

Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms.

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>

22 October 2003

Abstract: A vector autoregression with deterministic terms and with no restrictions to its characteristic roots is considered. Strong consistency results for the least squares statistics are presented. This extends earlier results where deterministic terms have not been considered. In addition the convergence rates are improved compared with earlier results.

Keywords: *Least squares estimator, Strong consistency, Vector autoregression.*

1 Introduction

The strong consistency of least squares estimators in an vector autoregression with deterministic terms is studied. Autoregressions generally have three types of asymptotic behaviour in that they may be stationary, a random walk type process or an explosive process. For the econometric analysis of non-explosive time series it usually suffices to use weak consistency and weak convergence arguments as in the work by Phillips (1991) and Johansen (1996). When a time series has explosive features it is mathematically more natural to use strong consistency arguments exploiting that explosive processes tend to follow persistent trajectories with probability one.

The first results showing strong consistency for explosive first order autoregressions were due to Rubin (1950) and Anderson (1959), with some generalisations by for instance Fuller, Hasza and Goebel (1981) and Jeganathan (1988). A general strong consistency result for vector autoregressions was given by Lai and Wei (1985) and this is generalised here to a situation with deterministic terms as seen in econometric models. The employed techniques are to a large extent derived using methods presented by Lai and Wei (1982, 1983a,b, 1985) and Wei (1992).

The paper is organised so that §2 presents the model and an overview of the main results. The proof follows in §3-10 and will be outlined in §2.

The following notation is used throughout the paper: For a matrix α let $\alpha^{\otimes 2} = \alpha\alpha'$, whereas $\alpha \otimes \beta$ is the Kronecker product and equals for example $(\alpha_{11}\beta, \alpha_{12}\beta)$ if $\alpha \in \mathbf{R}^{1 \times 2}$. Further $\text{diag}(\alpha_1, \dots, \alpha_n)$ is a block diagonal matrix with diagonal blocks α_j . When α is symmetric then $\lambda_{\min}(\alpha)$ and $\lambda_{\max}(\alpha)$ are the smallest and the largest

eigenvalue respectively. The choice of norm is the spectral norm $\|\alpha\| = \{\lambda_{\max}(\alpha^{\otimes 2})\}^{1/2}$ implying that $\|\alpha^{-1}\| = \{\lambda_{\min}(\alpha^{\otimes 2})\}^{-1/2}$. While $\mathbf{E}(\varepsilon_t|\mathcal{F}_{t-1})$ 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 abbreviation *a.s.* is used for properties holding almost surely.

2 The autoregressive model and main results

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

$$X_t = \sum_{j=1}^k A_j X_{t-j} + \mu D_{t-1} + \varepsilon_t, \quad \text{for } t = 1, \dots, T, \quad (2.1)$$

where D_t is a deterministic term and ε_t an innovation term.

The innovations are required to satisfy the local Marcinkiewicz-Zygmund conditions for convergence of explosive processes introduced by Lai and Wei (1983a). These are that (ε_t) is a martingale difference sequence with respect to an increasing sequence of σ -fields (\mathcal{F}_t) with the properties that some conditional moments of higher order are bounded and that the conditional variance has positive definite limit points.

Assumption 2.1 *For some $\gamma > 2$ it holds $\sup_t \mathbf{E}(\|\varepsilon_t\|^{2+\gamma} | \mathcal{F}_{t-1}) < \infty$ a.s.*

Assumption 2.2 *$\liminf_{t \rightarrow \infty} \lambda_{\min} \mathbf{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) > 0$ a.s.*

Each of the Assumptions 2.1, 2.2 exclude the possibility that the innovations could be autoregressive conditional heteroscedastic (ARCH) as proposed by Engle (1982). Therefore these conditions would probably be perceived as too strong for non-explosive situations, but for general autoregressions they are convenient.

The deterministic term D_t is a vector of terms such as a constant, a linear trend, or seasonal dummies. Inspired by Johansen (2000) the deterministic terms are required to satisfy the difference equation

$$D_t = \mathbf{D}D_{t-1}, \quad (2.2)$$

where \mathbf{D} has characteristic roots on the complex unit circle and linearly independent coordinates as described by the following assumption.

Assumption 2.3 *$|\text{eigen}(\mathbf{D})| = 1$ and $\text{rank}(D_1, \dots, D_{\dim \mathbf{D}}) = \dim \mathbf{D}$.*

With this type of deterministic terms the time series can be written conveniently in companion form. The stacked process $\mathbf{X}_{t-1} = (X'_{t-1}, \dots, X'_{t-k})'$ satisfies

$$\mathbf{X}_t = \mathbf{B}\mathbf{X}_{t-1} + \mu D_{t-1} + e_{X,t},$$

when defining matrices $\mathbf{B}, \boldsymbol{\mu}$ and a process $e_{X,t}$ as

$$\mathbf{B} = \begin{Bmatrix} A_1 \cdots A_{k-1} & A_k \\ I_{p(k-1)} & 0 \end{Bmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu} \mathbf{D} \\ 0 \end{pmatrix}, \quad e_{X,t} = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix},$$

while $S_t = (\mathbf{X}_t, D_t)$ that combines \mathbf{X}_t with the deterministic process D_t satisfies

$$S_t = \mathbf{S}S_{t-1} + e_{S,t}, \quad (2.3)$$

where \mathbf{S} and $e_{S,t}$ are defined as

$$\mathbf{S} = \begin{pmatrix} \mathbf{B} & \boldsymbol{\mu} \\ 0 & \mathbf{D} \end{pmatrix}, \quad e_{S,t} = \begin{pmatrix} e_{X,t} \\ 0 \end{pmatrix}.$$

The main object of interest is the least squares estimator for the parameters $A_1, \dots, A_k, \boldsymbol{\mu}$ which has the form

$$(\hat{A}_1, \dots, \hat{A}_k, \hat{\boldsymbol{\mu}}) - (A_1, \dots, A_k, \boldsymbol{\mu}) = \sum_{t=1}^T \varepsilon_t S'_{t-1} \left(\sum_{t=1}^T S_{t-1}^{\otimes 2} \right)^{-1},$$

where $S_{t-1}^{\otimes 2}$ is the outer product $S_{t-1} S'_{t-1}$. The partial estimator for the dynamic parameters A_1, \dots, A_k can correspondingly be written in terms of the residuals $(\mathbf{X}_t | D_t)$ from regressing the companion vector \mathbf{X}_t on the deterministic terms D_t as

$$(\hat{A}_1, \dots, \hat{A}_k) - (A_1, \dots, A_k) = \sum_{t=1}^T \varepsilon_t (\mathbf{X}_{t-1} | D_{t-1})' \left\{ \sum_{t=1}^T (\mathbf{X}_{t-1} | D_{t-1})^{\otimes 2} \right\}^{-1},$$

while the least squares variance estimator satisfies

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T (X_t | S_{t-1})^{\otimes 2} = \frac{1}{T} \sum_{t=1}^T (\varepsilon_t | S_{t-1})^{\otimes 2}.$$

The first main result gives a bound for a studentised version of the joint estimator.

Theorem 2.4 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then*

$$\left(\sum_{t=1}^T S_{t-1}^{\otimes 2} \right)^{-1/2} \sum_{t=1}^T S_{t-1} \varepsilon'_t \stackrel{a.s.}{=} o \left\{ T^{(1-\xi)/2} \right\} \quad \text{for all } \xi < \gamma / (2 + \gamma).$$

The remainder term can be decomposed as a sum of the following terms:

$$\begin{aligned} O(\max_{t \leq T} \|\varepsilon_t\|) &= o \left\{ T^{(1-\xi)/2} \right\} && \text{if } \max |\text{eigen}(\mathbf{B})| > 1, \\ &O\{(\log T)^{1/2}\} && \text{if } \mathbf{B} \text{ has eigenvalues of length 1,} \\ &O\{(\log \log T)^{1/2}\} && \text{if } \min |\text{eigen}(\mathbf{B})| < 1. \end{aligned}$$

Special cases have been proved by Pötscher (1989, Lemma A.1) for $\dim \mathbf{D} = 0$ and $\xi = 0$ and by Nielsen (2001a) who proves a univariate version holding in probability.

The proof of Theorem 2.4 relies on a separation of the stochastic and deterministic components using that $(\sum_{s=1}^T S_{s-1}^{\otimes 2})^{-1/2} S_{t-1}$ can be rewritten as

$$\left(\sum_{s=1}^T D_{s-1}^{\otimes 2}\right)^{-1/2} D_{t-1} + \left\{\sum_{s=1}^T (X_{s-1}|D_{s-1})^{\otimes 2}\right\}^{-1/2} (X_{t-1}|D_{t-1})$$

by partitioned inversion. For the least squares estimator itself a more complete understanding of the interaction between the deterministic components and unit root processes appearing in the denominator matrix is needed. Such results are not available as yet and the following consistency results are therefore only partial although they do represent some improvement over previous results and include a complete description of the pure stationary and purely explosive cases where $|\text{eigen}(\mathbf{B})| \neq 1$.

Theorem 2.5 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then*

$$(\hat{A}_1, \dots, \hat{A}_k) - (A_1, \dots, A_k) \stackrel{a.s.}{=} O\{(T^{-1} \log T)^{1/2}\}.$$

If \mathbf{B} and \mathbf{D} have no common eigenvalues then

$$(\hat{A}_1, \dots, \hat{A}_k, \hat{\mu}) - (A_1, \dots, A_k, \mu) \stackrel{a.s.}{=} o(T^{-\xi/2}) \quad \text{for all } \xi < \gamma/(2 + \gamma).$$

The issue of strong consistency of the least squares estimator was first discussed for a univariate, explosive, Gaussian first order autoregression, with $\dim \mathbf{X} = 1$, $\dim \mathbf{D} = 0$, by Rubin (1950) and Anderson (1959). Lai and Wei (1985, Theorem 4) studied the special case without deterministic terms, so $\dim \mathbf{D} = 0$, and gave a weaker result with $\xi = 0$. A related generalisation has previously been presented by Duflo, Senoussi and Touati (1991, Theorem 1) in the case where the explosive roots have multiplicity one whereas their Theorem 2 seems false in suggesting that the least square estimator for \mathbf{B} otherwise is inconsistent.

A direct consequence of Theorem 2.4 concerning the studentised least squares estimator is that the least squares variance estimator can be estimated consistently.

Corollary 2.6 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then*

$$\hat{\Omega} \stackrel{a.s.}{=} \frac{1}{T} \sum_{t=1}^T \varepsilon_t^{\otimes 2} + o(T^{-\xi}) \quad \text{for all } \xi < \gamma/(2 + \gamma).$$

While the Assumptions 2.1, 2.2 suffice to ensure that the sequence $(T^{-1} \sum_{t=1}^T \varepsilon_t^{\otimes 2})$ is relatively compact with positive definite limit points as argued by Lai and Wei (1985), some further structure is needed to get convergence to an interpretable matrix. In the light of the Assumptions 2.1, 2.2 it is convenient to apply the following sufficient condition used by Chan and Wei (1988).

Assumption 2.7 $E(\varepsilon_t^{\otimes 2} | \mathcal{F}_{t-1}) = \Omega$ *a.s. where Ω is positive definite.*

This gives rise to the following convergence result.

Theorem 2.8 *Suppose Assumptions 2.1, 2.7 are satisfied. Then*

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_t^{\otimes 2} \stackrel{a.s.}{=} \Omega + o(T^{-\zeta}), \quad \text{for all } \zeta < \min\left(\frac{\gamma}{2+\gamma}, \frac{1}{2}\right).$$

Corollary 2.6 and Theorem 2.8 lead to an immediate result for the least squares variance estimator.

Corollary 2.9 *Suppose Assumptions 2.1, 2.3, 2.7 are satisfied. Then*

$$\hat{\Omega} \stackrel{a.s.}{=} \frac{1}{T} \sum_{t=1}^T \varepsilon_t^{\otimes 2} + o(T^{-\xi}) \quad \text{for all } \xi < \min\left(\frac{\gamma}{2+\gamma}, \frac{1}{2}\right).$$

There is potential for many other econometric applications of Theorem 2.4. An example is lag length determination where it is conceptually important to establish the lag length before determining the location of the characteristic roots, see Pötscher (1989), Wei (1992) and Nielsen (2001b). Other examples are unit root testing (Nielsen 2001a) and cointegration analysis (Nielsen, 2000) where the asymptotic inference results also can be used without knowledge about the characteristic roots.

The remainder of the paper gives the proofs of these results. To a large extent the proofs follow those of Lai and Wei (1983b and 1985), but with many modifications because of the deterministic term. The proof is outlined as follows. In §3 the process \mathbf{X}_t is decomposed into autoregressive components U_t, V_t, W_t with characteristic roots outside, on, and inside the unit circle, respectively. The order of magnitude of the deterministic process, D_t , and the process itself, \mathbf{X}_t , is discussed in §4 and §5. The next sections are concerned with the order of magnitude of the denominator matrix $\mathbf{M}_T = \sum_{t=1}^T S_{t-1}^{\otimes 2}$. As a first step the sample correlation of U_t and D_t is considered in §6. The order of magnitude of the largest and the smallest eigenvalue of \mathbf{M}_T is then discussed in §7 and §8. The next step is to discuss sample correlations of all possible combinations of the processes U_t, V_t, W_t, D_t in §9 and finally the main results are proved in §10.

3 Decompositions of the process

The process X_t is decomposed in two ways to facilitate the subsequent asymptotic analysis. The first decomposition concerns the stochastic part of process while the second decomposition disentangles deterministic and stochastic parts of the process.

The first decomposition separates the eigenvalues of the companion matrix \mathbf{B} for the stochastic part of the process. Following Herstein (1975, p. 308) there exists a regular, real matrix M which transforms \mathbf{B} into a real block diagonal matrix with

blocks \mathbf{U} , \mathbf{V} and \mathbf{W} having eigenvalues with absolute value less than one, equal to one and larger than one, respectively. That is

$$\begin{pmatrix} M & 0 \\ 0 & I_{\dim \mathbf{D}} \end{pmatrix} S_t = \begin{pmatrix} U_t \\ V_t \\ W_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 & 0 & \mu_U \\ 0 & \mathbf{V} & 0 & \mu_V \\ 0 & 0 & \mathbf{W} & \mu_W \\ 0 & 0 & 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} U_{t-1} \\ V_{t-1} \\ W_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} e_{U,t} \\ e_{V,t} \\ e_{W,t} \\ 0 \end{pmatrix}.$$

For the purpose of proving the results of §2 it can be assumed without loss of generality that $\mathbf{B} = \text{diag}(\mathbf{U}, \mathbf{V}, \mathbf{W})$ is block diagonal and $S_t = (U'_t, V'_t, W'_t, D'_t)'$.

The second decomposition seeks to separate the stochastic and deterministic components and is based on two arguments. The processes U_t, W_t, D_t are first separated by a similarity transformation using that the matrices \mathbf{U}, \mathbf{W} and \mathbf{D} have no common eigenvalues, while the processes V_t, D_t have to be discussed in more detail since the matrices \mathbf{V} and \mathbf{D} may in general have common eigenvalues.

The processes U_t, W_t are linear functions of the deterministic process D_t and they are first shown to satisfy the relationships

$$U_t = \tilde{U}_t + \tilde{\mu}_U D_t \quad \text{where} \quad \tilde{U}_t = \mathbf{U} \tilde{U}_{t-1} + e_{U,t}, \quad (3.1)$$

$$W_t = \tilde{W}_t + \tilde{\mu}_W D_t \quad \text{where} \quad \tilde{W}_t = \mathbf{W} \tilde{W}_{t-1} + e_{W,t}, \quad (3.2)$$

with $\tilde{U}_0 = U_0 - \tilde{\mu}_U D_0$ and $\tilde{W}_0 = W_0 - \tilde{\mu}_W D_0$. The argument is the same in both cases. Taking U_t as an example consider the companion matrix for the vector (U'_t, D'_t) and apply a similarity transformation of the form

$$M = \begin{pmatrix} I_{\dim \mathbf{U}} & -\tilde{\mu}_U \\ 0 & I_{\dim \mathbf{D}} \end{pmatrix} \quad \text{so} \quad M \begin{pmatrix} \mathbf{U} & \mu_U \\ 0 & \mathbf{D} \end{pmatrix} M^{-1} = \begin{pmatrix} \mathbf{U} & \mu_U + \mathbf{U} \tilde{\mu}_U - \tilde{\mu}_U \mathbf{D} \\ 0 & \mathbf{D} \end{pmatrix}.$$

The result (3.1) then follows by arguing that $\tilde{\mu}_U$ can be chosen so $\mu_U + \mathbf{U} \tilde{\mu}_U - \tilde{\mu}_U \mathbf{D} = 0$ which is a consequence of the next Lemma 3.1.

Lemma 3.1 *Consider matrices $D \in \mathbf{R}^{d \times d}$, $N \in \mathbf{R}^{n \times n}$ and the equation $\tilde{\mu}D - N\tilde{\mu} = \mu$ for some $\mu, \tilde{\mu} \in \mathbf{R}^{n \times d}$. The solution $\tilde{\mu}$ is unique for all μ if and only if the matrices D and N have no common eigenvalues.*

Proof of Lemma 3.1. Following Magnus and Neudecker (1999, p.30) the equation of interest can be written as $\text{vec}(\tilde{\mu}D - N\tilde{\mu}) = (D' \otimes I_n - I_d \otimes N) \text{vec} \tilde{\mu} = \text{vec} \mu$. A unique solution exists when $D' \otimes I_n - I_d \otimes N$ has full rank.

Two properties of Kronecker products are needed. First, the two matrices $D' \otimes I_n$ and $I_d \otimes N$ 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 $D' \otimes I_n - I_d \otimes N$ is unitarily similar to a triangular matrix with diagonal elements given by $d_i - n_j$ where d_i and n_j are eigenvalues of D and N respectively. Hence, if D and N have no common eigenvalues then $D' \otimes I_n - I_d \otimes N$ has full rank.

Now suppose D and N have a common eigenvalue λ and let x and y be associated eigenvectors so $D'x = \lambda x$ and $Ny = \lambda y$. Since $(D' \otimes I_n - I_d \otimes N) \text{vec}(yx') = 0$ then $\text{vec}(yx')$ is an eigenvector for $D' \otimes I_n - I_d \otimes N$ associated with the eigenvalue 0. ■

In the special case where \mathbf{B} has no eigenvalues in common with \mathbf{D} the same argument can be made for the entire process \mathbf{X}_t as for the U_t, W_t processes. That is

$$\mathbf{X}_t = \tilde{\mathbf{X}}_t + \tilde{\mu}_X D_t \quad \text{where} \quad \tilde{\mathbf{X}}_t = \mathbf{B}\tilde{\mathbf{X}}_{t-1} + e_{X,t}. \quad (3.3)$$

When it comes to the general situation where \mathbf{V} and \mathbf{D} are allowed to have common eigenvalues it is convenient first to discuss the special case where \mathbf{V} and \mathbf{D} have their eigenvalues at one and are both Jordan matrices

$$\begin{pmatrix} \Lambda & E & & \\ & \ddots & \ddots & \\ & & \ddots & E \\ & & & \Lambda \end{pmatrix}, \quad (3.4)$$

with $(\Lambda, E) = (1, 1)$. In that situation V_t will be shown to satisfy

$$V_t = \tilde{V}_t + \tilde{\mu}_V \tilde{D}_t \quad \text{where} \quad \tilde{V}_t = \mathbf{V}\tilde{V}_{t-1} + e_{V,t} \quad \text{and} \quad \tilde{D}_t = \tilde{\mathbf{D}}\tilde{D}_{t-1}, \quad (3.5)$$

with $\tilde{V}_0 = V_0$ and where $\tilde{\mathbf{D}}$ is of the form (3.4) with $\dim \tilde{\mathbf{D}} = \dim \mathbf{V} + \dim \mathbf{D}$. To see this write the process (V_t, D_t) in companion form as

$$\begin{pmatrix} V_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mu_V \\ 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} V_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} e_{V,t} \\ 0 \end{pmatrix},$$

which has solution

$$\begin{pmatrix} V_t \\ D_t \end{pmatrix} = \sum_{j=0}^{t-1} \begin{pmatrix} \mathbf{V} & \mu_V \\ 0 & \mathbf{D} \end{pmatrix}^j \begin{pmatrix} e_{V,t-j} \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{V} & \mu_V \\ 0 & \mathbf{D} \end{pmatrix}^t \begin{pmatrix} V_0 \\ D_0 \end{pmatrix}. \quad (3.6)$$

Two properties of the companion matrix will be need. First, it satisfies

$$(I_{\dim \mathbf{V}}, 0) \begin{pmatrix} \mathbf{V} & \mu_V \\ 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} I_{\dim \mathbf{V}} \\ 0 \end{pmatrix} = \mathbf{V}.$$

Secondly, since \mathbf{V} and \mathbf{D} are both Jordan blocks of the form (3.4) with eigenvalues at one a matrix M exists so that

$$\mathbf{M} = M \begin{pmatrix} \mathbf{V} & \mu_V \\ 0 & \mathbf{D} \end{pmatrix} M^{-1}$$

is a block diagonal matrix with diagonal elements that are Jordan blocks of the form (3.4) with $(\Lambda, E) = (1, 1)$. Extracting the marginal solution for V_t from (3.6) and using these properties shows

$$V_t = \sum_{j=0}^{t-1} \mathbf{V}^j e_{V,t-j} + \mathbf{V}^t V_0 + (I_{\dim \mathbf{v}}, 0) M^{-1} \mathbf{M}^t M \begin{pmatrix} 0 \\ D_0 \end{pmatrix},$$

which in turn can be written as $\tilde{V}_t + \tilde{\mu}_V \tilde{\mathbf{D}}^t \tilde{D}_0$ for some $\tilde{\mu}_V$ and \tilde{D}_0 .

In general \mathbf{V} and \mathbf{D} can have eigenvalues anywhere on the unit circle. Suppose these occur at l distinct complex pairs $\exp(i\theta_j)$ and $\exp(-i\theta_j)$ for $0 \leq \theta_j \leq \pi$ which of course reduce to a single eigenvalue of 1 or -1 if θ_j equals 0 or π . Following Herstein (1975, p. 308) and using Assumption 2.3 to \mathbf{D} there exists regular, real similarity transformations M_V and M_D that block-diagonalise \mathbf{V} and \mathbf{D} as

$$\begin{aligned} M_V \mathbf{V} M_V^{-1} &= \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_l) & \text{where} & & \mathbf{V}_j &= \text{diag}(\mathbf{V}_{j,1}, \dots, \mathbf{V}_{j,m_j}), \\ M_D \mathbf{D} M_D^{-1} &= \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_l), \end{aligned}$$

where the sub-blocks $\mathbf{V}_{j,m}$ and \mathbf{D}_j are real Jordan matrices of the form (3.4), and where (Λ, E) is one of the pairs

$$(1, 1), \quad (-1, 1), \quad \text{or} \quad \left\{ \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \text{ for } 0 < \theta < \pi. \quad (3.7)$$

Using the same argument as above it therefore holds in general that V_t has representation (3.5) where \tilde{V}_t has sub-components $\tilde{V}_{j,t}$ and $\tilde{V}_{j,m,t}$ of dimension $\dim \mathbf{V}_j$ and $\dim \mathbf{V}_{j,m}$ respectively and \tilde{D}_t has sub-components $\tilde{D}_{j,t}$ of dimension $\dim \tilde{\mathbf{D}}_j = \dim \mathbf{D}_j + \max_m \dim \mathbf{V}_{j,m}$.

Combining the results (3.1), (3.2), (3.5) with the notation

$$\tilde{\mathbf{X}}_t = \begin{pmatrix} \tilde{U}_t \\ \tilde{V}_t \\ \tilde{W}_t \end{pmatrix}, \quad \tilde{\mu} = \begin{pmatrix} (0, \tilde{\mu}_U) \\ \tilde{\mu}_V \\ (0, \tilde{\mu}_W) \end{pmatrix},$$

shows that the process without loss of generality can be represented as

$$\mathbf{X}_t = \tilde{\mathbf{X}}_t + \tilde{\mu} \tilde{D}_t \quad \text{where} \quad \tilde{\mathbf{X}}_t = \mathbf{B} \tilde{\mathbf{X}}_{t-1} + e_{X,t}, \quad \tilde{D}_t = \tilde{\mathbf{D}} \tilde{D}_{t-1}. \quad (3.8)$$

It is convenient also to introduce the dimensions

$$\delta_j = \dim \mathbf{D}_j / \dim \Lambda_j, \quad \tilde{\delta}_j = \dim \tilde{\mathbf{D}}_j / \dim \Lambda_j, \quad \nu_{j,m} = \dim \mathbf{V}_{j,m} / \dim \Lambda_j,$$

as well as the maxima

$$\delta = \max_j \delta_j, \quad \tilde{\delta} = \max_j \tilde{\delta}_j, \quad \nu = \max_{j,m} \nu_{j,m}.$$

4 Limiting results for the deterministic component

In the following the order of magnitude of the deterministic process D_t and the denominator matrix $\sum_{t=1}^T D_t^{\otimes 2}$ will be described.

The main result is formulated using normalisation matrices

$$N_D = \text{diag}(N_{\delta_1}, \dots, N_{\delta_l}) \quad \text{where} \quad N_n = \text{diag}\{(\Lambda_j/T)^{n-1}, \dots, (\Lambda_j, T)^0\}. \quad (4.1)$$

Theorem 4.1 *Suppose Assumption 2.3 is satisfied. Then it holds*

- (i) $\max_{t \leq T} \|N_D D_t\| = O(1)$ and in particular $\|D_T\| = O(T^{\delta-1})$.
- (ii) $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T (N_D D_{t-1})^{\otimes 2}$ is positive definite.
- (iii) $\max_{t < T} D'_t (\sum_{s=1}^T D_{s-1}^{\otimes 2})^{-1} D_t = O(T^{-1})$.

Two lemmas concerning the order of $D_{j,t}$ and $\sum_{t=1}^T D_{j,t}^{\otimes 2}$ are needed.

Lemma 4.2 *Let $D_{j,1,0}$ denote the last element of the initial vector $D_{j,0}$. Then*

$$N_{\delta_j} D_{j,t} = f(\delta_j, t/T) \otimes \Lambda_j^t D_{j,1,0} + O(T^{-1}), \quad \text{uniformly in } t,$$

where f is the polynomial vector $f(n+1, u) = \{u^n/(n!), \dots, u^0/(0!)\}'$.

Proof of Lemma 4.2. The process $D_{j,t}$ satisfies $D_{j,t} = \mathbf{D}_j^t D_{j,0}$ where

$$\mathbf{D}_j^t = \begin{Bmatrix} \Lambda^t & \text{b}(t, 1)\Lambda^{t-1} & \dots & \text{b}(t, \delta_j - 1)\Lambda^{t-\delta_j+1} \\ 0 & \Lambda^t & & \text{b}(t, \delta_j - 2)\Lambda^{t-\delta_j+2} \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \Lambda^t \end{Bmatrix},$$

and $\text{b}(\cdot, \cdot)$ is the binomial coefficient. The desired result then follows by noting that $\text{b}(t, n) = t^n/n! + O(t^{n-1})$ uniformly in t while $\|\Lambda^n\| \leq \|\Lambda\|^n = 1$. ■

Lemma 4.3

- (i) $T^{-1} \sum_{t=1}^T (N_{\delta_j} D_{j,t-1})^{\otimes 2} = (\|D_{j,1,0}\|^2 / \dim \Lambda_j) \int_0^1 f(\delta_j, u)^{\otimes 2} du \otimes E_j + O(T^{-1})$.
- (ii) *If Assumption 2.3 is satisfied then the limiting matrix in (i) is positive definite.*
- (iii) $T^{-1} \sum_{t=1}^T (N_{\delta_j} D_{j,t-1})(N_{\delta_m} D_{m,t-1})' = O(T^{-1})$ for $j \neq m$.

Proof of Lemma 4.3. (i) By Lemma 4.2 the cross product satisfies

$$\frac{1}{T} \sum_{t=1}^T \left(\frac{t-1}{T}\right)^{p+q-2} \Lambda_j^{t-1} D_{j,1,0} D'_{j,1,0} (\Lambda_j^{t-1})' + O(T^{-1}).$$

Trigonometric identities show that $\Lambda_j^t D_{j,1,0} D'_{j,1,0} (\Lambda_j^t)'$ equals $E_j \|D_{j,1,0}\|^2 / \dim \Lambda_j + R_j$ where $R_j = 0$ for $\dim \Lambda_j = 1$ and $R_j = \cos(2\theta t) A + \sin(2\theta t) B$ for some constant matrices A, B when $\dim \Lambda_j = 2$. When $\dim \Lambda_j = 1$ the desired result follows immediately, whereas when $\dim \Lambda_j = 2$ it follows from the result

$$T^{-1} \sum_{t=1}^T (t/T)^{p+q-2} \cos(2\theta t + a) = O(T^{-1}) \quad \text{for } 0 < \theta < \pi,$$

for any constant a , see Gradshteyn and Ryzhik (1965, 2.633.2).

(ii) Note that the vector $f(n, u)$ can be expressed as a non-singular linear transformation of the first n Legendre polynomials, $p(n, u)$ say, which have the property that $\int_0^1 p(n, u)^{\otimes 2} du = I_n$ giving the positive definiteness.

(iii) Use the same type of arguments as in (i), noting that $\Lambda_j^t D_{j,1,0} D'_{m,1,0} (\Lambda_m^t)'$ equals $\cos(2\theta t) A + \sin(2\theta t) B$ for some constant matrices A, B . ■

Proof of Theorem 4.1. (i) The result follows from Lemma 4.2 by stacking the processes $N_{\delta_j} D_{j,t}$ and using the triangle inequality.

(ii) Lemma 4.3 implies that $T^{-1} \sum_{t=1}^T (N_D D_t)^{\otimes 2}$ converges to a block diagonal matrix with positive definite diagonal elements.

(iii) The desired result follows from (i), (ii) and replacing each D_t with $N_D D_t$. ■

5 The order of magnitude of the process

In the following the order of magnitude of the process \mathbf{X}_t is investigated. This is a generalisation of Lai and Wei (1985, Theorem 1) where the case without deterministic components is considered. Subsequently a convergence result is given for the explosive component W_t .

Theorem 5.1 *Suppose Assumptions 2.1, 2.3 are satisfied. Then, for $\xi < \gamma / (2 + \gamma)$,*

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

Proof of Theorem 5.1. By (3.8) it holds $\mathbf{X}_t = \tilde{\mathbf{X}}_t + \tilde{\mu} \tilde{\mathbf{D}}_t$. Lai and Wei (1985, Theorem 1) show the results for the purely stochastic component $\tilde{\mathbf{X}}_t$ whereas the order of the deterministic component $\tilde{\mathbf{D}}_t$ follows from Lemma 4.2. ■

When studying the process \tilde{W}_t Lai and Wei (1985) use the following generalisation of the Marcinkiewicz-Zygmund Theorem.

Theorem 5.2 *(Lai and Wei, 1983a, Corollaries 3 and 4)*

Suppose Assumptions 2.1, 2.2 are satisfied. Then for any sequence of matrices A_t the series $\sum_{t=1}^T A_t \varepsilon_t$ converges a.s. if and only if the series $\sum_{t=1}^T \|A_t\|$ converges.

If this holds, and $A_t \neq 0$ for infinitely many t then $\mathbf{P}(\sum_{t=1}^{\infty} A_t \varepsilon_t = Y) = 0$ 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 5.3 *Suppose Assumption 2.1, 2.2, 2.3 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 $\mathbf{P}(a'W = 0) = 0$ for any $a \in \mathbf{R}^{\dim W}$.

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

Proof of Corollary 5.3. (i) The decomposition (3.2) and Theorem 4.1, *i* show that $W_T = \tilde{W}_T + o(\mathbf{W}^T)$. The result then follows from Theorem 5.2, see also Lai and Wei (1985, Lemma 2).

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

6 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 §7 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 results of Lai and Wei concerning the matrices $T^{-1} \sum_{t=1}^T \varepsilon_t^{\otimes 2}$ and $T^{-1} \sum_{t=1}^T \tilde{U}_t^{\otimes 2}$ are stated. The results give conditions for relative compactness of sequences 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 6.1 (Lai and Wei, 1985, Theorem 2, equation 3.7, Example 3).

Suppose Assumption 2.1 is satisfied. Then, with probability one, the matrix sequences

$$\left(\frac{1}{T} \sum_{t=1}^T \varepsilon_t^{\otimes 2} : T \geq 1 \right), \quad \left\{ \frac{1}{T} \sum_{t=1}^T \mathbf{E}(\varepsilon_t^{\otimes 2} | \mathcal{F}_{t-1}) : T \geq 1 \right\}$$

are relatively compact with the same limit points.

If in addition Assumption 2.2 is satisfied the limit points are positive definite.

Since $e_{U,t}$ is a linear combination of $e_{X,t}$ the sequence $(T^{-1} \sum_{t=1}^T e_{U,t}^{\otimes 2} : T \geq 1)$ is therefore relatively compact. In addition the following results can be shown.

Theorem 6.2 (*Lai and Wei, 1985, Theorem 2, Example 3*).

Suppose Assumptions 2.1, 2.2 are satisfied. Then it holds with probability one that

$$\begin{aligned} \mathbf{E} \text{ is a limit point of } \left(\frac{1}{T} \sum_{t=1}^T e_{X,t}^{\otimes 2} : T \geq 1 \right) &\Rightarrow \\ \sum_{j=0}^{q-1} \mathbf{B}^j \mathbf{E} (\mathbf{B}^j)' \text{ is a limit point of } \left\{ \sum_{j=0}^{q-1} \mathbf{B}^j \frac{1}{T} \sum_{t=1}^T e_{X,t}^{\otimes 2} (\mathbf{B}^j)' : T \geq 1 \right\} & \\ \text{and positive definite if } q \geq k \dim X, \text{ where } k \text{ is the lag length,} & \end{aligned}$$

as well as it holds

$$\begin{aligned} \mathbf{E}_U \text{ is a limit point of } \left(\frac{1}{T} \sum_{t=1}^T e_{U,t}^{\otimes 2} : T \geq 1 \right) &\Rightarrow \\ \sum_{t=0}^{\infty} \mathbf{U}^t \mathbf{E}_U (\mathbf{U}^t)' \text{ is positive definite and a limit point of } \left(\frac{1}{T} \sum_{t=1}^T \tilde{U}_t^{\otimes 2} : T \geq 1 \right). & \end{aligned}$$

Before turning to the sample correlation of \tilde{U}_t and D_t it is useful to cite the following univariate result by Wei (1985).

Lemma 6.3 (*Wei 1985, Lemma 2*)

Suppose Assumption 2.1 is satisfied. Let (x_t) be a sequence of random variables adapted to (\mathcal{F}_t) with $s_T^2 = \sum_{t=1}^T x_t^2$. Assume $x_T^2 = o(s_T^{2-\eta})$ a.s. for some $\eta > 0$. Then

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

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

Theorem 6.4 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then, for all $\eta > 0$,*

$$\left(\sum_{t=1}^T \tilde{U}_{t-1}^{\otimes 2} \right)^{-1/2} \left(\sum_{t=1}^T \tilde{U}_{t-1} D'_{t-1} \right) \left(\sum_{t=1}^T D_{t-1}^{\otimes 2} \right)^{-1/2} \stackrel{\text{a.s.}}{=} o(T^{\eta-1/2}).$$

Proof of Theorem 6.4. Theorem 6.2 shows that $\| \sum_{t=1}^T \tilde{U}_{t-1}^{\otimes 2} \|^{-1/2}$ is $O(T^{-1/2})$ so it suffices to show that $(\sum_{t=1}^T \tilde{U}_{t-1} D'_{t-1})(\sum_{t=1}^T D_{t-1}^{\otimes 2})^{-1/2}$ is $o(T^\eta)$.

The main contribution arises from the sum $(\sum_{t=T^\alpha}^T \tilde{U}_t D'_t)(\sum_{t=1}^T D_t^{\otimes 2})^{-1/2}$ for an α satisfying $1 > \alpha > 0$. With this in mind and using $\tilde{U}_t = \mathbf{U} \tilde{U}_{t-1} + e_{U,t}$ the object of interest can be written as

$$\left(\sum_{t=1}^{T^\alpha} \tilde{U}_{t-1} D'_{t-1} + \mathbf{U}^{T^\alpha} \sum_{t=T^\alpha}^{T-1} \tilde{U}_{t-T^\alpha} D'_t + \sum_{t=T^\alpha}^{T-1} \sum_{s=0}^{T^\alpha-1} \mathbf{U}^s e_{U,t-s} D'_t \right) \left(\sum_{t=1}^T D_{t-1}^{\otimes 2} \right)^{-1/2}. \quad (6.1)$$

The first two terms in (6.1) are $o(T^\eta)$. To see this bound their norm by

$$\left(T^\alpha \max_{1 \leq t \leq T^\alpha} \|\tilde{U}_t\| + T \|\mathbf{U}\|^{T^\alpha} \max_{1 \leq t \leq T} \|\tilde{U}_t\| \right) \left\{ \max_{t < T} D'_t \left(\sum_{s=1}^T D_{s-1}^{\otimes 2} \right)^{-1} D_t \right\}^{1/2},$$

and use Theorems 4.1, 5.1 and that $\|\mathbf{U}\|^{T^\alpha}$ decreases exponentially.

The third term in (6.1) is $o(T^\eta)$. To see this use $D_t = \mathbf{D}^s D_{t-s}$ and the normalisation N_D given in (4.1) to rewrite it as

$$\sum_{s=0}^{T^\alpha-1} \mathbf{U}^s \left(\sum_{t=T^\alpha}^{T-1} e_{U,t-s} D'_{t-s} N'_D \right) \left\{ \sum_{t=-s}^{T-1-s} (N_D D_t)^{\otimes 2} \right\}^{-1/2}.$$

The norm of this expression is bounded by

$$\sum_{s=0}^{T^\alpha-1} \|\mathbf{U}\|^s \left\| \sum_{t=T^\alpha}^{T-1} e_{U,t-s} D'_{t-s} N'_D \right\| \left\| \sum_{t=1}^{T-T^\alpha} (N_D D_{t-1})^{\otimes 2} \right\|^{-1/2}.$$

Since the sum $\sum_{s=0}^{T^\alpha-1} \|\mathbf{U}\|^s$ converges it suffices to show that the last two components can be approximated uniformly by a variable which is $o(T^\eta)$.

The sum $\sum_{t=1}^{T-T^\alpha} (N_D D_{t-1})^{\otimes 2}$ is approximately equal to $\sum_{t=1}^T (N_D D_{t-1})^{\otimes 2}$ which converges to a positive definite matrix, see Theorem 4.1. The norm of the approximation error, $\sum_{t=T-T^\alpha+1}^T (N_D D_{t-1})^{\otimes 2}$, is bounded by $T^\alpha \max_{t < T} \|N_D D_t\|^2 = O(T^{\alpha-1}) = o(1)$, due to Theorem 4.1.

In a similar way $\sum_{t=T^\alpha}^{T-1} e_{U,t-s} D'_{t-s}$ is approximately $\sum_{t=1}^{T-1} e_{U,t} D'_t$, which is not dependent on s . Considering each element of this matrix and applying Lemma 6.3 shows that this is $o(T^\eta N_D^{-1})$. The approximation error can be bounded by

$$2T^\alpha \max_{t \leq T} \|e_{U,t}\| \left\{ \max_{t < T} D'_t \left(\sum_{s=1}^T D_{s-1}^{\otimes 2} \right)^{-1} D_t \right\}^{1/2}.$$

Using Theorems 4.1, 5.1 this is seen to be $o(T^{\alpha-\xi/2})$ which is $o(1)$ for a small α . ■

Some immediate consequences of these results are the following examples.

Example 6.5 *Suppose Assumptions 2.1, 2.7 are satisfied. Then Theorems 6.1 and 6.2 imply $T^{-1} \sum_{t=1}^T \varepsilon_t^{\otimes 2} \rightarrow \Omega$ a.s. and $T^{-1} \sum_{t=1}^T \tilde{U}_t^{\otimes 2} \rightarrow \sum_{t=0}^\infty \mathbf{U}^t \Omega (\mathbf{U}^t)'$ a.s.*

Example 6.6 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then Theorems 6.2, 6.4 and equation (3.1) imply that the sequence of matrices*

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

is relatively compact with positive definite limit points. Moreover, this series converges almost surely if $T^{-1} \sum_{t=1}^T \varepsilon_t^{\otimes 2}$ is convergent. According to Example 6.5 this is for instance the case if the additional Assumption 2.7 is satisfied.

7 The largest eigenvalue of the denominator matrix

The order of magnitude of the largest eigenvalue of the denominator matrix $\mathbf{M}_T = \sum_{s=1}^T S_{t-1}^{\otimes 2}$ can now be described. This is followed by a convergence result for the purely explosive case and a bound for the rate of convergence of sum of powers of \tilde{U}_t .

First, the largest eigenvalue of \mathbf{M}_T is considered in the following generalisation of Lai and Wei (1985, Corollary 1).

Theorem 7.1 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then*

$$\lambda_{\max}(\mathbf{M}_T) \stackrel{a.s.}{=} \begin{cases} O(T) + O(T^{2\delta-1}) & \text{if } |\text{eigen}(\mathbf{B})| < 1, \\ O(T^{2\rho} \log \log T) + O(T^{2\tilde{\delta}-1}) & \text{if } |\text{eigen}(\mathbf{B})| \leq 1, \\ O\{T^{2\rho-2} \max |\text{eigen}(\mathbf{B})|^{2T}\} & \text{for general } \mathbf{B}. \end{cases}$$

Proof of Theorem 7.1. If $\max |\text{eigen}(\mathbf{B})| < 1$ then $\mathbf{X}_t = \tilde{U}_t + \mu_u D_t$ by (3.1). Lai and Wei (1985, Corollary 1) show $\lambda_{\max}(\sum_{t=1}^T \tilde{U}_t^{\otimes 2}) = O(T)$ *a.s.* and the result then follows from Theorems 4.1, 6.4. If $\max |\text{eigen}(\mathbf{B})| \geq 1$ the result follows directly from Theorem 5.1. ■

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

Corollary 7.2 *Suppose Assumption 2.1, 2.2, 2.3 are satisfied and $\min |\text{eigen}(\mathbf{B})| > 1$ so $\mathbf{M}_T = \sum_{t=1}^T W_{t-1}^{\otimes 2}$ and recall the definition of W in Corollary 5.3. Then*

$$\mathbf{W}^{-T} \mathbf{M}_T (\mathbf{W}^{-T})' \stackrel{a.s.}{\rightarrow} \mathbf{F}_W = \sum_{t=1}^{\infty} (\mathbf{W}^{-t} W)^{\otimes 2},$$

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

$$\begin{aligned} \lim_{T \rightarrow \infty} T^{-1} \log \lambda_{\min}(\mathbf{M}_T) &\stackrel{a.s.}{=} 2 \log \min |\text{eigen}(\mathbf{W})|, \\ \lim_{T \rightarrow \infty} T^{-1} \log \lambda_{\max}(\mathbf{M}_T) &\stackrel{a.s.}{=} 2 \log \max |\text{eigen}(\mathbf{W})|. \end{aligned}$$

Proof of Corollary 7.2. Let R_T denote the difference between the matrices $\mathbf{W}^{-T} \sum_{t=1}^T W_t^{\otimes 2} (\mathbf{W}^{-T})'$ and $\mathbf{W}^{-T} \sum_{t=1}^T \tilde{W}_t^{\otimes 2} (\mathbf{W}^{-T})'$. The decomposition (3.2) shows

$$R_T = O\left(\max_{t < T} \|D_t\| \sum_{t=1}^T \|\mathbf{W}^{-T} W_{t-1}\|\right) + O\left(T \max_{t \leq T} \|D_t\|^2 \|\mathbf{W}^{-T}\|^2\right),$$

which vanishes for large T due to Theorem 4.1,*i* and Corollary 5.3,*i*. The desired result is then a direct consequence of Lai and Wei (1983b, Theorem 2). ■

While Theorem 7.1 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.

Theorem 7.3 *Suppose Assumption 2.1 is satisfied. Then, for all $\eta > 0$ and $\zeta < \gamma$*

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

Proof of Theorem 7.3. For notational convenience define $e_{U,0} = \tilde{U}_0$. Using Hölder's inequality it follows that

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

Summation over t then gives the following bound

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

Changing summation index in the double sum this can be bounded further by

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

The first two sums converge, while the third term can be decomposed as

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

The latter term is of order $O(T) = o(T^{1+\eta})$ by Assumption (2.1). The first term is a martingale. Normalised by $T^{1+\eta}$ it converges to zero *a.s.* on the set where

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

see Hall and Heyde (1980, Theorem 2.18). Minkowski's inequality shows that this sum is finite if the sum $\sum_{j=0}^{\infty} j^{-(1+\eta)} \mathbf{E}(\|e_{U,j}\|^{2+\gamma} | \mathcal{F}_{t-1})$ is finite. Assumption (2.1) ensures this is the case. ■

8 The smallest eigenvalue of the denominator matrix

Three results are given concerning the order of the inverse of the denominator matrix, $\mathbf{M}_T = \sum_{t=1}^T S_{t-1}^{\otimes 2}$, of the least square estimator in the non-explosive case. Using the techniques of Chan and Wei (1988) it can be proved that $T^{-1}\mathbf{M}_T$ is bounded from below in a weak convergence sense. The first result goes some way toward an almost sure version of this result in showing that the partial denominator matrix

$T^{-1} \sum_{t=1}^T (\mathbf{X}_t | D_t)^{\otimes 2}$ is bounded from below while the second result shows that the joint matrix $T^{-1} \mathbf{M}_T$ is bounded from below in the special case where \mathbf{S} and \mathbf{D} have no common eigenvalues. In combination these results can be used to establish the third result concerning the order of $\max_{T^\alpha \leq t \leq T} S'_t \mathbf{M}_T^{-1} S_t$ without actually establishing the order of \mathbf{M}_T^{-1} and this will suffice to prove the main theorems.

The first result concerns the partial denominator matrix $\sum_{t=1}^T (\mathbf{X}_t | D_t)^{\otimes 2}$ and is related to Lai and Wei (1985, Theorem 3).

Theorem 8.1 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied and $|\text{eigen}(\mathbf{B})| \leq 1$. Then*

$$\liminf_{T \rightarrow \infty} \lambda_{\min} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{X}_{t-1}^{\otimes 2} \right) \geq \liminf_{T \rightarrow \infty} \lambda_{\min} \left\{ \frac{1}{T} \sum_{t=1}^T (\mathbf{X}_{t-1} | D_{t-1})^{\otimes 2} \right\} \stackrel{a.s.}{>} 0.$$

To prove Theorem 8.1 the following Lemma is needed. This Lemma ensures that Lai and Wei (1982, Lemma 1) concerning the order of magnitude of normalised least squares estimators can be used.

Lemma 8.2 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then $T_0 = \inf(T : \mathbf{M}_T \text{ is invertible}) < \infty$ a.s.*

Proof of Lemma 8.2. It suffices to show that $u' \mathbf{M}_T u > 0$ for all $u \in \mathbf{R}^{\dim \mathbf{S}}$ so $u \neq 0$ and some T . Since $u' \mathbf{M}_T u = u' \sum_{t=1}^T S_{t-1}^{\otimes 2} u = u(S_0, \dots, S_{T-1})^{\otimes 2} u$ it is equivalent that $(S_0, \dots, S_{T-1})R$ spans $\mathbf{R}^{\dim \mathbf{S}}$ for some invertible matrix R .

The decomposition (3.8) shows that $\mathbf{X}_t = \tilde{\mathbf{X}}_t + \tilde{\mu} \tilde{D}_t$ where $\tilde{\mathbf{X}}_t = \mathbf{B} \tilde{\mathbf{X}}_{t-1} + e_{X,t}$ and $\tilde{D}_t = \tilde{\mathbf{D}} \tilde{D}_{t-1}$. The Cayley-Hamilton Theorem, see Herstein (1975, p.334), implies that if $\det(\lambda I_{\dim \tilde{\mathbf{D}}} - \tilde{\mathbf{D}}) = \sum_{j=0}^{\dim \tilde{\mathbf{D}}} d_j \lambda^{\dim \tilde{\mathbf{D}} - j}$ with $d_0 = 1$ is the characteristic polynomial of $\tilde{\mathbf{D}}$ then $\sum_{j=0}^{\dim \tilde{\mathbf{D}}} d_j \tilde{D}_{t-j} = \sum_{j=0}^{\dim \tilde{\mathbf{D}}} d_j \tilde{\mathbf{D}}^{\dim \tilde{\mathbf{D}} - j} \tilde{D}_{t - \dim \tilde{\mathbf{D}}} = 0$ and in particular $\sum_{j=0}^{\dim \tilde{\mathbf{D}}} d_j \mathbf{X}_{t-j} = \sum_{j=0}^{\dim \tilde{\mathbf{D}}} d_j \tilde{\mathbf{X}}_{t-j} = z_t$, say. Define

$$R = \left(\begin{array}{ccc|c} d_{\dim \tilde{\mathbf{D}}} & & & I_{\dim \tilde{\mathbf{D}}} \\ \vdots & \ddots & & \\ d_1 & & d_{\dim \tilde{\mathbf{D}}} & \\ d_0 & \ddots & & 0 \\ & \ddots & d_1 & \\ & & d_0 & \end{array} \right),$$

and partition R as a (2×2) -block matrix so the upper right block is a $\dim \tilde{\mathbf{D}}$ -dimensional square matrix. The above properties then show that $(S_1, \dots, S_T)R$ is an upper triangular (2×2) -block matrix. The lower right block is $(D_0, \dots, D_{\dim \tilde{\mathbf{D}}-1})$ which spans $\mathbf{R}^{\dim \mathbf{D}}$ by Assumption 2.3. It is left to prove that the upper left block $(z_{\dim \tilde{\mathbf{D}}}, \dots, z_{2 \dim \tilde{\mathbf{D}}-1})$ spans $\mathbf{R}^{\dim \mathbf{X}}$. The process z_t is a linear combination of the

process $(\tilde{X}'_t, \dots, \tilde{X}'_{t-\dim \tilde{\mathbf{D}}-\dim \mathbf{X}})'$ which satisfies a first order autoregression without deterministic terms. The desired result then follows from Lai and Wei (1985, Theorem 3) using Assumptions 2.1, 2.2. ■

Proof of Theorem 8.1. Let $m = \dim \mathbf{X}$. Using the model equation (2.3) and that $D_{t-1-j} = \mathbf{D}^{-1-j} D_t$ the process \mathbf{X}_t can be rewritten as

$$\mathbf{X}_t = \sum_{j=0}^{m-1} \mathbf{B}^j e_{X,t-j} + \sum_{j=0}^{m-1} \mathbf{B}^j \boldsymbol{\mu} \mathbf{D}^{-1-j} D_t + \mathbf{B}^m \mathbf{X}_{t-m},$$

where $e_{X,0} = \mathbf{X}_0$, $e_{X,-t} = -\boldsymbol{\mu} D_{t-1}$ and $\mathbf{X}_{-t} = 0$ for $t > 0$. It follows that

$$\sum_{t=0}^{T-1} (\mathbf{X}_t | D_t)^{\otimes 2} \geq \sum_{t=0}^{T-1} (\mathbf{X}_t | D_t, \mathbf{X}_{t-m})^{\otimes 2} = \sum_{t=0}^{T-1} \left(\sum_{j=0}^{m-1} \mathbf{B}^j e_{X,t-j} \middle| D_t, \mathbf{X}_{t-m} \right)^{\otimes 2}. \quad (8.1)$$

It is now argued that the lower bound in (8.1) satisfies

$$\sum_{t=0}^{T-1} \left(\sum_{j=0}^{m-1} \mathbf{B}^j e_{X,t-j} \middle| D_t, \mathbf{X}_{t-m} \right)^{\otimes 2} \stackrel{a.s.}{=} \sum_{t=0}^{T-1} \left(\sum_{j=0}^{m-1} \mathbf{B}^j e_{X,t-j} \right)^{\otimes 2} + O(\log T). \quad (8.2)$$

The norm of the difference between the left hand side and the first term on the right is bounded by

$$\left\| \left\{ \sum_{t=1}^j e_{X,t-j} \begin{pmatrix} \mathbf{X}_{t-m} \\ D_t \end{pmatrix}' + \sum_{t=1}^{T-j} e_{X,t} \begin{pmatrix} \mathbf{X}_{t+j-m} \\ D_{t+j} \end{pmatrix}' \right\} \left\{ \sum_{t=1}^{T-j} \begin{pmatrix} \mathbf{X}_{t+j-m} \\ D_{t+j} \end{pmatrix}^{\otimes 2} \right\}^{-1/2} \right\|^2.$$

The first sum is finite, so when normalised by the denominator term it is seen to be $O(1)$ due to Lemma 8.2. The normalised second term is $O\{(\log T)^{1/2}\}$ due to Lai and Wei (1982, Lemma 1) which can be used because of Lemma 8.2.

Using Lai and Wei (1982, Lemma 1) once again it follows that

$$\sum_{t=1}^T \left(\sum_{j=0}^{m-1} \mathbf{B}^j e_{X,t-j} \right)^{\otimes 2} \stackrel{a.s.}{=} \sum_{j=0}^{m-1} \sum_{t=1}^T (\mathbf{B}^j e_{X,t-j})^{\otimes 2} + O(\log T). \quad (8.3)$$

The proof is completed by combining (8.2), (8.3) and Theorem 6.2. ■

When \mathbf{B} and \mathbf{D} have no common eigenvalues Theorem 8.1 can be extended.

Theorem 8.3 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied and \mathbf{B} and \mathbf{D} have no common eigenvalues. Then $\liminf_{T \rightarrow \infty} \lambda_{\min}(T^{-1} \mathbf{M}_T) > 0$ a.s.*

Proof of Theorem 8.3. Due to the representation $\mathbf{X}_t = \tilde{\mathbf{X}}_t + \tilde{\mu}_X D_t$ given in (3.3) it suffices to show the result for sums of squares of

$$\tilde{S}_t = \begin{pmatrix} I_{\dim \mathbf{X}} & \tilde{\mu}_X \\ 0 & I_{\dim D} \end{pmatrix} S_t = \begin{pmatrix} \tilde{\mathbf{X}}_t \\ D_t \end{pmatrix}.$$

If $\det(\lambda I_{\dim \mathbf{B}} - \mathbf{B}) = \sum_{j=0}^{\dim \mathbf{B}} b_j \lambda^{\dim \mathbf{B}-j}$ is the characteristic polynomial of \mathbf{B} the Cayley-Hamilton Theorem, see Herstein (1975, p.334), implies $\sum_{j=0}^{\dim \mathbf{B}} b_j \mathbf{B}^{\dim \mathbf{B}-j} = 0$ and hence

$$y_t = \sum_{j=0}^{\dim \mathbf{B}} b_j S_{t-j} = \left\{ \begin{array}{l} \sum_{s=0}^{\dim \mathbf{B}-1} \sum_{j=0}^s b_j \mathbf{B}^{s-j} e_{X,t-s} \\ \left(\sum_{j=0}^{\dim \mathbf{B}} b_j \mathbf{D}^{\dim \mathbf{B}-j} \right) D_{t-\dim \mathbf{B}} \end{array} \right\}.$$

Since \mathbf{B} and \mathbf{D} have no common eigenvalues then $\det(\sum_{j=0}^{\dim \mathbf{B}} b_j \mathbf{D}^{\dim \mathbf{B}-j}) \neq 0$. It follows that $\liminf_{T \rightarrow \infty} \lambda_{\min}(T^{-1} \sum_{t=\dim \mathbf{D}+1}^T y_t^{\otimes 2}) > 0$ a.s. using Theorems 4.1, 6.2, 6.4 and Lai and Wei (1985, equation 3.19). The argument is finished as in proof of Lai and Wei (1985, Theorem 3). ■

The final and more technical result addresses the order of $S_t' \mathbf{M}_T^{-1} S_t$. Lai and Wei (1985, Lemma 4) show that $\max_{t \leq T} S_t' \mathbf{M}_T^{-1} S_t$ vanishes when $|\text{eigen}(\mathbf{B})| \leq 1$ and $\dim \mathbf{D} = 0$. For the subsequent analysis it suffices to take a maximum over just $T^\alpha \leq t \leq T$ for some $0 < \alpha < 1$ requiring that $0 < |\text{eigen}(B)|$, but allowing $\dim \mathbf{D} > 0$.

Theorem 8.4 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied and $0 < |\text{eigen}(\mathbf{B})| \leq 1$. Then $\max_{T^\alpha \leq t < T} S_t' \mathbf{M}_T^{-1} S_t = o(T^{-\zeta/4})$ a.s. for all α, ζ so $0 < \alpha < 1$ and $\zeta < \min\{\gamma/(2 + \gamma), 1/2\}$.*

Theorem 8.4 will be proved in a few steps following Lai and Wei (1983b). The first step is to strengthen their Lemma 3.

Lemma 8.5 *Let (a_t) be a sequence of non-negative numbers satisfying*

(i) *there exists $C > 0$ and $\kappa > 0$ such that $a_{t+1} \leq a_t + Ct^{-\kappa}$ for all large t .*

(ii) *$\sum_{t=1}^T a_t = o(T^\delta)$ for all $\delta > 0$,*

Then $a_T = o(T^{-\rho})$ for all $\rho < \min(1, \kappa/2)$.

Proof of Lemma 8.5. Condition (i) implies that for every $0 < \rho < 1$ then it holds $\min_{T > t \geq T-T^\rho} a_t \geq a_T - 2CT^{\rho-\kappa}$ for all large T . In particular, choosing ρ to satisfy $0 < \rho < \min(1, \kappa/2)$, it is seen that

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

Combining this with (ii) it follows that $a_T \leq T^{-\rho} (\sum_{t=1}^T a_t + 2C) = o(T^{\delta-\rho})$. Since δ can be chosen arbitrarily small this proves the desired result. ■

The second step is generalise Lai and Wei (1983b, Lemma 6ii).

Lemma 8.6 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied and $\max |\text{eigen}(\mathbf{B})| \leq 1$. Define T_0 as in Lemma 8.2. Then*

- (i) $S'_{t-1} \mathbf{M}_t^{-1} S_{t-1} \leq 1$ for $t \geq T_0$,
- (ii) $\sum_{t=T_0}^T S'_{t-1} \mathbf{M}_t^{-1} S_{t-1} \stackrel{a.s.}{=} O(\log T)$.

Proof of Lemma 8.6. The proof is the same as that of Lai and Wei (1983b, Lemma 6ii) using the generalisations of their Theorem 3 and Lemma 6,i presented above in Theorem 7.1 and Lemma 8.2. ■

The third step is to generalise Lai and Wei (1983b, Lemma 7).

Lemma 8.7 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied, $0 < |\text{eigen}(\mathbf{B})| \leq 1$. Then*

- (i) $\|\mathbf{M}_T^{1/2} \mathbf{S}' \mathbf{M}_{T+1}^{-1} \mathbf{S} \mathbf{M}_T^{1/2}\| \leq 1 + O\{(T^{-1} \log T)^{1/2}\}$ a.s.
- (ii) $S'_T \mathbf{M}_{T+1}^{-1} S_T \leq S'_{T-1} \mathbf{M}_T^{-1} S_{T-1} + o(T^{-\zeta/2})$ a.s. for all $\zeta < \min\{\gamma/(2+\gamma), 1/2\}$.

Proof of Lemma 8.7. (i) Using Lai and Wei (1983b, Lemma 5i) in the same way as in the proof of Lai and Wei (1983b, Lemma 7i) it holds

$$\left\| \mathbf{M}_T^{1/2} \mathbf{S}' \mathbf{M}_{T+1}^{-1} \mathbf{S} \mathbf{M}_T^{1/2} \right\| \leq (1 - \|Q_W + Q'_W\|)^{-1} \leq (1 - 2 \|Q_W\|)^{-1},$$

where

$$Q_W = \mathbf{M}_T^{-1/2} \left(\sum_{t=1}^T S_{t-1} e'_{S,t} \right) (\mathbf{S}')^{-1} \mathbf{M}_T^{-1/2}.$$

This expression can be rewritten using the identities

$$e'_{S,t} = e'_{X,t} \boldsymbol{\nu}, \quad \boldsymbol{\nu} (\mathbf{S}')^{-1} = (\mathbf{B}')^{-1} \boldsymbol{\nu}, \quad \boldsymbol{\nu} \mathbf{M}_T^{-1/2} = \left\{ \sum_{t=1}^T (\mathbf{X}_{t-1} | D_{t-1})^{\otimes 2} \right\}^{-1/2} \boldsymbol{\nu}, \quad (8.4)$$

for $\boldsymbol{\nu} = (I_{\dim \mathbf{X}}, 0)$. The desired result follows by noting that $\mathbf{M}_T^{-1/2} \sum_{t=1}^T S_{t-1} e'_{X,t}$ is $O\{(\log T)^{1/2}\}$ according to Lai and Wei (1982, Lemma 1), which can be used because of Lemma 8.2, while the term $(\mathbf{B}')^{-1} \left\{ \sum_{t=1}^T (\mathbf{X}_{t-1} | D_{t-1})^{\otimes 2} \right\}^{-1/2}$ is $O(T^{-1/2})$ by Theorem 8.1.

(ii) Noting that $S_T = \mathbf{S} S_{T-1} + e_{S,T}$ it holds

$$\mathbf{M}_{T+1}^{-1/2} S_T = \left\{ \mathbf{M}_T^{-1/2} \mathbf{S}^{-1} \mathbf{M}_{T+1} (\mathbf{S}^{-1})' \mathbf{M}_T^{-1/2} \right\}^{-1/2} \mathbf{M}_T^{-1/2} S_{T-1} + \mathbf{M}_T^{-1/2} e_{S,T}.$$

The norm of the first term is less than $(S'_{T-1} \mathbf{M}_T^{-1} S_{T-1})^{1/2} [1 + O\{(T^{-1} \log T)^{1/2}\}]^{1/2}$ by (i). This is in turn bounded by $(S'_{T-1} \mathbf{M}_T^{-1} S_{T-1})^{1/2} + O\{(T^{-1} \log T)^{1/4}\}$ due to Lemma 8.6,i. The second term equals $\boldsymbol{\nu}' \left\{ \sum_{t=1}^T (\mathbf{X}_{t-1} | D_{t-1})^{\otimes 2} \right\}^{-1/2} e_{X,T}$ according to the identities (8.4) and is seen to be $o(T^{-\xi/2})$ by Theorems 5.1, 8.1. ■

Theorem 8.4 can now be proved.

Proof of Theorem 8.4. The Lemmas 8.6,*ii* and 8.7,*ii* show that the conditions of Lemma 8.5 are satisfied for the sequence $S'_t \mathbf{M}_t^{-1} S_t$ with $\kappa = \zeta/2$ and therefore $t^{\zeta/4} S'_t \mathbf{M}_{t+1}^{-1} S_t = o(1)$ for large t . For $t > T_0$ then $\mathbf{M}_t^{-1} > \mathbf{M}_{t+1}^{-1}$ so $S'_t \mathbf{M}_T^{-1} S_t \leq S'_t \mathbf{M}_{t+1}^{-1} S_t$ and thus for all $\epsilon > 0$ and almost every outcome a T_1 exists so for all t, T so $T > t \geq T_1$ it holds $S'_t \mathbf{M}_T^{-1} S_t \leq S'_t \mathbf{M}_{t+1}^{-1} S_t < \epsilon$. This in turn implies that for all $\epsilon > 0$ and almost every outcome a T_1 exists so for all T so $T > T_1$ it holds $\max_{T_0 \leq t < T} S'_t \mathbf{M}_T^{-1} S_t \leq S'_t \mathbf{M}_{t+1}^{-1} S_t < \epsilon$ as desired. ■

9 Sample correlations

It has already been established in §6 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 $\tilde{U}_t, V_t, D_t, W_t$ are studied. A first result concerns the sample correlation of W_t and U_t, D_t .

Theorem 9.1 , *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then*

$$\begin{aligned} \left(\sum_{t=1}^T W_{t-1}^{\otimes 2} \right)^{-1/2} \left(\sum_{t=1}^T W_{t-1} D'_{t-1} \right) \left(\sum_{t=1}^T D_{t-1}^{\otimes 2} \right)^{-1/2} &\stackrel{a.s.}{=} O(T^{-1/2}), \\ \left(\sum_{t=1}^T W_{t-1}^{\otimes 2} \right)^{-1/2} \left(\sum_{t=1}^T W_{t-1} \tilde{U}'_{t-1} \right) \left(\sum_{t=1}^T \tilde{U}_{t-1}^{\otimes 2} \right)^{-1/2} &\stackrel{a.s.}{=} o(T^{-\xi/2}) \text{ for all } \xi < \frac{\gamma}{2 + \gamma}. \end{aligned}$$

The bound for the sample correlation of W_t and \tilde{U}_t should be viewed in the light of the results of Anderson (1959). He found that $Y_T = (\sum_{t=1}^T W_{t-1}^{\otimes 2})^{-1/2} \sum_{t=1}^T W_{t-1} \varepsilon_t$ is convergent when the innovations ε_t are independent, identically distributed but in general divergent. The stated result combined with Theorem 6.1 shows that the order of Y_T is at most $o\{T^{(1-\xi)/2}\}$.

Proof of Theorem 9.1. The norms of the two expression are bounded by

$$m^{1/2} \left\| \sum_{t=1}^T (\mathbf{W}^{-T} W_{t-1})^{\otimes 2} \right\|^{-1/2} \sum_{t=1}^T \|\mathbf{W}^{-T} W_{t-1}\|, \quad (9.1)$$

where m is either of

$$m_D = \max_{t < T} D'_t \left(\sum_{s=1}^T D_{s-1}^{\otimes 2} \right)^{-1} D_t, \quad m_U = \left\| \left(\sum_{t=1}^T \tilde{U}_{t-1}^{\otimes 2} \right)^{-1} \right\| \max_{t < T} \|\tilde{U}_t\|^2.$$

The last two terms of (9.1) are convergent according to Corollaries 5.3, 7.2. It holds $m_D = O(T^{-1})$ by Theorem 4.1, whereas $m_U = o(T^{-\xi})$ since Theorem 5.1 shows \tilde{U}_t^2 so $\max_{t < T} \|\tilde{U}_t\|^2$ is $o(T^{1-\xi})$, while the denominator term is $O(T^{-1})$ by Example 6.6. ■

For the sample correlation between W_t and (V'_t, D'_t) a different type of proof is needed using the results of §8. This is because the order of the smallest eigenvalue of (V'_t, D'_t) is unknown.

Theorem 9.2 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then, for all $\zeta < \min\{\gamma/(2 + \gamma), 1/2\}$, it holds*

$$\left(\sum_{t=1}^T W_{t-1}^{\otimes 2}\right)^{-1/2} \left\{ \sum_{t=1}^T W_{t-1} \begin{pmatrix} V_{t-1} \\ D_{t-1} \end{pmatrix}' \right\} \left\{ \sum_{t=1}^T \begin{pmatrix} V_{t-1} \\ D_{t-1} \end{pmatrix}^{\otimes 2} \right\}^{-1/2} \stackrel{a.s.}{=} o(T^{-\zeta/8}).$$

Proof of Theorem 9.2. For convenience define

$$S_t = \begin{pmatrix} V_t \\ D_t \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{V} & \mu_V \\ 0 & \mathbf{D} \end{pmatrix}, \quad e_{S,t} = \begin{pmatrix} e_{V,t} \\ 0 \end{pmatrix}, \quad \mathbf{M}_T = \sum_{t=1}^T S_{t-1}^{\otimes 2}. \quad (9.2)$$

Since $\sum_{t=1}^T (\mathbf{W}^{-T} W_t)^{\otimes 2}$ is convergent according to Corollary 7.2 it suffices to show $R_T = \mathbf{W}^{-T} \sum_{t=1}^T W_{t-1} S_{t-1}' \mathbf{M}_T^{-1/2}$ is $o(1)$. Follow Anderson (1959, Theorem 2.2) in writing

$$\mathbf{W}^{-T} W_{t-1} = \mathbf{W}^{-T} (W_{t-1} - \mathbf{W}^{t-T} W_{T-1}) + \mathbf{W}^{t-T} (\mathbf{W}^{-T} W_{T-1}).$$

Since $W_{t-1} - \mathbf{W}^{t-T} W_{T-1} = -\sum_{s=1}^{T-t} \mathbf{W}^{-s} e_{W,t-1+s}$ it follows, for any $0 < \alpha < 1$, that

$$\begin{aligned} R_T &= \mathbf{W}^{-T} \sum_{t=1}^{T^\alpha-1} W_{t-1} S_{t-1}' \mathbf{M}_T^{-1/2} - \mathbf{W}^{-T} \sum_{t=T^\alpha}^T \sum_{s=1}^{T-t} \mathbf{W}^{-s} e_{W,t-1+s} S_{t-1}' \mathbf{M}_T^{-1/2} \\ &\quad + \sum_{t=T^\alpha}^T \mathbf{W}^{t-T} (\mathbf{W}^{-T} W_{T-1}) S_{t-1}' \mathbf{M}_T^{-1/2}. \end{aligned} \quad (9.3)$$

The first two terms in (9.3) vanish exponentially fast. Their norm is less than

$$\|\mathbf{W}\|^{T^\alpha-T} \left(\max_{t < T^\alpha} \|\mathbf{W}^{-t} W_t\| + \max_{t \leq T} \|e_{W,t}\| \sum_{s=T^\alpha}^T \sum_{r=1}^{T-s} \|\mathbf{W}\|^{-r} \right) \max_{t < T} \|S_t\| \left\| \mathbf{M}_T^{-1/2} \right\|,$$

where $\|\mathbf{W}\|^{T^\alpha-T}$ vanishes exponentially, $\mathbf{M}_T^{-1} = O(1)$ by Lemma 8.2 and the remaining terms are of polynomial order according to Theorem 5.1, Corollary 5.3.

The final term in (9.3) is $o(T^{-\zeta/8})$. Its norm is less than

$$\left\| \mathbf{W}^{-T} W_{T-1} \right\| \left(\sum_{t=1}^T \|\mathbf{W}\|^{t-T} \right) \left(\max_{T^\alpha \leq t < T} S_t' \mathbf{M}_T^{-1} S_t \right)^{1/2},$$

where the first two components converge, see Corollary 5.3, and the last component is $o(T^{-\zeta/8})$ by Theorem 8.4. ■

Remark 9.3 *The bottleneck in the proof of Theorem 9.2 is the order of magnitude of $\max_{T^\alpha < t \leq T} S_t' \mathbf{M}_T^{-1} S_t$. By extending the weak convergence results of Chan and Wei (1988) it can be proved that this term is $O_P(T^{-1})$ when $|\text{eigen}(\mathbf{B})| \leq 1$ implying that the sample correlation between W_t and (V_t', D_t') is $O_P(T^{-1/2})$.*

Wei (1992, Theorem A.1) consider the sample correlation between U_t and V_t in the univariate case $\dim X = 1$ when D_t is absent and ε_t is a martingale difference sequence satisfying Assumptions 2.1, 2.7. That result can be generalised and strengthened by a proof resembling that of Theorem 6.4.

Theorem 9.4 *Suppose Assumptions 2.1, 2.2, 2.3 are satisfied. Then, for all $\xi < \gamma/(2 + \gamma)$,*

$$\left(\sum_{t=1}^T \tilde{U}_{t-1}^{\otimes 2} \right)^{-1/2} \left\{ \sum_{t=1}^T \tilde{U}_{t-1} \begin{pmatrix} V_{t-1} \\ D_{t-1} \end{pmatrix}' \right\} \left\{ \sum_{t=1}^T \begin{pmatrix} V_{t-1} \\ D_{t-1} \end{pmatrix}^{\otimes 2} \right\}^{-1/2} \stackrel{a.s.}{=} o(T^{-\xi/2}).$$

Proof of Theorem 9.4. Define S_t , \mathbf{S} , $e_{S,t}$, \mathbf{M}_T as in (9.2). Theorem 6.2 shows $\|(\sum_{t=1}^T \tilde{U}_t^{\otimes 2})^{-1/2}\|$ is $O(T^{-1/2})$ so it suffices to show $R_T = \sum_{t=1}^T \tilde{U}_{t-1} S'_{t-1} \mathbf{M}_T^{-1/2}$ is $o(T^{1/2})$. Inspired by the proof of Theorem 6.4 and Wei (1992, Theorem A.1) write $\tilde{U}_t = \sum_{s=1}^{T^\alpha} \mathbf{U}^{s-1} e_{U,t-s} + \mathbf{U}^{T^\alpha} \tilde{U}_{t-1-T^\alpha}$ for some $0 < \alpha < 1$ so that

$$R_T = \left(\sum_{t=1}^{T^\alpha} \tilde{U}_{t-1} S'_{t-1} + \mathbf{U}^{T^\alpha} \sum_{t=T^\alpha+1}^T \tilde{U}_{t-1-T^\alpha} S'_{t-1} + \sum_{t=T^\alpha+1}^T \sum_{s=1}^{T^\alpha} \mathbf{U}^{s-1} e_{U,t-s} S'_{t-1} \right) \mathbf{M}_T^{-1/2}. \quad (9.4)$$

The first term in (9.4) is $o\{T^{(1-\xi)/2}\}$. The norm of the sum $\sum_{t=1}^{T^\alpha} \tilde{U}_{t-1} S'_{t-1}$ is bounded by $T^\alpha \max_{t < T^\alpha} \|\tilde{U}_t\| \max_{t < T^\alpha} \|S_t\|$ which is of the desired order when α is chosen small enough and using Theorem 5.1 while $\mathbf{M}_T^{-1/2}$ is bounded due to the positive definiteness of \mathbf{M}_T stated in Lemma 8.2.

The second term in (9.4) is $o(1)$. By Cauchy-Schwartz' inequality its norm is bounded by $\|\mathbf{U}\|^{T^\alpha} (\sum_{t=0}^T \tilde{U}_t^{\otimes 2})^{1/2} (\sum_{t=T^\alpha+1}^T S'_{t-1} \mathbf{M}_T^{-1} S_{t-1})^{1/2}$ where $\|\mathbf{U}\|^{T^\alpha}$ vanishes exponentially and the other terms are $O\{(T \log T)^{1/2}\}$ due to Theorem 6.2 and Lemma 8.6, *ii*.

The third term of (9.4) is $o\{T^{(1-\xi)/2}\}$. Its norm is bounded by

$$\sum_{s=1}^{T^\alpha} \|\mathbf{U}\|^{s-1} \left\| \sum_{t=T^\alpha+1}^T e_{U,t-s} S'_{t-1} \mathbf{M}_T^{-1/2} \right\|,$$

according to the triangle inequality. Hölder's inequality implies

$$\left\| \sum_{t=T^\alpha+1}^T e_{U,t-s} S'_{t-1} \mathbf{M}_T^{-1/2} \right\| \leq \left(\sum_{t=T^\alpha+1}^T \|e_{U,t-s}\|^p \right)^{1/p} \left\{ \sum_{t=T^\alpha+1}^T (S'_{t-1} \mathbf{M}_T^{-1} S_{t-1})^{q/2} \right\}^{1/q},$$

for $2 < p < 2 + \gamma$ and $p^{-1} + q^{-1} = 1$. Since $q/2 < 1$ and $T^\alpha > T_0$ for large T then $(S'_t \mathbf{M}_T^{-1} S_t)^{q/2} \leq S'_t \mathbf{M}_T^{-1} S_t \leq S'_t \mathbf{M}_t^{-1} S_t$ so the last term is $o(T^\eta)$ for all $\eta > 0$ according to Lemma 8.6, *ii*. Due to Theorem 7.3 then $\sum_{t=T^\alpha+1}^T \|e_{U,t-s}\|^p \leq \sum_{t=1}^T \|e_{U,t}\|^p$ is $o(T^\eta)$

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

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

	$(U_t D_t)$	$(V_t D_t)$	$(W_t D_t)$
$(U_t D_t)$	1	$o(T^{-\xi/2})$	$o(T^{-\xi/2})$
$(V_t D_t)$		1	$o(T^{-\zeta/8})$
$(W_t D_t)$			1

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

uniformly in s . Overall the sum in t is therefore $o\{T^{(1-\xi)/2}\}$ uniformly in s . The desired result then follows since $\sum_{s=1}^{T^\alpha} \|\mathbf{U}\|^s$ converges. ■

The Tables 1 and 2 give an overview of the sample correlation results of Theorems 6.4, 9.1, 9.2, 9.4. All pairs of $\tilde{U}_t, V_t, W_t, D_t$ have been considered except for V_t, D_t which has non-negligible sample correlation when \mathbf{V} and \mathbf{D} have common characteristic roots. To produce these tables it is used that the marginal sample correlation, $C(x, y)$, of processes x_t, y_t relates to the joint correlation by

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

according to the formula for partitioned inversion, as well as by

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

As a consequence of the results summarised in Table 2 the condition $|\text{eigen}(\mathbf{B})| \leq 1$ can be eliminated in Theorem 8.1 concerning the lower bound for $\sum_{t=1}^T (\mathbf{X}_t|D_t)^{\otimes 2}$.

Corollary 9.5 *Suppose Assumption 2.1, 2.2, 2.3 are satisfied. Then*

$$\liminf_{T \rightarrow \infty} \lambda_{\min} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{X}_{t-1}^{\otimes 2} \right) \geq \liminf_{T \rightarrow \infty} \lambda_{\min} \left\{ \frac{1}{T} \sum_{t=1}^T (\mathbf{X}_{t-1}|D_{t-1})^{\otimes 2} \right\} \stackrel{a.s.}{>} 0.$$

Proof of Corollary 9.5. Let $R_t = (U'_t, V'_t)'$. Using a similarity transformation M as described in §3 and the results in Table 2 shows $\sum_{t=1}^T (\mathbf{X}_{t-1}|D_{t-1})^{\otimes 2}$ equals

$$(M^{-1})' \left\{ \begin{array}{cc} \sum_{t=1}^T (R_{t-1}|D_{t-1})^{\otimes 2} & 0 \\ 0 & \sum_{t=1}^T (W_{t-1}|D_{t-1})^{\otimes 2} \end{array} \right\} M^{-1} \{1 + o(1)\},$$

a.s. Apply Theorem 8.1 to the upper left block and Corollary 7.2 and Theorem 9.1 to the lower left block. ■

10 Proofs of main results

The proofs of the main results in Theorem 2.4, 2.5 and 2.8 now follows. The first of these results concerns the studentised least squares estimator.

Proof of Theorem 2.4. The process S_t is a linear combination of $R_t = (U'_t, V'_t, D'_t)'$ and W_t . As a consequence of Theorems 9.1 and 9.4 the sample correlation of R_t and W_t vanishes asymptotically, see also Table 1. The vector of interest therefore equals

$$\left\{ \sum_{t=1}^T \begin{pmatrix} R_{t-1}^{\otimes 2} & 0 \\ 0 & W_{t-1}^{\otimes 2} \end{pmatrix} \right\}^{-1/2} \sum_{t=1}^T \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} \varepsilon'_t \{1 + o(1)\},$$

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

For the explosive component note that the norm of $(\sum_{t=1}^T W_{t-1}^{\otimes 2})^{-1/2} \sum_{t=1}^T W_{t-1} \varepsilon'_t$ is bounded by

$$\left\| \sum_{t=1}^T (\mathbf{W}^{-T} W_{t-1})^{\otimes 2} \right\|^{-1/2} \left(\sum_{t=1}^T \|\mathbf{W}^{-T} W_{t-1}\| \right) \max_{t \leq T} \|\varepsilon_t\|,$$

where the first two terms are convergent because of the Corollaries 7.2, 5.3. The order of the last term is given in Theorem 5.1.

For the non-explosive part with $\max |\text{eigen}(\mathbf{B})| \leq 1$ Lai and Wei (1982, Lemma 1) together with Theorem 7.1 shows the desired result.

For $\max |\text{eigen}(\mathbf{B})| < 1$ then Lemma 6.3 combined with Theorems 4.1, 6.2 and 6.4 shows the result. ■

By combining Theorem 2.4 with results for the denominator matrix established in §7,8 the strong consistency result for the least squares estimator can now be proved.

Proof of Theorem 2.5. Consider first the partial estimator. Transforming \mathbf{X}_t into (R'_t, W'_t) with $R_t = (U'_t, V'_t)'$ using a similarity transformation M as described in §3 shows that $(\hat{A}_1, \dots, \hat{A}_k) - (A_1, \dots, A_k)$ equals

$$\sum_{t=1}^T \varepsilon_t \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} \Big| D_t \left\{ \sum_{t=1}^T \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} \Big| D_t \right\}^{\otimes 2} \Big\}^{-1} M^{-1}.$$

The sample correlation between $(R_{t-1}|D_t)$ and $(W_{t-1}|D_t)$ vanishes asymptotically so it suffices to prove the result for the two special cases where $\max |\text{eigen}(\mathbf{B})| \leq 1$ so $\dim \mathbf{W} = 0$ and where $\min |\text{eigen}(\mathbf{B})| > 1$ so $\dim \mathbf{R} = 0$. In the first case the desired order follows from Theorems 2.4, 8.1 while in the second case the statistic vanishes exponentially fast due to Theorem 2.4 and Corollary 7.2.

The second result for the full estimator when \mathbf{B} and \mathbf{D} have no common eigenvalues follows from Theorems 2.4, 8.3. ■

Proof of Theorem 2.8. Assumption 2.7 shows that $m_t = a'(\varepsilon_t^2 - \Omega)b$ for arbitrary $\dim X$ -vectors a and b . Hall and Heyde (1980, Theorem 2.18) show that if $1 \leq p \leq 2$ then $\sum_{t=1}^T m_t = o(T^{1-\zeta})$ *a.s.* on the set $\{\sum_{t=1}^{\infty} t^{p(\zeta-1)} \mathbf{E}(\|m_t\|^p | \mathcal{F}_{t-1}) < \infty\}$. This set has probability one if $p \leq 1 + \gamma/2$ and $p(\zeta - 1) < -1$ according to Assumption 2.1. These restrictions are satisfied when $\zeta < \min\{\gamma/(2 + \gamma), 1/2\}$. ■

11 Acknowledgments

Comments from S. Johansen are gratefully acknowledged.

12 References

- Anderson, T.W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676-687.
- Chan, N.H., and Wei, C.Z. (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.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987-1008.
- Fuller, W.A., Hasza, D.P. and Goebel, J.J. (1981). Estimation of the parameters of stochastic difference equations. *Annals of Statistics* 9, 531-543.
- Gradshteyn, I.S. and Ryzhik, I.M. (1965). *Table of Integrals, Series and Products*. Academic Press, New York.
- 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.
- Jeganathan, P. (1988). On the strong approximation of the distributions of estimators in linear stochastic models, I and II: stationary and explosive AR models. *Annals of Statistics* 16, 1283-1314.
- 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. (1982). 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. (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.
- Mirsky, L. (1961). *An introduction to linear algebra*. Corrected print. Oxford University Press. Reprinted 1990 by Dover, New York.
- 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 criteria for order determination in a general vector autoregression. Mimeo.
- Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.
- 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. (1992). On predictive least squares principles. *Annals of Statistics* 20, 1-42.
- Wei, C.Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *Annals of Statistics* 13, 1498-1508.