8.2.

	$\tilde{U}_t$	$V_{r,t}$	$V_{s,t}$	W_t	$D_{r,t}$	$D_{s,t}$
$\tilde{U}_t$	1	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{O}\left(T^{-\xi/2}\right)$	$\mathrm{o}\left(T^{\eta-1/2}\right)$	$\mathrm{o}\left(T^{\eta-1/2}\right)$
$V_{r,t}$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	1	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	1	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$
$V_{s,t}$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	1	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	1
W_t	$\mathrm{O}\left(T^{-\xi/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	1	$\mathrm{O}\left(T^{-1/2}\right)$	$\mathrm{O}\left(T^{-1/2}\right)$
$D_{r,t}$	$\mathrm{o}\left(T^{\eta-1/2}\right)$	1	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	$\mathrm{O}\left(T^{-1/2}\right)$	1	$\mathrm{O}\left(T^{-1}\right)$
$D_{s,t}$	$\mathrm{o}\left(T^{\eta-1/2}\right)$	$\mathrm{o}_{\mathsf{P}}\left(T^{\eta-1/2}\right)$	1	$\mathrm{O}\left(T^{-1/2}\right)$	$\mathrm{O}\left(T^{-1}\right)$	1

Table 4: Asymptotic order of magnitude of pairwise correlations, where $\eta > 0$ and $\xi < \gamma/(2+\gamma)$.

The proof of Theorem 11.1 will be sketched in the following. The case where X is a univariate and D is absent has been analysed by Chan and Wei (1988) and then extended to include D_t by Nielsen (2001a). Since the argument for the general multivariate case is merely a tedious exercise in notation a detailed proof is omitted. It has some interest though to give an overview of the arguments of Chan and Wei (1988) so as to see where the assumption (10.1) is used.

The first step is an invariance principle for partial sums of the innovations.

Lemma 11.2 (*Chan and Wei, 1988, Theorem 2.2*)
Suppose Assumptions 1.4, 10.1 are satisfied and that $\dim X = 1$. *Then*

$$\frac{\sqrt{2}}{\sqrt{T}}\left(\sum_{t=1}^{[Tu]} \varepsilon_t \cos t\theta_1, \sum_{t=1}^{[Tu]} \varepsilon_t \sin t\theta_1, \ldots, \sum_{t=1}^{[Tu]} \varepsilon_t \cos t\theta_l, \sum_{t=1}^{[Tu]} \varepsilon_t \sin t\theta_l\right) \xrightarrow{\mathsf{D}} B_u$$

on the space $D[0,1]^{2l}$ of right continuous functions with limits from the left, where B is a standard Brownian motion.

It is worth noting the marginal convergence of a element such as $\sum_{t=1}^{[Tu]} \varepsilon_t \cos t\theta_1$ can be established under weaker conditions. For instance if it holds that

$$\frac{1}{T}\sum_{t=1}^{T} \mathsf{E}(\varepsilon_t^{\otimes 2}|\mathcal{F}_{t-1}) \overset{a.s.}{=} \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}\varepsilon_t^{\otimes 2} + \mathrm{o}\,(1)$$

together with Assumptions 1.4, 1.5 the invariance principle of Brown (1971, Theorem 3) can be used. Alternatively, the invariance principle for mixingales by McLeish (1977) could be used. The tricky bit is to show the joint convergence of the components. This holds if the average of the conditional covariances of the innovations is asymptotically negligible. To see this Chan and Wei (1988) use Assumption 10.1 and a trigonometric argument to show, for instance,

$$\frac{1}{T}\sum_{t=1}^{[Tu]} \mathsf{E}\left(\varepsilon_t \cos t\theta_1 \varepsilon_t \sin t\theta_1 \middle| \mathcal{F}_{t-1}\right) \overset{a.s.}{=} \Omega \frac{1}{T}\sum_{t=1}^{[Tu]} \cos t\theta_1 \sin t\theta_1 \to 0.$$

The second step is to look at the sample correlations of the processes $V_{1,t}, \ldots, V_{l,t}$ for the univariate case. While the above Lemma 11.2 shows that these processes have quite a bit in common with random walks their sample correlations have a complete different behaviour. For two random walks constructed from two sets of independent innovations the sample correlation will converge in distribution whereas the trigonometric construction of $V_{1,t}, \ldots, V_{l,t}$ ensures that their sample correlations are negligible. The argument is very similar to that made for the deterministic component D_t in Section 3. As long as $\dim X = 1$ and $\dim D = 0$ it can be assumed without loss of generality that $\mathbf{V}_r$ is of Jordan form (3.2) with (Λ, I) given by (3.3) which leads to the following result.

Lemma 11.3 (*Chan and Wei, 1988, Section 3*)
Suppose Assumptions 1.4, 10.1 are satisfied and that $\dim X = 1$ *and* $\dim D = 0$. *Let* S_t *be given by* $\tilde{U}_t$ *and components* $V_{q,t}$ *so* $\theta_q \neq \theta_r$. *Then for* $\eta > 1/2$ (!)

$$\left(\sum_{t=1}^{T} S_t^{\otimes 2}\right)^{-1/2} \left(\sum_{t=1}^{T} S_t V_{r,t}'\right) \left(\sum_{t=1}^{T} V_{r,t}^{\otimes 2}\right)^{-1/2} = \mathrm{o}_{\mathsf{P}}\left(T^{\eta - 1/2}\right).$$

The third and final step is to show that the result actually holds for $\eta > 0$ and to allow explosive and deterministic components. This extension is given by Nielsen (2001a, Lemma A4, B1).

These three steps can be extended to give Theorem 11.1. For the univariate case Lemma 11.3 has been extended by Chan (1989) to situations where $\dim D \neq 0$.

In the proof of Lemma 11.3 a normalisation, $N_{r,T}$ say, is found ensuring that $\sum_{t=1}^{T}(N_{r,T}V_{r,t})^{\otimes 2}$ converges in distribution to a random matrix which is positive definite $a.s.$ In this purely random situation $N_{r,T}^{-1} = \mathrm{O}(T^{-1})$ whereas in the general situation including deterministic terms the inverse of the normalisation term is $\mathrm{O}(T^{-1/2})$ as found by Chan (1989), see also Lemma 3.4, so

$$\left\{\sum_{t=1}^{T}\begin{pmatrix} V_{r,t} \\ D_{r,t} \end{pmatrix}^{\otimes 2}\right\}^{-1} = \mathrm{O}_{\mathsf{P}}\left(T^{-1}\right) \quad \text{and} \quad \left\{\sum_{t=1}^{T}(V_{r,t}|D_{r,t})^{\otimes 2}\right\}^{-1} = \mathrm{O}_{\mathsf{P}}\left(T^{-2}\right). \quad (11.1)$$

The Corollary 8.5, giving an almost sure lower bound for $\sum_{t=1}^{T}(\mathbf{X}_t|D_t)^{\otimes 2}$, can therefore be generalised as

$$\left\{\sum_{t=1}^{T}\begin{pmatrix} \mathbf{X}_t \\ D_t \end{pmatrix}^{\otimes 2}\right\}^{-1} = \mathrm{O}_{\mathsf{P}}\left(T^{-1}\right).$$

Together with Theorem 9.1 this implies that the least squares estimator for μ is at least weakly consistent as long as Assumptions 1.4, 1.6, 10.1 are satisfied.

References

Brown, B.M. (1971). Martingale Central Limit Theorems. *Annals of Mathematical Statistics* 42, 59-66.

Chan, N.H. (1989) Asymptotic inference for unstable autoregressive time series with drifts. *Journal of Statistical Planning and Inference* 23, 301-312.

Chan, N.H., and C.Z. Wei (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* 16, 367-401.

Duflo, M., Senoussi, R. and Touati, R. (1991). Propriétés asymptotiques presque sûre de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 27, 1-25.

Hall, P., and Heyde, C.C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, San Diego.

Herstein, I.N. (1975). *Topics in Algebra*, 2nd edition. New York: Wiley.

Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models.* 2nd printing. Oxford University Press.

Johansen, S. (2000). A Bartlett correction factor for tests on the cointegrating relations. *Econometric Theory* 16, 740-778.

Lai, T.L. and Wei, C.Z. (1982a). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* 10, 154-166.

Lai, T.L. and Wei, C.Z. (1982b). Asymptotic properties of projections with applications to stochastic regression problems. *Journal of Multivariate Analysis* 12, 346-370.

Lai, T.L. and Wei, C.Z. (1983a) A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics, Academia Sinica* 11, 1-13.

Lai, T.L. and Wei, C.Z. (1983b) Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analysis* 13, 1-23.

Lai, T.L. and Wei, C.Z. (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P.R. Krishnaiah, ed., *Multivariate Analysis VI*, Elsevier Science Publishers, 375-393.

Magnus, J.R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics.* Revised edition. Chichester: Wiley.

McLeish, D.L. (1977). On the invariance principle for nonstationary mixingales. *Annals of Probability* 5, 616-621.

Mirsky, L. (1961). *An introduction to linear algebra.* Corrected print. Oxford University Press.

Nielsen, B. (2000). The asymptotic distribution of likelihood ratio test statistics for cointegration in unstable vector autoregressive processes. Discussion paper.

Nielsen, B. (2001a). The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica,* **69**, 211-219.

Nielsen, B. (2001b). Weak consistency of criterions for order determination in a general vector autoregression. Mimeo.

Pötscher, B.M. (1989). Model selection under nonstationarity: Autoregressive models and stochastic linear regression models. *Annals of Statistics* 17, 1257-1274.

Rubin, H. (1950). Consistency of maximum-likelihood estimates in the explosive case. *Statistical Inference in Dynamic Economic Models* (ed. T.C. Koopmans), 356-364, New York: Wiley.

Wei, C.Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *Annals of Statistics* 13, 1498-1508.