# Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination

Bent Nielsen
*Department of Economics, University of Oxford*

27 November 2008

A vector autoregression is singular when explosive characteristic roots have geometric multiplicity larger than one. The singular component is a mixingale. Martingale decompositions are constructed for sample moments involving the singular component. This permits weak and strong analysis in the case of martingale difference innovations. While least squares estimators are shown to be inconsistent in the singular case, procedures for lag length determination are shown to have the same asymptotic properties in regular and singular cases.

*Keywords*: inconsistency, lag length determination, martingale decomposition, mixingale, singular vector autoregression, triangular Toeplitz matrices.

## 1 Introduction

In general, vector autoregressions can have stationary roots, unit roots, regular and singular explosive components, as well as deterministic components. The singularity arises when explosive roots have geometric multiplicity larger than one. In some applications most components may be present. This could be the case for hyperinflationary data, see Nielsen (2005b, 2008), or when stock markets have rational bubbles, see Engsted (2006). A broader question is to what extent investigators can work with vector autoregressions without prior knowledge of the location of the roots. These issues have been addressed for regular vector autoregressions by Lai and Wei (1985) and Nielsen (2005a, 2006). The aim of the paper is to present a unified theory of least squares estimators and lag order determination methods covering both the regular and the singular case.

Several results are given here. First, the decomposition of explosive components into regular and singular components is analysed in detail using a commutation property of triangular Toeplitz matrices. This would permit a Granger-Johansen type representation of the vector autoregression. Secondly, the asymptotic properties of the singular component are analysed. It has zero conditional expectation, but it is not adapted so it is not a martingale difference, but a mixingale. Martingale decompositions are found for the sample moments involving the singular process. Thirdly, least squares estimators of singular vector autoregressions are shown to be inconsistent, since the singular explosive component and the innovation process are, in general, correlated. This formalises results indicated by Anderson (1959) and Duflo, Senoussi and Touati (1991), see also Duflo (1997). The inconsistency does, however, not arise for certain purely explosive triangular systems as found by Phillips and Magdalinos (2008). Finally, the inconsistency is shown not to affect lag order determination, which can be carried out without knowledge of the parameters. In particular the likelihood ratio test for lag length is asymptotically $\chi^2$ for regular as well as singular process. The lag order estimator found by minimising information criteria with increasing penalty is weakly consistent and a Hannan-Quinn type bound on the penalty that ensures strong consistency can be established.

The paper is organised so that §2 introduces the model and presents a decomposition of the processes into stationary, unit root, and explosive components. In §3 a further decomposition into regular and singular explosive component is made. In §4 martingale decompositions are given for sample moments involving the singular explosive components. In §5 the least squares estimators are discussed. In §6 lag order determination is discussed.

The following notation is used throughout the paper: For a matrix $\alpha$ let $\alpha^{\otimes 2} = \alpha\alpha'$. When $\alpha$ has full column rank then $\overline{\alpha} = \alpha(\alpha'\alpha)^{-1}$. When $\alpha$ is symmetric then $\lambda_{\min}(\alpha)$ and $\lambda_{\max}(\alpha)$ are the smallest and the largest eigenvalue respectively. For matrices $||\alpha|| = \{\lambda_{\max}(\alpha^{\otimes 2})\}^{1/2}$ is the spectral norm, implying that $||\alpha^{-1}|| = \{\lambda_{\min}(\alpha^{\otimes 2})\}^{-1/2}$. If $\alpha$ and $\beta$ are both semi-definite matrices then $\alpha \geq \beta$ if $\alpha - \beta$ is positive semi-definite. While $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1})$ is a conditional expectation the residual of the least squares regression of $Y_t$ on $Z_t$ is denoted $(Y_t|Z_t) = Y_t - \sum_{s=1}^T Y_s Z_s' (\sum_{s=1}^T Z_s^{\otimes 2})^{-1} Z_t$. The abbreviation $a.s.$ is used for properties holding almost surely.

## 2 The autoregressive model, its decomposition, and further notation

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

$$X_t = \sum_{j=1}^k A_j X_{t-j} + \mu D_{t-1} + \varepsilon_t, \qquad \text{for } t = 1, \ldots, T, \tag{2.1}$$

$$D_t = \mathbf{D} D_{t-1}, \tag{2.2}$$

where $D_{t-1}$ is a deterministic term and $\varepsilon_t$ an innovation term.

For the analysis of explosive processes the local Marcinkiewicz-Zygmund result of Lai and Wei (1983) is needed. This requires that the innovations are martingales with conditionally bounded moments. That is, for an increasing sequence of $\sigma$-fields, $(\mathcal{F}_t)$, let $(\varepsilon_t, \mathcal{F}_t)$ be a martingale difference sequence so $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ which satisfies the following assumptions.

**Assumption A** *For some* $\gamma > 0$ *it holds that* $\sup_t \mathsf{E}(\|\varepsilon_t\|^{2+\gamma} |\mathcal{F}_{t-1}) < \infty$ *a.s.*

**Assumption B** $\liminf_{t\to\infty} \lambda_{\min} \mathsf{E}(\varepsilon_t^{\otimes 2}|\mathcal{F}_{t-1}) > 0$ *a.s.*

When manipulating the sample moments involving the singular process a stronger assumption than B is needed.

**Assumption C** $\mathsf{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = \Omega > 0$ *a.s.*

The deterministic term $D_t$ is a vector of terms such as a constant, a linear trend, or periodic functions like seasonal dummies. This is achieved if $\mathbf{D}$ has characteristic roots on the unit circle. Moreover, $D_t$ is assumed to have linearly independent coordinates.

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

For the least squares analysis the time series can be written conveniently in companion form. First, define $\mathbf{X}_{t-1} = (X_{t-1}', \ldots, X_{t-k}')'$ with associated parameter matrices and innovations

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

Secondly, define the companion vector $S_t = (\mathbf{X}_t', D_t')'$ and

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

The companion vector $S_t$ then satisfies a first order autoregression

$$S_t = \mathbf{S} S_{t-1} + e_{S,t}. \tag{2.3}$$

Following, for instance, Nielsen (2005a, §3) the companion process $S_t$ can be decomposed into stationary, unit root and explosive processes. By a similarity transformation, see Herstein (1975, p.308), then a real, invertible matrix $M$ exists so

$$\begin{pmatrix} M & 0 \\ 0 & I_{\dim \mathbf{D}} \end{pmatrix} S_t = \begin{pmatrix} \tilde{U}_t \\ V_t \\ \tilde{W}_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 & 0 & 0 \\ 0 & \mathbf{V} & 0 & \mu_V \\ 0 & 0 & \mathbf{W} & 0 \\ 0 & 0 & 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} \tilde{U}_{t-1} \\ V_{t-1} \\ \tilde{W}_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} e_{U,t} \\ e_{V,t} \\ e_{W,t} \\ 0 \end{pmatrix}, \tag{2.4}$$

in which the absolute values of the eigenvalues of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are smaller than one, equal to one, and larger than one, respectively. The decoration on the notation $\tilde{U}_t, \tilde{W}_t$ is chosen to be consistent with the notation of Nielsen (2005a). The unit root process can be written as

$$V_t = \tilde{V}_t + \tilde{\mu}_V \tilde{D}_t \qquad \text{where} \qquad \tilde{V}_t = \mathbf{V}\tilde{V}_{t-1} + e_{V,t}, \quad \tilde{D}_t = \tilde{\mathbf{D}}\tilde{D}_{t-1}, \tag{2.5}$$

and $\tilde{\mathbf{D}}$ has dimension $\dim \tilde{\mathbf{D}} = \dim \mathbf{V} + \dim \mathbf{D}$ and the same eigenvalues as $\mathbf{D}$ with the same geometric multiplicity, but possibly larger algebraic multiplicity.

Following Duflo, Senoussi and Touati (1991) vector autoregressions are defined to be regular or singular according to the following criterion.

**Definition 1** *A vector autoregression is* ***regular*** *if all explosive roots of* $\mathbf{B}$ *have geometric multiplicity one. That is: for all* $\rho \in \mathbb{C}$ *so* $|\rho| > 1$ *and* $\det(\mathbf{B} - \rho I_{\dim \mathbf{B}}) = 0$ *then* $\operatorname{rank}(\mathbf{B} - \rho I_{\dim \mathbf{B}}) = (\dim \mathbf{B}) - 1$.
*Otherwise, the vector autoregression is* ***singular****.*

**Example 2.1** *The matrices*

$$\mathbf{B}_1 = \begin{pmatrix} \rho & 1 \\ 0 & \rho \end{pmatrix}, \qquad \mathbf{B}_2 = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$$

*both have eigenvalue* $\rho$ *with algebraic multiplicity 2. The geometric multiplicities are 1 and 2, respectively. Thus, for* $|\rho| > 1$, *the matrix* $\mathbf{B}_1$ *is associated with a regular vector autoregression, while* $\mathbf{B}_2$ *is associated with a singular vector autoregression.*

The next theorem shows that the issue of singularity cannot arise when the vector autoregression is univariate.

**Theorem 2.2** *If* $\rho$ *is a root of* $\mathbf{B}$ *then the geometric multiplicity of* $\rho$ *is at most* $p = \dim X$. *That is,* $\operatorname{rank}(\mathbf{B} - \rho I_{\dim \mathbf{B}}) \geq \dim \mathbf{B} - p$.

**Proof of Theorem 2.2.** If $\rho = 0$ note that $\mathbf{B} - \rho I_{\dim \mathbf{B}} = \mathbf{B}$. The lower $\dim \mathbf{B} - p$ rows of $\mathbf{B}$ have full row rank by construction. If $\rho \neq 0$ define the $\dim \mathbf{B} \times (\dim \mathbf{B} - p)$-matrix $N = (0, I_{\dim \mathbf{B} - p})'$. Then $N'(\mathbf{B} - \rho I_{\dim \mathbf{B}})N$ is a lower triangular matrix with all diagonal elements equal to $-\rho \neq 0$, hence, it has full rank. ∎

**Remark 2.3** *In the singular case the matrix* $\sum_{t=1}^{T} (\mathbf{W}^{-T} \tilde{W}_t)^{\otimes 2}$ *has a singular limit as pointed out by Anderson (1959). Duflo, Senoussi and Touati (1991), see also Duflo (1997, p.68, 127), characterised the situations in which the singularity arises. In the singular case the least squares estimator is inconsistent, see Theorems 5.1, 5.2 below. The possibility of singularities was overlooked by Lai and Wei (1985), so their results only apply to regular vector autoregressions. The same applies to the work of Nielsen (2005a, 2006).*

For assessing certain sample correlations the result of Lai and Wei (1982) is used in the regular case. That result is formulated for martingale differences and does not immediately carry over to the martingale approximations appearing in the singular case. Therefore a constraint on the unit root parameters is needed for some of the strong results.

**Assumption E** *If the process is singular the parameters satisfy one of conditions*
*(i)* $\mathbf{V} = 1$ *and* $\dim \mathbf{D} = 0$.
*(ii)* $\dim \mathbf{V} = 0$.

## 3 The explosive component

The explosive component $\tilde{W}_t$ is decomposed further. A regular explosive component grows at an exponential rate, while a singular component, $Z_t$ say, is identified. In §4 the singular component is shown to obey a Law of Large Numbers.

Some aspects of the explosive component have been analysed by Phillips and Magdalinos (2008). They considered a triangular system in which the regressor is a first order vector autoregression with purely explosive and diagonal first order autoregressive coefficient, where the innovations are independent and identically distributed and without deterministics.

The explosive component has decomposition

$$\tilde{W}_t = \mathbf{W}^t W - Z_t, \tag{3.1}$$

where, assuming the existence of an infinite sequence $e_{W,t}$,

$$Z_t = \sum_{j=1}^{\infty} \mathbf{W}^{-j} e_{W,t+j} \qquad \text{and} \qquad W = \tilde{W}_0 + Z_0. \tag{3.2}$$

Assuming A, B the random vector $Z_t$ is well-defined due to a Marcinkiewicz-Zygmund result and $Z_t$ has continuous distribution, see Lai and Wei (1983). The process $Z_t$ was denoted $\zeta_{t+1}$ by Phillips and Magdalinos (2008, equation 22).

In the following it is shown that in the singular case the process $\mathbf{W}^t W$ has linearly dependent coordinate processes. In addition to this property, $\mathbf{W}^t W$ is shown to have some simple properties stemming from properties of triangular Toeplitz matrices. These properties are unrelated to the location of the eigenvalue of $\mathbf{W}$, so for the discussion in the rest of this section the eigenvalues of $\mathbf{W}$ need not be explosive. In the asymptotic analysis of vector autoregressions these properties are, however, only of relevance in the explosive case.

To describe the singularity some notation is needed. For the matrix $\mathbf{W}$, which for the sake of the argument could be any real square matrix, let $m = m_r + 2m_c$ denote the number of distinct eigenvalues, so $\varphi_1 \neq \cdots \neq \varphi_{m_r} \in \mathbb{R}$ while $\varphi_{m_r+1}, \ldots, \varphi_m$ are complex pairs of the form $\psi_j \exp(\pm i\theta_j)$. Further, let $n_j$ denote the dimension of the largest Jordan block associated with $\varphi_j$ and let $n = \sum_{j=1}^m n_j$. Define vectors $\lambda_t \in \mathbb{R}^n$ as the concatenation of $m_r$ vectors of the form

$$(c_{t,n_j-1}\varphi_j^{t-n_j+1}, \ldots, c_{t,0}\varphi_j^t)', \tag{3.3}$$

where $c_{t,k} = t(t-1)\cdots(t-k+1)/(k!)$ and $m_c$ vectors of the form

$$\{c_{t,n_j-1}\psi_j^{t-n_j+1} \begin{pmatrix} \cos{(t-n_j+1)\,\theta_j} \\ \sin{(t-n_j+1)\,\theta_j} \end{pmatrix}', \ldots, c_{t,0}\psi_j^t \begin{pmatrix} \cos t\theta_j \\ \sin t\theta_j \end{pmatrix}'\}'. \tag{3.4}$$

Finally, let $J_n$ denote the $(n \times n)$-matrix with block diagonal structure where the diagonal blocks are these largest Jordan blocks. With this notation the decomposition of the explosive component can be formulated. A proof follows below.

**Theorem 3.1** *(i) Consider the process $\tilde{W}_t$ given by* (3.1)*. Let $n$ be the sum of the dimensions of the largest Jordan blocks associated with the distinct eigenvalues of $\mathbf{W}$. Then, for some $w \in \mathbb{R}^{\dim \mathbf{W} \times n}$ which is a function of the random vector $W$ and for a determistic $\lambda_t \in \mathbb{R}^n$ with components of the form* (3.3), (3.4) *it holds*

$$\tilde{W}_t = w\lambda_t - Z_t.$$

*(ii) Assuming A, B then* $\mathsf{P}(\mathrm{rank}(w) = n) = 1$.

For a regular vector autoregression then $n = \dim \mathbf{W}$ and $w$ is an invertible matrix with probability one. The process $\lambda_t$ has exponential growth whereas the component $Z_t$ is a mixingale obeying a Law of Large Numbers as shown in §4.2. The process $\tilde{W}_t$ is then explosive. The deterministic and exponentially growing process $\lambda_t$ is well-understood through the analysis of Lai and Wei (1985). In the singular case where $n < \dim \mathbf{W}$ then a matrix $w_\perp$ exists so $(w, w_\perp)$ is invertible and $w_\perp' w = 0$ so that

$$w_\perp' \tilde{W}_t = -w_\perp' Z_t.$$

In particular, the normalised sum of squares of $\tilde{W}_t$, that is $\sum_{t=1}^{T} (\mathbf{W}^{-T} \tilde{W}_t)^{\otimes 2}$, converges to a singular matrix.

The relation $w_\perp' \tilde{W}_t$ is in effect a stochastically co-explosive relation, where the co-explosive vectors $w_\perp$ are stochastic. Combing Theorem 3.1 with the Granger-Johansen representation for co-explosive processes in Nielsen (2005b) indicates a way forward for analysing processes with both co-explosive and stochastically co-explosive relations.

The proof of Theorem 3.1 hinges on a Jordan decomposition. Powers of Jordan matrices are triangular Toeplitz matrices. It is convenient to start with establishing properties of such matrices.

For vectors $a = (a_1, \ldots, a_n)'$ and $x = (x_1, \ldots, x_n)'$ introduce operators for triangular Toeplitz matrices and for reordering of vectors, that is

$$\mathrm{tt}(a) = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ & \ddots & & \vdots \\ & & \ddots & a_2 \\ & & & a_1 \end{pmatrix}, \qquad \downarrow x = \begin{pmatrix} x_n \\ \vdots \\ x_1 \end{pmatrix}.$$

**Lemma 3.2** *Let $a = (a_1, \ldots, a_n)'$ and $b = (b_1, \ldots, b_n)$.*
*(i) If $A = \mathrm{tt}(a)$, $B = \mathrm{tt}(b)$ then $AB = BA = \mathrm{tt}(\sum_{j=1}^{1} a_j b_{1+1-j}, \ldots, \sum_{j=1}^{n} a_j b_{n+1-j})$.*
*(ii) $\mathrm{tt}(a)b = \{\mathrm{tt}(\downarrow b)\}(\downarrow a)$.*
*(iii) When $J = \mathrm{tt}(\mu, 1, 0, \ldots, 0)$ then $J^t = \mathrm{tt}(\mu^t c_{t,0}, \ldots, \mu^{t-n+1} c_{t,n-1})$.*
*(iv) The results in (i)-(iii) also hold for block triangular Toeplitz matrices where the blocks $a_j$, $b_j$ and $\mu$ then have complex structure like*

$$a_j = \begin{pmatrix} a_{j1} & a_{j2} \\ -a_{j2} & a_{j1} \end{pmatrix},$$

*while 1 becomes an identity matrix.*

**Proof of Lemma 3.2.** $(i)$, $(ii)$ Write the matrices out to inspect.
$(iii)$ See for instance Varga (2000, p. 13).
$(iv)$ Use the properties

$$\begin{pmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{pmatrix}\begin{pmatrix} b_1 & b_2 \\ -b_2 & b_1 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 \\ -b_2 & b_1 \end{pmatrix}\begin{pmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{pmatrix},$$
$$\begin{pmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

■

**Lemma 3.3** *Suppose* $\mathbf{W}$ *is a real square matrix (possibly with unrestricted eigenvalues). Let* $n$ *be the sum of the dimensions of the largest Jordan blocks associated with the distinct eigenvalues of* $\mathbf{W}$. *Recall the definition of* $\lambda_t$ *and* $J_n$ *in connection with* (3.3), (3.4). *Then, for all* $W \in \mathbb{R}^{\dim \mathbf{W}}$ *there exists a* $w \in \mathbb{R}^{\dim \mathbf{W} \times n}$ *so*
$(i)$ $\mathbf{W}^t W = w\lambda_t$ *for* $t = \ldots, -1, 0, 1, \ldots$
$(ii)$ $\mathbf{W}^t w = w(J_n)^t$.

**Proof of Lemma 3.3.** $(i)$ The real Jordan decomposition, see Herstein (1975, p.308), shows that there exists an invertible real matrix $N$ so $N\mathbf{W}N^{-1} = J$ is a block diagonal matrix with blocks of the form $\mathrm{tt}(\tilde{\lambda}_1)$, where $\tilde{\lambda}_t$ is of the form (3.3) or (3.4).

By Lemma 3.2$(iii, iv)$ then $J^t$ is block diagonal, with blocks of the form $\mathrm{tt}(\downarrow \tilde{\lambda}_t)$.

Suppose $J$ has one Jordan block, so $\lambda_t = \tilde{\lambda}_t$. Then $N\mathbf{W}^t W = N\mathbf{W}^t N^{-1} N W = J^t N W$. Since $J^t = \mathrm{tt}(\downarrow \lambda_t)$ and $NW$ a vector then $J^t NW = \mathrm{tt}\{\downarrow (NW)\}\lambda_t$ by Lemma 3.2$(ii, iv)$. This implies the desired result with $Nw = \mathrm{tt}\{\downarrow (NW)\}$.

For general $J$, applying Lemma 3.2$(ii, iv)$ for each block shows that $(i)$ holds for each block. Concatenating vertically the $Nw$ matrices for blocks with the same eigenvalues gives the expression $(i)$.

If the Jordan blocks of $J$ with the same eigenvalues are clustered together then the matrix $Nw$ will have a block diagonal structure with one block for each distinct eigenvalue of $J$. Each block will be a vertical concatenation of triangular Toeplitz matrices, possibly padded with zero vectors, with dimensions comformable with the blocks of $J$.

$(ii)$ Since $N$ of $(i)$ is invertible it is equivalent to show $N\mathbf{W}^t w = Nw(J_n)^t$. Note that $N\mathbf{W}^t w = (N\mathbf{W}^t N^{-1})(Nw) = J^t(Nw)$. As outlined in $(i)$ the blocks of $J^t$ and $Nw$ have triangular Toeplitz structure, so they commute due to Lemma 3.2$(i, iv)$. Collecting the blocks of $J^t$ with the same eigenvalues then gives the desired result. For instance,

$$\begin{pmatrix} \rho & 1 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho \end{pmatrix}\begin{pmatrix} w_1 & w_2 \\ 0 & w_1 \\ 0 & w_3 \end{pmatrix} = \begin{pmatrix} w_1 & w_2 \\ 0 & w_1 \\ 0 & w_3 \end{pmatrix}\begin{pmatrix} \rho & 1 \\ 0 & \rho \end{pmatrix},$$

where the right hand side expression has the same span as $w$. ■

**Proof of Theorem 3.1.** $(i)$ Apply Lemma 3.3$(i)$ to (3.1).

$(ii)$ Due to Lai and Wei (1983, Corollary 4) for all vectors $a \in \mathbb{R}^{\dim \mathbf{W}}$ and all random variables $Y$ that are $\mathcal{F}_t$-measurable for some $t$ then $\mathsf{P}(a'W = Y) = 0$. The coordinates of the matrix $w$ are given as products $b_{ij}'W$, for some deterministic vectors $b_{ij}$. The matrix $w$ has reduced rank if some vector $c$ exists so $wc = 0$. This is not possible due to the continuity property of $W$. ∎

Phillips and Magdalinos (2008) found, in the special case described above, that $w_\perp' Z_t$ satisfies a first order autoregression (equation 20, using notation $z_{1t}$). This also holds in the general setup.

**Theorem 3.4** *The process $Z_t$ satisfies*
$(i)$ *the equation $Z_t = \mathbf{W}Z_{t-1} - e_{W,t}$,*
$(ii)$ *the triangular system*

$$\begin{pmatrix} \overline{w}' \\ w_\perp' \end{pmatrix} Z_t = \begin{pmatrix} \overline{w}'\mathbf{W}w_\perp & \overline{w}'\mathbf{W}\overline{w}_\perp \\ 0 & w_\perp'\mathbf{W}\overline{w}_\perp \end{pmatrix} \begin{pmatrix} \overline{w}' \\ w_\perp' \end{pmatrix} Z_{t-1} - \begin{pmatrix} \overline{w}' \\ w_\perp' \end{pmatrix} e_{W,t}.$$

**Proof of Theorem 3.4.** $(i)$ From (3.2) then

$$\mathbf{W}Z_{t-1} = \mathbf{W}\sum_{j=1}^{\infty} \mathbf{W}^{-j} e_{W,t-1+j} = e_{W,t} + \sum_{j=2}^{\infty} \mathbf{W}^{-(j-1)} e_{W,t-1+j} = e_{W,t} + Z_t.$$

$(ii)$ Pre-multiply the equation $Z_t = \mathbf{W}Z_{t-1} - e_{W,t}$ by $(\overline{w}, w_\perp)'$ and post-multiply $\mathbf{W}$ by the identity $I = w\overline{w}' + \overline{w}_\perp w_\perp'$ to get

$$\begin{pmatrix} \overline{w}' \\ w_\perp' \end{pmatrix} Z_t = \begin{pmatrix} \overline{w}'\mathbf{W}w_\perp & \overline{w}'\mathbf{W}\overline{w}_\perp \\ w_\perp'\mathbf{W}w & w_\perp'\mathbf{W}\overline{w}_\perp \end{pmatrix} \begin{pmatrix} \overline{w}' \\ w_\perp' \end{pmatrix} Z_{t-1} - \begin{pmatrix} \overline{w}' \\ w_\perp' \end{pmatrix} e_{W,t}.$$

Due to Lemma 3.3$(iii)$ then $\mathbf{W}w \in \mathrm{span}(w)$ so $w_\perp'\mathbf{W}w = 0$. ∎

## 4 Sample moments involving the singular process

The singular component $Z_t$ is an innovation in the sense that $\mathsf{E}(Z_t|\mathcal{F}_{t-1}) = 0$ but it is not a martingale difference as it is not $\mathcal{F}_s$-measurable for any $s$. It can be shown to be a mixingale with exponentially declining mixingale numbers. Sample moments involving the singular process are analysed directly through martingale approximations rather than by exploiting mixingale results. At first the order of magnitude of $Z_t$ is established.

### 4.1 The order of magnitude of the singular process

A result concerning the order of magnitude of the mixingale $Z_t$ is now given. At first a general result is formulated

**Theorem 4.1** *Let $(m_t, \mathcal{F}_t)$ be a martingale difference, that is $\mathsf{E}(m_t|\mathcal{F}_{t-1}) = 0$ a.s. and $m_t$ is $\mathcal{F}_t$-measurable. Suppose $\sup_t \mathsf{E}(||m_t||^\alpha|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\alpha > 1$. Let $n_t = \sum_{j=1}^\infty a_j m_{t+j}$, for constants $a_j$ so $\sum_{j=1}^\infty ||a_j|| < \infty$. Then*
*$(i)$ $\sup_t \mathsf{E}(||n_t||^\alpha|\mathcal{F}_{t-1}) < \infty$ a.s.*
*$(ii)$ $||n_T|| = \mathrm{o}(T^\zeta)$ a.s. for all $\zeta > 1/\alpha$.*

For the proof a variant of Lai and Wei (1983, Lemma 2) is needed.

**Lemma 4.2** *Let $(m_t, \mathcal{F}_t)$ be a martingale difference so $\sup_t \mathsf{E}(||m_t||^\alpha|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\alpha > 1$.*
*$(i)$ Then for every $0 < \eta < 1$ there exists positive integers $t_0$ and $K$ and a martingale difference sequence $(\tilde{m}_t, \tilde{\mathcal{F}}_t)$ so $\tilde{\mathcal{F}}_t \subset \mathcal{F}_t$ and $\mathsf{E}(\tilde{m}_t|\tilde{\mathcal{F}}_{t-1}) = 0$ satisfying, for all $t \geq t_0$,*

$$\mathsf{E}(||\tilde{m}_t||^\alpha|\breve{\mathcal{F}}_{t-1}) \leq K^\alpha \qquad a.s. \tag{4.1}$$

*and $\mathsf{P}(m_t = \tilde{m}_t$ for all $t \geq t_0) \geq 1 - \eta$.*
*$(ii)$ For constants $a_j$ so $\sum_{j=1}^\infty ||a_j|| < \infty$ define $\tilde{n}_t = \sum_{j=1}^\infty a_j \tilde{m}_{t+j}$. Then, it holds $\mathsf{E}(||\tilde{m}_t||^\alpha|\tilde{\mathcal{F}}_t) \leq (K \sum_{j=1}^\infty ||a_j||)^\alpha$ a.s.*

**Proof of Lemma 4.2.** $(i)$ Follow the proof of Lemma 2$(i)$ of Lai and Wei (1983) for the univariate case and their §4 for the multivariate case.

$(ii)$ The triangle inequality and the spectral norm inequality $||AB|| \leq ||A||||B||$ imply

$$||\tilde{n}_t|| = ||\sum_{j=1}^\infty a_j \tilde{m}_{t+j}|| \leq \sum_{j=1}^\infty ||a_j \tilde{m}_{t+j}|| \leq \sum_{j=1}^\infty ||a_j||||\tilde{m}_{t+j}||.$$

Since $\sum_{j=1}^\infty ||a_j||$ is finite then by Jensen's inequality

$$||\tilde{n}_t||^\alpha \leq (\sum_{j=1}^\infty ||a_j||)^{\alpha-1} \sum_{j=1}^\infty ||a_j||||\tilde{m}_{t+j}||^\alpha. \tag{4.2}$$

Note, that, by taking iterated expectations and using (4.1), for $j \geq 1$,

$$\mathsf{E}(||\tilde{m}_{t+j}||^\alpha|\tilde{\mathcal{F}}_t) = \mathsf{E}\{\mathsf{E}(||\tilde{m}_{t+j}||^\alpha|\tilde{\mathcal{F}}_{t+j-1})|\tilde{\mathcal{F}}_t\} \leq K^\alpha.$$

The desired results follows by combining these results. ∎

The proof of the first part of Theorem 4.1 is a variant of Lai and Wei (1983, Corollary 2). The proof of the second part is inspired by Lai and Wei (1985, Theorem 1), but uses the conditional Borel-Cantelli lemma of Chen (1978) which does not require the process to be adapted.

**Proof of Theorem 4.1.** $(i)$ Follow the proof of Corollary 2 of Lai and Wei (1983). Assume the contrary that for some $\eta > 0$

$$\mathsf{P}\{\sup_t \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) = \infty\} = 2\eta > 0. \tag{4.3}$$

For this $\eta$ there exists, due to Lemma 4.2, integers $t_0$ and $K$ and a process $\tilde{Z}_t$ and a filtration $\tilde{\mathcal{F}}_t$ with $\mathcal{F}_t \subset \tilde{\mathcal{F}}_t$ so that $\mathsf{E}(||\tilde{n}_t||^\alpha|\tilde{\mathcal{F}}_t) < K_n^\alpha$ and on a set $\breve{\Delta}$ so $\mathsf{P}(\breve{\Delta}) \geq 1-\eta$ then $n_t = \tilde{n}_t$ for $t > t_0$.

On the set $\breve{\Delta}$ it holds that

$$\sup_t \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) = \max\{\max_{t<t_0} \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t), \sup_{t\geq t_0} \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t)\}.$$

Since $n_t = \tilde{n}_t$ for $t > t_0$ then

$$\mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) = \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) = \mathsf{E}\{\mathsf{E}(||n_t||^\alpha|\tilde{\mathcal{F}}_t)|\mathcal{F}_t\} \leq K_n^\alpha \qquad \text{a.s.},$$

so that $\sup_{t\geq t_0} \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) \leq K_n^\alpha$ a.s. Moreover, for each $\omega \in \breve{\Delta}$ then the maximum over $t < t_0$ is a maximum of a finite number of elements, so it is bounded. Thus, $\mathsf{P}\{\sup_t \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) < \infty\} \geq 1-\eta$, which contradicts (4.3).

$(ii)$ As $\zeta$ is defined on an open set it suffices to show that $||n_t|| = \mathrm{O}(t^\zeta)$ $a.s.$ This holds if $\sum_{t=1}^\infty 1(||n_t|| > t^\zeta) = \sum_{t=1}^\infty 1(||n_t||^\alpha > t^{\zeta\alpha}) < \infty$ $a.s.$ The conditional Borel-Cantelli lemma of Chen (1978), which does not require $n_t$ to be $\mathcal{F}_t$-measurable shows this holds $a.s.$ on the the set where $\mathcal{I} = \sum_{t=1}^\infty \mathsf{P}(||n_t||^\alpha > t^{\zeta\alpha}|\mathcal{F}_t) < \infty$. Now, by the Markov inequality

$$\mathsf{P}(||n_t||^\alpha > t^{\zeta\alpha}|\mathcal{F}_t) \leq \frac{1}{t^{\zeta\alpha}} \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) \leq \frac{1}{t^{\zeta\alpha}} c \qquad a.s.$$

where $c = \sup_t \mathsf{E}(||n_t||^\alpha|\mathcal{F}_t) < \infty$ $a.s.$ by part $(i)$. Thus, $\mathcal{I}$ is bounded by $c\sum_{t=1}^\infty t^{-\zeta\alpha}$, which is finite when $\zeta\alpha > 1$. ■

Turning to the process $Z_t$ a consequence of Theorem 4.1 is as follows. Using that for $2\zeta = 1-\xi$ and $\zeta > (2+\gamma)^{-1}$ then $\xi < \gamma/(2+\gamma)$.

**Corollary 4.3** *Assuming A then* $\sup_t \mathsf{E}(||Z_t||^{2+\gamma}|\mathcal{F}_t) < \infty$ *and* $||Z_t|| = \mathrm{o}\{t^{(1-\xi)/2}\}$ *a.s. for all* $\xi < \gamma/(2+\gamma)$.

## 4.2 Martingale decompositions

Martingale decompositions are established for $\sum_{t=1}^T Z_t$, $\sum_{t=1}^T Z_t U_t$, $\sum_{t=1}^T (Z_t^{\otimes 2} - \mathsf{E}Z_t^{\otimes 2})$. For $Z_t$ this follows by manipulating the autoregressive equation of Theorem 3.4$(i)$. For $Z_t U_t$ the argument involves the conditional Borel Cantelli Theorem of Chen (1978) which does not require the process to be adapted. For $Z_t^{\otimes 2} - \mathsf{E}Z_t^{\otimes 2}$ those arguments are combined with a Beveridge-Nelson-type decomposition for variances of linear processes as exploited in Phillips and Solo (1992) although forward in time.

An alternative approach would be to show that $Z_t$, $Z_t U_t$, $Z_t^{\otimes 2} - \mathsf{E}Z_t^{\otimes 2}$ are mixingales with exponentially declining mixingale numbers. Mixingale limit results, which are proved through martingale decompositions, could then be used. That approach has two drawbacks: A further assumption that $\sup_t \mathsf{E}||\varepsilon_t||^{2+\delta}$ is bounded seems needed, and yet only some of the desired strong limit results are available.

**Theorem 4.4** *Assuming A then, for all* $\xi < \gamma/(2+\gamma)$,

$$(\mathbf{W}-I)\sum_{t=1}^{T} Z_t = \sum_{t=1}^{T} e_{W,t+1} + \mathrm{o}\{T^{(1-\xi)/2}\} \qquad a.s.$$

**Proof of Theorem 4.4.** Reorganising the equation in Theorem 3.4$(i)$ gives $(\mathbf{W}-I)Z_t = \Delta Z_{t+1} + e_{W,t+1}$. Cumulating yields $(\mathbf{W}-I)\sum_{t=1}^{T} Z_t = (\sum_{t=1}^{T} e_{W,t+1} + Z_{T+1} - Z_1)$. By Corollary 4.3 then $Z_t$ is $\mathrm{o}\{T^{(1-\xi)/2}\}$ $a.s.$ ∎

For sample covariances involving $Z_t$ define

$$Y_t = \begin{pmatrix} \mathbf{W}Z_{t-1} \\ \varepsilon_t \end{pmatrix}, \qquad \mathbf{Y} = \begin{pmatrix} \mathbf{W}^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \qquad e_{Y,t} = \begin{pmatrix} e_{W,t} \\ \varepsilon_t \end{pmatrix}, \tag{4.4}$$

so $Y_t = \sum_{j=0}^{\infty} \mathbf{Y}^j e_{Y,t+j}$. Assuming C the expectation of $Y_t^{\otimes 2}$ is

$$\Omega_{YY} = \mathsf{E}Y_t^{\otimes 2} = \begin{pmatrix} \mathbf{W}\Omega_{ZZ}\mathbf{W}' & \mathbf{W}\Omega_{Z\varepsilon} \\ \Omega_{\varepsilon Z}\mathbf{W}' & \Omega \end{pmatrix}, \tag{4.5}$$

where

$$\Omega_{ZZ} = \mathsf{E}Z_t^{\otimes 2} = \mathsf{E}(\sum_{j=1}^{\infty} \mathbf{W}^{-j} e_{W,t+j})^{\otimes 2} = \sum_{j=1}^{\infty} \mathbf{W}^{-j}\mathsf{E}(e_{W,t})(\mathbf{W}^{-j})', \tag{4.6}$$

$$\Omega_{Z\varepsilon} = \mathsf{E}Z_t e'_{S,t+1} = \mathsf{E}(\sum_{j=1}^{\infty} \mathbf{W}^{-j} e_{W,t+j})\varepsilon'_{t+1} = \mathbf{W}^{-1}\mathsf{Cov}(e_{W,t},\varepsilon_t). \tag{4.7}$$

Here, $\Omega_Y$, $\Omega_{ZZ}$ are well defined and positive definite by the argument of Lai and Wei (1985, Example 3), while $\Omega_{Z\varepsilon}$ is non-zero as $e_{W,t}$ is a function of $\varepsilon_t$.

**Theorem 4.5** *Let* $R_{t-1} = (\varepsilon'_{t-1}, U'_{t-1}, V'_{t-1}N'_{V,t}, D'_{t-1}N'_{D,t})$. *Assuming A, D then, for all* $\xi < \gamma/(2+\gamma)$,

$$\sum_{t=1}^{T} Y_t R'_{t-1} = \sum_{t=1}^{T}\sum_{j=0}^{t-1} \mathbf{Y}^j e_{Y,t} R'_{t-1-j} + \mathrm{o}(T^{1-\xi}) \qquad a.s.$$

**Proof of Theorem 4.5.** Write $Y_t R'_{t-1} = \sum_{j=0}^{\infty} \mathbf{Y}^j e_{Y,t+j} R'_{t-1}$. Split the sum in two sums, of which the first sums to $T-t$ and the second from $T-t+1$. This yields

$$\mathcal{I}_T = \sum_{t=1}^{T} Y_t R'_{t-1} = (\sum_{t=1}^{T}\sum_{j=0}^{T-t} + \sum_{t=1}^{T}\sum_{j=T-t+1}^{\infty})\mathbf{Y}^j e_{Y,t+j} R'_{t-1} = \mathcal{I}_{1,T} + \mathcal{I}_{2,T}.$$

For $\mathcal{I}_{1,T}$ rearrange using $s = j+t$ to get the leading term.
To prove $\mathcal{I}_{2,T} = \mathrm{o}(T^{\zeta})$ $a.s.$ with $2\zeta = 1-\xi$ write, with $s = j-T+t$,

$$\mathcal{I}_{2,T} = \sum_{t=1}^{T}\sum_{j=T-t+1}^{\infty} \mathbf{Y}^j e_{Y,t+j} R'_{t-1} = \sum_{t=1}^{T} \mathbf{Y}^{T-t} \sum_{s=1}^{\infty} \mathbf{Y}^s e_{Y,T+s} R'_{t-1}.$$

As $\zeta$ is defined on an open set it suffices to show that $||\mathcal{I}_{2,T}|| = \mathrm{O}(T^{\zeta})$ $a.s.$ This holds if $\sum_{T=1}^{\infty} 1(||\mathcal{I}_{2,T}|| > T^{\zeta}) = \sum_{t=1}^{\infty} 1(||\mathcal{I}_{2,T}||^{2\alpha} > T^{2\alpha\zeta}) < \infty$ $a.s.$ By the conditional Borel-Cantelli lemma of Chen (1978) this holds $a.s.$ on the the set where $\sum_{t=1}^{\infty} \mathsf{P}(||\mathcal{I}_{2,T}||^{2\alpha} > T^{2\alpha\zeta}|\mathcal{F}_T) < \infty$.

Now, by the Markov inequality

$$\mathsf{P}(||\mathcal{I}_{2,T}||^{2\alpha} > T^{2\alpha\zeta}|\mathcal{F}_T) \leq \frac{1}{T^{2\alpha\zeta}} \mathsf{E}(||\mathcal{I}_{2,T}||^{2\alpha}|\mathcal{F}_T). \tag{4.8}$$

It will be desired that $2\alpha < 2 + \gamma$. However, the expectation $\mathsf{E}(||\mathcal{I}_{2,T}||^{2\alpha})$ may be undefined if $2\alpha \geq 2 + \gamma$. In that case apply the truncation argument in the proof of Lai and Wei (1982, Lemma 2): Choose constants $a_t$ so $\mathsf{P}(||R_t||^{2\alpha} > a_t) < t^{-2}$. By the Borel-Cantelli Lemma, see Breiman (1968, p.41), then $\mathsf{P}(R_t = R_t^* \text{ for large } t) = 1$ where $R_t^* = R_t$ if $||R_t||^{2\alpha} < a_t$ and zero otherwise. To bound $\mathsf{E}(||\mathcal{I}_{2,T}||^{2\alpha}|\mathcal{F}_T)$ apply the inequality (4.2) and note $R_{t-1}$ is $\mathcal{F}_T$-measurable so

$$\mathsf{E}(||\mathcal{I}_{2,T}||^{2\alpha}|\mathcal{F}_T) \leq c_1^{2\alpha-1} \sum_{t=1}^{T} ||\mathbf{Y}||^{T-t} \sum_{s=1}^{\infty} ||\mathbf{Y}||^{s} ||R'_{t-1}||^{2\alpha} \mathsf{E}(||e_{Y,T+s}||^{2\alpha}|\mathcal{F}_T).$$

By Assumption A then $\sup_t \mathsf{E}(||e_{Y,T+s}||^{2\alpha}|\mathcal{F}_T) < \infty$ $a.s.$ for $2\alpha < 2+\gamma$, while Nielsen (2005a, Theorems 4.1, 5.1) assuming A, D imply that $\max_{t\leq T} ||R'_{t-1}||^{2\alpha} = \mathrm{o}(T^{\alpha(1-\varphi)})$ for all $\varphi < \gamma/(2+\gamma)$. Thus, for large $T$ and all $c_2 > 0$ then $\mathsf{P}(||\mathcal{I}_{2,T}||^{2\alpha} > T^{2\alpha\zeta}|\mathcal{F}_T) \leq c_2 T^{\alpha(1-\varphi)-2\alpha\zeta}$, so it is necessary that $\alpha(1-\varphi) - 2\alpha\zeta < -1$. This condition along with $2\alpha < 2+\gamma$ and $\varphi < \gamma/(2+\gamma)$ implies the desired bound for $\xi$. ∎

**Theorem 4.6** *Assuming A, C then, for all* $\xi < \gamma/(2+\gamma)$,

$$\sum_{t=1}^{T} Y_t^{\otimes 2} = \sum_{j=0}^{\infty} \mathbf{Y}^j (\sum_{t=1}^{T} e_{Y,t}^{\otimes 2})(\mathbf{Y}^j)' + \sum_{t=1}^{T} (m_t + m_t') + \mathrm{o}(T^{1-\xi}) \qquad a.s.,$$

*where* $m_t = \sum_{s=1}^{t-1} \sum_{\ell=1}^{s} \mathbf{Y}^{s-\ell} e_{Y,t-\ell} e'_{Y,t} (\mathbf{Y}^s)'$.

**Proof of Theorem 4.6.** *First*, decompose $Y_t^{\otimes 2} = (\sum_{j=0}^{\infty} \mathbf{Y}^j e_{Y,t+j})^{\otimes 2} = \mathcal{I}_{1,t} + \mathcal{I}_{2,t} + \mathcal{I}'_{2,t}$, where

$$\mathcal{I}_{1,t} = \sum_{j=0}^{\infty} \mathbf{Y}^j e_{Y,t+j}^{\otimes 2} (\mathbf{Y}^j)', \qquad \mathcal{I}_{2,t} = \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \mathbf{Y}^j e_{Y,t+j} e'_{Y,t+j+r} (\mathbf{Y}^{j+r})'$$

*Secondly*, write $\mathcal{I}_{1,t} = \mathcal{I}_{11,t} + \mathcal{I}_{12,t}$, where

$$\mathcal{I}_{11,t} = \sum_{j=0}^{\infty} \mathbf{Y}^j e_{Y,t}^{\otimes 2} (\mathbf{Y}^j)', \qquad \mathcal{I}_{12,t} = \sum_{j=1}^{\infty} \mathbf{Y}^j (e_{Y,t+j}^{\otimes 2} - e_{Y,t}^{\otimes 2})(\mathbf{Y}^j)'.$$

Here $\sum_{t=1}^{T} \mathcal{I}_{11,t}$ is the leading component. The term $\mathcal{I}_{12,t}$ can be written as

$$\mathcal{I}_{12,t} = \sum_{j=1}^{\infty} \mathbf{Y}^j \sum_{\ell=1}^{j} \Delta(e_{Y,t+\ell}^{\otimes 2} - \Omega_Y)(\mathbf{Y}^j)' = \Delta y_{12,t},$$

assuming C, where

$$y_{12,t}=\sum_{j=1}^{\infty}\mathbf{Y}^{j}\sum_{\ell=1}^{j}(e_{Y,t+\ell}^{\otimes 2}-\Omega_{Y})(\mathbf{Y}^{j})'=\sum_{\ell=1}^{\infty}\sum_{j=\ell}^{\infty}\mathbf{Y}^{j}(e_{Y,t+\ell}^{\otimes 2}-\Omega_{Y})(\mathbf{Y}^{j})'.$$

It follows that $\sum_{t=1}^{T}\mathcal{I}_{12,t}=y_{12,T}-y_{12,0}$. Assuming A, C then $m_{t+\ell}=e_{Y,t+\ell}^{\otimes 2}-\Omega_{Y}$ is a martingale difference satisfying $\sup_t \mathsf{E}(||m_t||^{1+\gamma/2}|\mathcal{F}_{t-1})<\infty$ *a.s.*, while $\mathbf{Y}^j$ has geometric decay. Theorem 4.1($ii$) then shows $y_{12,T}=\mathrm{o}(T^{1-\xi})$ *a.s.* for all $\xi<\gamma/(2+\gamma)$.

*Thirdly*, rewrite $\mathcal{I}_{2,t}$ using $s=j+r$

$$\mathcal{I}_{2,t}=\sum_{s=1}^{\infty}\sum_{j=0}^{s-1}\mathbf{Y}^{j}e_{Y,t+j}e_{Y,t+s}'(\mathbf{Y}^{s})'.$$

Split the sum in two

$$\mathcal{I}_{2,t}=(\sum_{s=1}^{T-t}\sum_{j=0}^{s-1}+\sum_{s=T-t+1}^{\infty}\sum_{j=0}^{s-1})\mathbf{Y}^{j}e_{Y,t+j}e_{Y,t+s}'(\mathbf{Y}^{s})'=\mathcal{I}_{21,t}+\mathcal{I}_{22,t}.$$

Here $\sum_{t=1}^{T}\mathcal{I}_{21,t}$ is the leading component. Rearrange the sum using $u=t+s$ and $\ell=s-j$ to get $\sum_{t=1}^{T}\mathcal{I}_{21,t}=\sum_{u=1}^{T}m_u$.

Further, follow the argument concerning the term $\mathcal{I}_{2,T}$ in the proof of Theorem 4.5 to see that $\sum_{t=1}^{T}\mathcal{I}_{22,t}=\mathrm{o}(T^{1-\xi})$. ■

### 4.3 Sample moments

The sample cross correlations of $\tilde{U}_{t-1},(\tilde{V}_{t-1}',\tilde{D}_{t-1}'),\lambda_{t-1},(Z_{t-1}',\varepsilon_t')$ turn out to vanish. Those not involving $Z_t$ have been studied in Nielsen (2005a). Those involving $Z_t$ are new. For convenience define, for instance, the sample correlation of $Z_{t-1}$ and $\tilde{U}_{t-1}$ as

$$\mathsf{c}_{zu}=\widehat{\mathsf{Cor}}(Z_{t-1},\tilde{U}_{t-1})=(\sum_{t=1}^{T}Z_{t-1}^{\otimes 2})^{-1/2}(\sum_{t=1}^{T}Z_{t-1}\tilde{U}_{t-1}')(\sum_{t=1}^{T}\tilde{U}_{t-1}^{\otimes 2})^{-1/2}. \tag{4.9}$$

Let $\mathsf{c}_{z(vd)},\mathsf{c}_{z\lambda}$ denote sample correlations of $Z_{t-1}$ with $(\tilde{V}_{t-1}',\tilde{D}_{t-1}')'$ and $\lambda_{t-1}$, respectively. Further, recall $\Omega_{YY}$ defined in (4.5), which is positive definite.

**Theorem 4.7** *Assuming A, C with* $\gamma>1$ *then*
($i$) $T^{-1}\sum_{t=1}^{T}Y_{t-1}^{\otimes 2}\to\Omega_{YY}$ *a.s.*
($ii$) $\mathsf{c}_{zu}=\mathrm{o}(T^{-\varphi})$ *a.s. for all* $\varphi<\{\gamma+\min(0,\gamma-2)\}/\{2(2+\gamma)\}$.
($iii,a$) $\mathsf{c}_{z(vd)}=\mathrm{o}(T^{-\psi})$ *a.s. for all* $\psi<\min\{\gamma/(2+\gamma),1/2)\}$ *assuming D, E.*
($iii,b$) $\mathsf{c}_{z(vd)}=\mathrm{o}_{\mathsf{P}}(T^{-\psi})$ *for all* $\psi<\min\{\gamma/(2+\gamma),1/2)\}$ *assuming D.*
($iv$) $\mathsf{c}_{z\lambda}=\mathrm{o}(T^{-\xi/2})$ *a.s. for all* $\xi<\gamma/(2+\gamma)$.

The proof of Theorem 4.7 exploits the martingale decompositions in Theorems 4.5, 4.6. The order of the martingales found in those theorems has to be established. Thus, in relation to Theorem 4.5 define the martingale differences

$$m_{\varepsilon U,t}=\sum_{j=0}^{t-1}\mathbf{Y}^{j}e_{Y,t}(\varepsilon_{t-1-j}',U_{t-1-j}'),\qquad m_{VD,t}=\sum_{j=0}^{t-1}\mathbf{Y}^{j}e_{Y,t}V_{t-1-j}'N_{V,t}',$$

and recall the martingale difference $m_t$ in Theorem 4.6.

**Lemma 4.8** *Assuming A, C then*
*(i)* $\sum_{t=1}^{T} m_{\varepsilon U,t} = \mathrm{o}(T^{1-\varphi})$ *a.s. for all* $\varphi < \{\gamma + \min(0, \gamma - 2)\}/\{2(2+\gamma)\}$.
*(ii)* $\sum_{t=1}^{T} m_{VD,t} = \mathrm{o}(T^{1-\psi})$ *a.s. for all* $\psi < \min\{\gamma/(2+\gamma), 1/2)\}$.
*(iii)* $\sum_{t=1}^{T} m_t = \mathrm{o}(T^{1-\varphi})$ *for all* $\varphi < \{\gamma + \min(0, \gamma - 2)\}/\{2(2+\gamma)\}$.

**Proof of Lemma 4.8.** $(i)$ a law of large numbers for $m_{\varepsilon U,t}$ has to be established. Since $Q_{t-1-j} = (\varepsilon'_{t-1-j}, U'_{t-1-j})$ are $\mathcal{F}_{t-1}$-measurable then $m_{\varepsilon U,t}$ is a martingale difference sequence. By Chow (1965, Theorem 5) then $\sum_{t=1}^{T} m_{\varepsilon U,t} = \mathrm{o}(T^{1-\varphi})$ on the set where $\sum_{t=1}^{T} \mathsf{E}(||t^{\varphi-1} m_{\varepsilon U,t}||^{\alpha}|\mathcal{F}_{t-1}) < \infty$ for some $1 \leq \alpha \leq 2$. By (4.2) then

$$||m_{\varepsilon U,t}||^{\alpha} \leq (\sum_{j=0}^{t-1} ||\mathbf{Y}||^{j})^{\alpha-1} \sum_{j=0}^{t-1} ||\mathbf{Y}||^{j} ||Q_{t-1-j}||^{\alpha} ||e_{Y,t}||^{\alpha}.$$

For an $\alpha$ so $\alpha < 1 + \gamma/2$ and $\alpha \leq 2$ consider $\mathcal{E}_t = \mathsf{E}(||t^{(\xi-1)/2} Q_{t-1-j}||^{\alpha} ||e'_{Y,t}||^{\alpha}|\mathcal{F}_{t-1})$. If the unconditional moment does not exist truncate as in the argument concerning the term $\mathcal{I}_{2,T}$ in the proof of Theorem 4.5. Assuming A then $\mathcal{E}_t \leq ||Q_{t-1-j}||^{\alpha}$ *a.s.*, uniformly in $t$. This bound is $\mathrm{o}\{t^{(1-\xi)/2}\}$ *a.s.* for all $\xi < \gamma/(2+\gamma)$, uniformly in $t$, see Lai and Wei (1985, Theorem 1) or Nielsen (2005a, Theorem 5.1). Hence, it has to hold that $\{1 - \varphi + (\xi - 1)/2\}\alpha > 1$ with the above constraints to $\alpha$.

$(ii)$ Follow the argument of $(i)$ noting that $N_{V,t} V_t = \mathrm{O}\{(\log\log t)^{1/2}\}$, see Lai and Wei (1985, Theorem 1) or Nielsen (2005a, Theorem 5.1), while $N_{D,t} D_t = \mathrm{O}(1)$, see Nielsen (2005a, Theorem 4.1). Hence, it has to hold that $(1-\psi)\alpha > 1$ with the above constraints to $\alpha$.

$(iii)$ Same argument as in $(i)$. ■

**Proof of Theorem 4.7.** $(i)$ Apply Theorem 4.6, assuming A, C. By Lemma 4.8$(i)$ with $\gamma > 1$ then the martingale terms $\sum_{t=1}^{T} m_t$ vanish. Thus, a Law of Large Numbers, see Lai and Wei (1985, Theorem 2, Example 3) or Nielsen (2005a, Theorem 6.1) gives the desired result.

$(ii)$ For the numerator apply Theorem 4.5 and Lemma 4.8$(i)$. By Lai and Wei (1985, Theorem 2, Example 3) assuming A, C, see also Nielsen (2005a, Theorem 6.2), $||(\sum_{t=1}^{T} \tilde{U}_{t-1}^{\otimes 2})^{-1}|| = \mathrm{O}(T^{-1})$ *a.s.*

$(iii, a)$ For the numerator apply Theorem 4.5 and Lemma 4.8$(ii)$. For the denominator two cases are covered.

*First*, assume E$(i)$. The Donsker and Varadhan's (1977) Law of the Iterated Logarithm for the integrated squared Brownian motion states

$$\liminf_{T\to\infty} \frac{\log\log T}{T^2} \int_0^T B_u^2 du \overset{a.s.}{=} \frac{1}{4}.$$

Thus, with $N_{V,T} = T^{-1/2}$, it holds $||\{\sum_{t=1}^{T} (N_{V,T} \tilde{V}_{t-1})^{\otimes 2}\}^{-1}|| = \mathrm{O}(T^{-1} \log\log T)$ *a.s.*

*Secondly,* assume E$(ii)$. By Nielsen (2005a, Theorem 4.1) assuming D it holds $||\{\sum_{t=1}^{T} (N_{D,T} \tilde{D}_{t-1})^{\otimes 2}\}^{-1}|| = \mathrm{O}(T^{-1})$.

$(iii,b)$ Following arguments as in Chan and Wei (1988), assuming D, it can be proved that the weak limit of $T^{-1}\sum_{t=1}^{T} R_{t-1}^{\otimes 2}$, where $R_{t-1} = (\tilde{V}_{t-1}' N_{V,T}', \tilde{D}_{t-1}' N_{D,T}')'$, is positive definite. Combine this with the arguments in $(iii,a)$.

$(iv)$ As in Nielsen (2005a, Theorem 9.1) use that

$$||\mathsf{c}_{z\lambda}|| \leq \{\max_{t\leq T} Z_{t-1}'(\sum_{t=1}^{T} Z_{t-1}^{\otimes 2})^{-1} Z_{t-1}\}^{1/2} (\sum_{t=1}^{T} ||\lambda_{t-1}||) ||(\sum_{t=1}^{T} \lambda_{t-1}^{\otimes 2})^{-1/2}||.$$

The terms involving $\lambda_{t-1}$ are convergent $a.s.$ due to Nielsen (2005a, Corollary 5.3, 7.2). By Corollary 4.3 then $Z_t = \mathrm{o}\{T^{(1-\xi)/2}\}$ $a.s.$ for all $\xi < \gamma/(2+\gamma)$. Since $||(\sum_{t=1}^{T} Z_{t-1}^{\otimes 2})^{-1}|| = \mathrm{O}(T^{-1})$ $a.s.$ by $(i)$ the desired order follows. ■

For easy reference the cross correlations of $\varepsilon_t, \tilde{U}_{t-1}, (\tilde{V}_{t-1}', \tilde{D}_t')', \lambda_{t-1}$ analysed in Nielsen (2005a, Theorems 2.4, 9.1, 9.2, 9.4), assuming A, C, D, are stated here. It holds for all $\xi < \gamma/(2+\gamma)$ and all $\zeta < \min\{2\gamma/(2+\gamma), 1\}$ that

$$\begin{aligned} \mathsf{c}_{\varepsilon u}^2, \mathsf{c}_{\varepsilon d}^2 &\overset{a.s.}{=} \mathrm{O}(\log\log T), \qquad \mathsf{c}_{\varepsilon(vd)}^2 \overset{a.s.}{=} \mathrm{O}(\log T) \\ \mathsf{c}_{\varepsilon\lambda}^2, \mathsf{c}_{u(vd)}^2, \mathsf{c}_{u\lambda}^2 &\overset{a.s.}{=} \mathrm{O}(\max_{1\leq t\leq T} \|\varepsilon_t\|^2) = \mathrm{o}(T^{1-\xi}) \\ \mathsf{c}_{\lambda d} &\overset{a.s.}{=} \mathrm{O}(T^{-1/2}), \qquad \mathsf{c}_{\lambda(vd)} \overset{a.s.}{=} \mathrm{o}(T^{-\zeta/4}), \qquad \mathsf{c}_{vd} \overset{a.s.}{=} \mathrm{O}(1). \end{aligned} \tag{4.10}$$

**Theorem 4.9** *Assuming A, C, D, E with $\gamma > 1$ then*

$$\liminf_{T\to\infty} \lambda_{\min}(T^{-1}\sum_{t=1}^{T} \mathbf{X}_{t-1}^{\otimes 2}) \geq \liminf_{T\to\infty} \lambda_{\min}\{T^{-1}\sum_{t=1}^{T} (\mathbf{X}_{t-1}|D_t)^{\otimes 2}\} > 0 \qquad a.s.$$

**Proof of Theorem 4.9.** Partitioned inversion gives the inequality. The regular case is covered in Nielsen (2005a, Corollary 9.5) assuming A, C, D. By Theorem 4.7 assuming A, C, D, E with $\gamma > 1$, the singular and regular component are asymptotically uncorrelated while $T^{-1}\sum_{t=1}^{T} Z_{t-1}^{\otimes 2}$ has a positive definite limit. ■

## 5 Consistency properties of the least squares estimator

The least squares estimator for the companion matrix $\mathbf{S}$ and the covariance matrix $\Omega$ are shown to be inconsistent for singular explosive processes. The inconsistency arises from the correlation of the processes $Z_{t-1}$ and $\varepsilon_t$. This issue is avoided in the triangular system of Phillips and Magdalinos (2008) due to an independence assumption.

Two results are given using weak and strong convergence, respectively. Let $n$ be the sum of the dimensions of the largest Jordan blocks associated with the distinct eigenvalues of $\mathbf{W}$ and define dimensions $s = \dim \mathbf{S}$, $y = \dim \mathbf{U} + \dim \mathbf{V} + n$, $d = \dim \mathbf{D}$, matrices $(\Omega_{ZZ}, \Omega_{\varepsilon Z}, \Omega_{SZ}) = \mathsf{Cov}\{(Z_{t-1}, \varepsilon_t, e_{S,t}), Z_{t-1}\}$, and random matrices

$$\begin{aligned} \tilde{\Omega} &= \Omega - \Omega_{\varepsilon Z} w_\perp (w_\perp' \Omega_{ZZ} w_\perp)^{-1} w_\perp' \Omega_{Z\varepsilon}, \\ \tilde{\mathbf{S}} &= \mathbf{S} + \{0_{s\times y}, \Omega_{SZ} w_\perp (w_\perp' \Omega_{ZZ} w_\perp)^{-1} w_\perp' M\}, \\ \tilde{\mathbf{S}}_{norm} &= \{0_{s\times y}, T^{1/2} \Omega_{SZ} w_\perp (w_\perp' \Omega_{ZZ} w_\perp)^{-1/2}, 0_{s\times d}\}. \end{aligned}$$

**Theorem 5.1** *Assuming $A$, $C$, $D$ with $\gamma > 1$ then*
(*i*) $\hat{\Omega} \xrightarrow{\mathsf{P}} \tilde{\Omega}$.
(*ii*) $(\hat{\mathbf{S}} - \mathbf{S})(\sum_{t=1}^{T} S_{t-1}^{\otimes 2})^{1/2} = \tilde{\mathbf{S}}_{norm}\{1 + \mathrm{o}_{\mathsf{P}}(1)\} + \mathrm{o}_{\mathsf{P}}(1)$.
(*iii*) $\hat{\mathbf{S}} \xrightarrow{\mathsf{P}} \tilde{\mathbf{S}}$.
(*iv*) $\mathsf{P}(rank(\widetilde{\Omega}) = \dim \widetilde{\Omega}) = 1$.
(*v*) *If* $\dim \mathbf{W} > n$ *then the matrix* $\tilde{\mathbf{S}}$ *satisfies* $\mathsf{P}(\tilde{\mathbf{S}} = 0) = 0$.
(*vi*) *If* $\dim \mathbf{D} = 0$ *then the eigenvalues of* $\hat{\mathbf{U}}$ *and* $\hat{\mathbf{V}}$ *are consistent and* $n$ *of the eigenvalues of* $\hat{\mathbf{W}}$ *are consistent, namely those of the largest Jordan blocks associated with each distinct eigenvalue. Thus* $\hat{\mathbf{S}}$ *has* $y = \dim \mathbf{U} + \dim \mathbf{V} + n$ *consistent eigenvalues and* $\dim \mathbf{W} - n$ *inconsistent eigenvalues.*

**Proof of Theorem 5.1.** (*i*) By the companion equation (2.3) then

$$T\hat{\Omega} = \sum_{t=1}^{T} \varepsilon_t^{\otimes 2} - \sum_{t=1}^{T} \varepsilon_t (MS_{t-1})' \{\sum_{t=1}^{T} (MS_{t-1})^{\otimes 2}\}^{-1} \sum_{t=1}^{T} (MS_{t-1}) \varepsilon_t'.$$

Due to uncorrelatedness of the regular components, $\tilde{U}_t, \tilde{V}_t, \lambda_t, D_t$, and the singular component, $Z_t$, established in Theorem 4.7 assuming A, C, D with $\gamma > 1$, then the matrix $Q_1 = \sum_{t=1}^{T} (MS_{t-1})^{\otimes 2}$ is asymptotically block diagonal. Moreover, the regular components are uncorrelated with the innovation $\varepsilon_t$, see (4.10), so $T\hat{\Omega}$ has leading term $\sum_{t=1}^{T} (\varepsilon_t | w_\perp' Z_{t-1})^{\otimes 2}$. The desired limits then arise from Theorem 4.7.

(*ii*) Same type of argument as in (*i*).

(*iii*) Rewrite $\hat{\mathbf{S}} - \mathbf{S} = Q_2 Q_1^{-1/2} M$ where $Q_2 = (\hat{\mathbf{S}} - \mathbf{S})(\sum_{t=1}^{T} S_{t-1}^{\otimes 2})^{1/2}$. The terms $Q_1$ and $Q_2$ were discussed in (*i*) and (*ii*). The regular component of the inverse, $Q_1^{-1}$, is $\mathrm{O}_{\mathsf{P}}(T^{-1})$, so the regular component of $\hat{\mathbf{S}} - \mathbf{S}$ vanishes. For the singular component use Theorem 4.7.

(*iv*) Due to (2.4) then $e_{W,t} = M_{W1} \varepsilon_t$ where $M_{W1} = (0, I_{\dim \mathbf{W}}) M (I_p, 0)'$. Thus, by (4.6), (4.7) then $\Omega_{\varepsilon Z} = \Omega M_{W1}' (\mathbf{W}^{-1})'$ and $\Omega_{ZZ} = \sum_{j=1}^{\infty} \mathbf{W}^{-j} M_{W1} \Omega M_{W1}' (\mathbf{W}^{-j})'$. The variance $\Omega_{ZZ}$ can be rewritten as $\mathbf{W}^{-1} M_{W1} \Omega M_{W1}' (\mathbf{W}^{-1})' + \mathbf{W}^{-1} \Omega_{ZZ} (\mathbf{W}^{-1})'$ where $\mathbf{W}^{-1} \Omega_{ZZ} (\mathbf{W}^{-1})'$ is positive definite since $\mathbf{W}$ is invertible and $\Omega_{ZZ}$ is positive definite, see (4.6). Consider two special cases.

*First*, suppose $p \geq \dim \mathbf{W} - n$. Define $A = w_\perp' \mathbf{W}^{-1} \Omega_{ZZ} (\mathbf{W}^{-1})' w_\perp$ as well as $B = M_{W1}' (\mathbf{W}^{-j})' w_\perp$. The matrices $A, B$ are random since $w$ is random. Then

$$\widetilde{\Omega} = \Omega - \Omega B (B' \Omega B + A)^{-1} B' \Omega.$$

Post-multiply by $(B, \Omega^{-1} B_\perp)$ where $B_\perp$ satisfies $B_\perp' B = 0$, $\mathrm{span}(B, B_\perp) = \mathbb{R}^p$. Then:

$$\widetilde{\Omega} B = \Omega B (B' \Omega B + A)^{-1} (B' \Omega B + A - B' \Omega B) = \Omega B (B' \Omega B + A)^{-1} A$$

has same rank as $B$ since $A$ and $B' \Omega B + A$ are invertible *a.s.*, while $\widetilde{\Omega} \Omega^{-1} B_\perp = B_\perp$. This shows that $\widetilde{\Omega}$ spans $\mathbb{R}^p$.

*Secondly*, suppose $p \leq \dim \mathbf{W} - n$. Let $B = w_\perp' \mathbf{W}^{-1} M_{W1}$. To cater explicitly for the situation where $B$ has reduced rank write $B = \xi\eta'$ where $\xi, \eta$ have full column rank. Then

$$\widetilde{\Omega} = \Omega - \Omega\eta\xi'(\xi\eta'\Omega\eta\xi' + A)^{-1}\xi\eta'\Omega.$$

Post-multiply by $(\Omega^{-1}\eta_\perp, \eta)$ to get $\widetilde{\Omega}\Omega^{-1}\eta_\perp = \eta_\perp$, while

$$\widetilde{\Omega}\eta = \Omega\eta - \Omega\eta\xi'(\xi\eta'\Omega\eta\xi' + A)^{-1}\xi\eta'\Omega\eta.$$

Post-multiplying the latter expression by $\xi'\overline{\xi} = I$, where $\overline{\xi} = \xi(\xi'\xi)^{-1}$ then gives

$$\widetilde{\Omega}\eta = \Omega\eta\xi'(\xi\eta'\Omega\eta\xi' + A)^{-1}A\overline{\xi}$$

which has the same rank as $\eta$.

$(v)$ The matrix $\Omega_{ZZ}$ is positive definite while $\Omega_{SZ}$ is non-zero due to (4.6), (4.7) using Assumptions A, C. Then use that $\mathsf{P}\{\text{rank}(w) = n\} = 1$ by Theorem 3.1$(ii)$ assuming A, B.

$(vi)$ The result in $(i)$ shows that the first two columns of

$$M(\hat{\mathbf{S}} - \mathbf{S})M^{-1} = \begin{pmatrix} \hat{\mathbf{U}} - \mathbf{U} & 0 & 0 \\ 0 & \hat{\mathbf{V}} - \mathbf{V} & 0 \\ 0 & 0 & \hat{\mathbf{W}} - \mathbf{W} \end{pmatrix}$$

vanish, so the eigenvalues of $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are consistent. The bias in the eigenvalues then arises from the limit of $\hat{\mathbf{W}} - \mathbf{W}$. This limit has bias $-\Omega_W(\mathbf{W}^{-1})'w_\perp(w_\perp'\Omega_Z w_\perp)^{-1}w_\perp'$ due to (4.7), which has rank $\dim \mathbf{W} - n$ *a.s.* This implies $(\hat{\mathbf{W}} - \mathbf{W})w \to 0$ *a.s.* By Lemma 3.3$(iii)$ then $\mathbf{W}w = wJ_n$ so $\overline{w}'\mathbf{W}w = J_n$, so that $n$ of the eigenvalues of $\hat{\mathbf{W}}$ are consistent. ■

A corresponding strong result applies, except that certain parameter restrictions are required for the unit root components.

**Theorem 5.2** *Assuming A, C, D, E with $\gamma > 1$ then*
*$(i)$ $\hat{\Omega} \to \tilde{\Omega}$ a.s.*
*$(ii)$ $(\hat{\mathbf{S}} - \mathbf{S})(\sum_{t=1}^T S_{t-1}^{\otimes 2})^{1/2} = \tilde{\mathbf{S}}_{norm}\{1 + \mathrm{o}(1)\} + \mathrm{o}(1)$ a.s.*
*$(iii)$ $\hat{\mathbf{S}} \to \tilde{\mathbf{S}}$ a.s. if $\dim \mathbf{D} = 0$.*

**Proof of Theorem 5.2.** Follow the proof of Theorem 5.1.

$(i, ii)$ Use $(iii, a)$ instead of $(iii, b)$ in Theorems 4.7 assuming E in addition.

$(iii)$ The regular component of the inverse, $Q_1^{-1}$, is $\mathrm{O}(T^{-1})$ *a.s.* when $\dim \mathbf{D} = 0$ due to Theorem 4.9 assuming E. ■

**Remark 5.3** *For regular vector autoregressions the term $w_\perp$ falls away and the bias term disappears. The results then correspond to those of Lai and Wei (1985) and Nielsen (2005a).*

**Remark 5.4** *For singular vector autoregressions the bias term is non-zero. Thus, the least squares estimators are inconsistent.*

**Example 5.5** *The bivariate, purely explosive case. Let* $p = 2$, $k = 1$, $\Omega = I_2$, $A_1 = \rho I_2$, $\dim \mathbf{D} = 0$, *so* $A_1 = \mathbf{S}$, $\boldsymbol{\iota}_S = I_2$, $\Omega_{\varepsilon W} = I_2$, $M_W = I_2$. *Then* $\Omega_Y = \sum_{j=1}^{\infty} \rho^{-2j} I_2 = (\rho^2 - 1) I_2$ *and*

$$\begin{aligned} \hat{A}_1 &\overset{a.s.}{\to} \rho I_2 - I_2 \rho^{-1}(\rho^2 - 1) w_{\perp}(w_{\perp}' w_{\perp})^{-1} w_{\perp}' \\ &= \rho w (w'w)^{-1} w + \rho^{-1} w_{\perp}(w_{\perp}' w_{\perp})^{-1} w_{\perp}', \end{aligned}$$

*which has eigenvalues at* $\rho$ *and* $\rho^{-1}$.

**Example 5.6** *The overfitted, explosive case. Let* $p = 2$, $k = 2$, $\Omega$ *unrestricted,* $A_1 = \rho I_2$, $A_2 = 0$, $\dim \mathbf{D} = 0$, *so*

$$\mathbf{B} = \begin{pmatrix} \rho I_2 & 0 \\ I_2 & 0 \end{pmatrix}, \qquad M = \begin{pmatrix} I & -\rho I \\ I & 0 \end{pmatrix},$$

$\mathbf{U} = 0$, $\mathbf{W} = \rho I_2$, $e_{U,t} = e_{W,t} = \varepsilon_t$, $\Omega_{\varepsilon W} = \Omega$ *and* $\Omega_Y = (\rho^2 - 1)^{-1}\Omega$. *Then*

$$\begin{aligned} \hat{A}_1 &\overset{a.s.}{\to} A_1 - (\rho - \rho^{-1}) \Omega w_{\perp}(w_{\perp}' \Omega w_{\perp})^{-1} w_{\perp}', \\ \hat{A}_2 &\overset{a.s.}{\to} A_2 = 0. \end{aligned}$$

*Thus, despite the inconsistency of the overall least squares estimator, the estimator for the over-fitted lag is consistent.*

## 6 Lag order determination

Lag order determination for vector autoregressions with deterministic terms is discussed in Nielsen (2006). As pointed out in Remark 2.3 the proofs only apply in the regular case. In the following it is shown that corresponding results hold in the singular case. That is, the order of a vector autoregression can be determined without knowledge about the location and the geometric multiplicity of the characteristic roots. The result is related to Example 5.6, which shows that the least squares estimators of the redundant lag coefficient matrices are zero.

The statistical model is now a $p$-dimensional vector autoregression of order $K$ so

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

conditional on the initial values $X_0, \ldots, X_{1-K}$. The effective sample is $X_1, \ldots, X_T$, for all sub-models with lag length $k < K$. The aim is to determine the largest non-trivial order for the time series, $k_0$ say, so $A_{k_0} \neq 0$ and $A_j = 0$ for $j > k_0$. Thus, it is convenient to give the variance estimator a subscript indicating the applied lag-length, that is $\hat{\Omega}_k$.

In the case of Gaussian innovations the likelihood ratio test statistic for testing that $A_k = 0$, in a model of order $k$ is

$$\mathsf{LR}(A_k = 0) = -T \log\det(\hat{\Omega}_{k-1}^{-1}\hat{\Omega}_k) = -T \log\det\{I_p - \hat{\Omega}_{k-1}^{-1}(\hat{\Omega}_{k-1} - \hat{\Omega}_k)\}. \qquad (6.1)$$

The result for the likelihood ratio statistic proved for the regular case by Nielsen (2006, Theorem 2.1, 2.2) also applies in the singular case.

**Theorem 6.1** *Assuming A, C, D with* $\gamma > 2$ *and* $k_0 < k$ *then* $\mathsf{LR}(A_k = 0)$ *is asymptotically* $\chi^2(p^2)$.

The lag length can also be determined maximising a penalised likelihood, or equivalently minimising information criteria of the type

$$\Phi_j = \log\det\hat{\Omega}_j + j\frac{f(T)}{T} \qquad j = 0, \ldots, K.$$

Then $k_0$ is estimated by the argument $\hat{k}$ that minimises $\Phi_j$. Several candidates for the penalty $f$ are applied. Akaike (1973) has $f(T) = 2p^2$, Schwarz (1978) has $f(T) = p^2 \log T$, while Hannan and Quinn (1979) and Quinn (1980) have $f(T) = 2p^2 \log\log T$. While these authors considered stationary autoregressions generalisations to unit root processes have been made by Paulsen (1984), Pötscher (1989) and Tsay (1984). Pötscher (1989) also considered explosive autoregressions. Nielsen (2006, Theorems 2.4, 2.5) established results concerning over-estimation and under-estimation of the estimator $\hat{k}$ in the regular case. These results can be generalised to the singular case. A small difference for the over-estimation results is that Assumption A is required with $\gamma > 1$ in the singular case, rather than just $\gamma > 0$.

**Theorem 6.2** *Assuming A, C, D with* $\gamma > 1$ *and* $f(T) = \mathrm{o}(T)$ *then*
*(i)* $\mathsf{P}(\hat{k} \geq k_0) \to 1$.
*(ii)* $\liminf_{T\to\infty} \hat{k} \geq k_0$ *a.s. assuming E in addition.*

For the under-estimation case it is convenient to separate weak and strong results. The weak result has the same conditions as Nielsen (2006, Theorem 2.5).

**Theorem 6.3** *Assuming A, C, D with* $\gamma > 2$ *and* $f(T) \to \infty$ *then* $\mathsf{P}(\hat{k} \leq k_0) \to 1$.

For the strong result the regular case is discussed fully in Nielsen (2006, Theorem 2.5) covering different degrees of parameter restrictions for the parameters $\mathbf{V}$, $\mathbf{D}$. For the singular case with $\mathbf{V}$, $\mathbf{D}$ restricted by Assumption E the following result can be formulated in the style of Hannan and Quinn (1979) and Quinn (1980).

**Theorem 6.4** *Assuming A, C, D, E with* $\gamma > 2$ *and* $\liminf_{T\to\infty}(2\log\log T)^{-1} f(T) > \{p + 2(\dim\mathbf{W} - n)\}^2$ *then* $\limsup_{T\to\infty} \hat{k} \leq k_0$ *a.s.*

For the proof of the above results some analysis of $\hat{\Omega}_{k-1} - \hat{\Omega}_k$ is needed. It is convenient to define, for any time series $R_t$,

$$Q\left(R_t\right) = \sum_{t=1}^{T} \varepsilon_t R_t' (\sum_{t=1}^{T} R_t^{\otimes 2})^{-1} \sum_{t=1}^{T} R_t \varepsilon_t'.$$

Thus, with $\mathbf{X}_{t-1} = (X_{t-1}', \ldots, X_{t-k+1}')'$ then

$$\hat{\Omega}_{k-1} = T^{-1} \sum_{t=1}^{T} (\varepsilon_t | \mathbf{X}_{t-1}, D_t)^{\otimes 2}, \qquad T(\hat{\Omega}_{k-1} - \hat{\Omega}_k) = Q(X_{t-k} | \mathbf{X}_{t-1}, D_t).$$

The next Lemma described the properties of $T(\hat{\Omega}_{k-1} - \hat{\Omega}_k)$.

**Lemma 6.5** *Assuming A, C, D with $\gamma > 2$ and $k_0 < k$ then there exists an $\{(p + \dim \mathbf{U}) \times p\}$-matrix $C$ with full column rank, so with $\hat{U}_t = C'\{\varepsilon_t', \tilde{U}_{t-1}'\}'$, and defining*

$$\begin{aligned} L &= \sum_{t=1}^{T} (\varepsilon_t | w_\perp' Z_{t-1}) \hat{U}_{t-1}' (\sum_{t=1}^{T} \hat{U}_{t-1}^{\otimes 2})^{-1} \sum_{t=1}^{T} \hat{U}_{t-1} (\varepsilon_t | w_\perp' Z_{t-1})', && (6.2) \\ R_V &= Q(V_{t-2} | D_t) - Q(V_{t-1} | D_t), && (6.3) \end{aligned}$$

*it holds*
*(i)* $T(\hat{\Omega}_{k-1} - \hat{\Omega}_k) = (L + R_V)\{1 + \mathrm{o}_\mathsf{P}(1)\} + \mathrm{o}_\mathsf{P}(1)$.
*(ii)* $T(\hat{\Omega}_{k-1} - \hat{\Omega}_k) = (L + R_V)\{1 + \mathrm{o}(1)\} + \mathrm{o}(1)$ *a.s. if Assumption E holds.*

The order of magnitude of the term $R_V$ is described in Nielsen (2006, Lemma 3.5).

To prove Lemma 6.5 some properties about the function $Q\left(R_t\right)$ are needed.

**Lemma 6.6** *Let $R_t = (R_{1,t}', R_{2,t}')'$ have sample correlation $\mathsf{c}_{R_1 R_2} = \mathrm{o}(1)$; see (4.9) for definition. Then*
*(i)* $Q(R_t) = \{Q(R_{1,t}) + Q(R_{2,t})\}\{1 + \mathrm{o}(1)\}$.
*(ii) If* $\mathsf{c}_{R_1 R_2}, \mathsf{c}_{\varepsilon R_2} = \mathrm{o}(T^{-1/4})$ *then* $Q(R_{1,t} | R_{2,t}) = \{Q(R_{1,t}) + \mathrm{o}_1\}\{1 + \mathrm{o}(1)\}$ *where* $\mathrm{o}_1 = \mathrm{o}(T^{-1} \sum_{t=1}^{T} \varepsilon_t^{\otimes 2})$.

**Proof of Lemma 6.6.** *(i)* Since $\mathsf{c}_{R_1 R_2} = \mathrm{o}\left(1\right)$ *a.s.* then

$$\sum_{t=1}^{T} R_t^{\otimes 2} = \begin{pmatrix} \sum_{t=1}^{T} R_{1,t}^{\otimes 2} & 0 \\ 0 & \sum_{t=1}^{T} R_{2,t}^{\otimes 2} \end{pmatrix} \{1 + \mathrm{o}\left(1\right)\}, \qquad a.s.$$

which leads to the desired result.

*(ii)* Write $Q\left(R_{1,t} | R_{2,t}\right)$ as $HH'$ where $H = H_1^{1/2} H_2 H_3^{1/2}$ with

$$H_1 = \sum_{t=1}^{T} \varepsilon_t^{\otimes 2}, \qquad H_2 = \mathsf{c}_{\varepsilon R_1} - \mathsf{c}_{\varepsilon R_2} \mathsf{c}_{R_2 R_1}, \qquad H_3 = 1 - \mathsf{c}_{R_1 R_2}^{\otimes 2}.$$

Since $\mathsf{c}_{R_1 R_2}, \mathsf{c}_{\varepsilon R_2} = \mathrm{o}(T^{-1/4})$ *a.s.* then

$$H_2 = \mathsf{c}_{\varepsilon R_1} - \mathrm{o}(T^{-1/2}), \qquad H_3 = 1 - \mathrm{o}(T^{-1/2}) \qquad a.s.$$

so $H = \{H_1^{1/2}\mathsf{c}_{\varepsilon R_1} + \mathrm{o}(T^{-1/2}H_1^{1/2})\}\{1+\mathrm{o}(1)\}$ giving the desired expression. ■

**Proof of Lemma 6.5.** $(i)$ The proof follows in various steps.
*First,* an algebraic argument. Due to Lemma 3.2 of Nielsen (2006) then

$$\begin{aligned} T(\hat\Omega_{k-1} - \hat\Omega_k) &= Q(X_{t-k}|\mathbf{X}_{t-1}, D_t) \\ &= Q\left(\mathbf{X}_{t-2}|D_t\right) - Q\left(\mathbf{X}_{t-1}|D_t\right) + Q\left(\varepsilon_{t-1}|\mathbf{X}_{t-2}, D_t\right). \end{aligned} \tag{6.4}$$

*Secondly,* analyse the terms in (6.4). Due to the uncorrelatedness established in Theorem 4.7 and (4.10), assuming A, C, D with $\gamma > 1$, then by Lemma 6.6$(i)$

$$Q\left(\mathbf{X}_{t-u}|D_t\right) = \{Q\left(\begin{array}{c} \tilde U_{t-u} \\ w_\perp' Z_{t-u} \end{array}\middle| D_t\right) + Q\left(V_{t-u}|D_t\right) + Q(w\lambda_{t-u}|D_t)\}\{1+\mathsf{o}_\mathsf{P}(1)\}.$$

Moreover, for $u = 1, 2$ the correlation between $D_t, V_{t-u}, \lambda_{t-u}$ and the terms $\varepsilon_t, \tilde U_{t-u}, Z_{t-u}$ is $\mathrm{o}(T^{-1/4})$ *a.s.* for $\gamma > 2$. By Lai and Wei (1985, Remark to Theorem 2) assuming A, C then $\sum_{t=1}^T \varepsilon_t^{\otimes 2} = \mathrm{O}(T)$. Then, by Lemma 6.6$(ii)$,

$$\begin{aligned} Q\left(\begin{array}{c} \tilde U_{t-u} \\ w_\perp' Z_{t-u} \end{array}\middle| D_t\right) &= Q\left(\begin{array}{c} \tilde U_{t-u} \\ w_\perp' Z_{t-u} \end{array}\right)\{1+\mathsf{o}_\mathsf{P}(1)\} + \mathsf{o}_\mathsf{P}(1), \\ Q\left(\varepsilon_{t-1}|\mathbf{X}_{t-2}, D_t\right) &= Q\left(\varepsilon_{t-1}|w_\perp' Z_{t-2}, U_{t-2}\right)\{1+\mathsf{o}_\mathsf{P}(1)\} + \mathsf{o}_\mathsf{P}(1). \end{aligned}$$

Insert the above results in (6.4), use $R_V = Q(V_{t-2}|D_t) - Q(V_{t-1}|D_t)$, and note that $Q(w\lambda_{t-2}) - Q(w\lambda_{t-1}) = \mathrm{o}(1)$ *a.s.* as in Nielsen (2006, equation 3.10) to get

$$\begin{aligned} T(\hat\Omega_{k-1} - \hat\Omega_k) &= \{Q\left(\begin{array}{c} \tilde U_{t-2} \\ w_\perp' Z_{t-2} \end{array}\right) - Q\left(\begin{array}{c} \tilde U_{t-1} \\ w_\perp' Z_{t-1} \end{array}\right) \\ &\quad + Q(\varepsilon_{t-1}|w_\perp' Z_{t-2}, \tilde U_{t-2}) + R_V\}\{1+\mathsf{o}_\mathsf{P}(1)\} + \mathsf{o}_\mathsf{P}(1). \end{aligned} \tag{6.5}$$

*Third,* by partial inversion, see Nielsen (2006, equation 3.4) then

$$\mathcal{I}_1 = Q\left(\begin{array}{c} \tilde U_{t-2} \\ w_\perp' Z_{t-2} \end{array}\right) + Q(\varepsilon_{t-1}|w_\perp' Z_{t-2}, \tilde U_{t-2}) = Q\left(\begin{array}{c} \varepsilon_{t-1} \\ \tilde U_{t-2} \\ w_\perp' Z_{t-2} \end{array}\right).$$

In the latter expression the index of $w_\perp' Z_{t-2}$ can be changed to $w_\perp' Z_{t-1}$. The argument is that $w_\perp' Z_{t-1} = (w_\perp' \mathbf{W} \overline{w}_\perp) w_\perp' Z_{t-2} - w_\perp' e_{W,t-1}$ due to Theorem 3.4$(ii)$. Since $e_{W,t}$ is a function of $\varepsilon_t$ then there is a bijective relation between $(\varepsilon_{t-1}', Z_{t-2}' w_\perp)$ and $(\varepsilon_{t-1}', Z_{t-1}' w_\perp)$ so $\mathcal{I}_1 = Q(\varepsilon_{t-1}', \tilde U_{t-2}', Z_{t-1}' w_\perp)$. Inserting in (6.5) gives

$$T(\hat\Omega_{k-1} - \hat\Omega_k) = \{Q\left(\begin{array}{c} \varepsilon_{t-1} \\ \tilde U_{t-2} \\ w_\perp' Z_{t-1} \end{array}\right) - Q\left(\begin{array}{c} \tilde U_{t-1} \\ w_\perp' Z_{t-1} \end{array}\right) + R_V\}\{1+\mathsf{o}_\mathsf{P}(1)\} + \mathsf{o}_\mathsf{P}(1). \tag{6.6}$$

*Fourthly,* Recall that $\tilde U_{t-1} = \mathbf{U}\tilde U_{t-2} + e_{U,t-1}$ where $e_{U,t-1} = M_U' \boldsymbol{\iota} \varepsilon_{t-1}$ for some matrix $M_U$ and $\boldsymbol{\iota} = (I_p, 0)'$. Since $M_U' \boldsymbol{\iota}$ has full row rank then $\tilde U_{t-1} = C_\perp'(\tilde U_{t-2}', \varepsilon_{t-1}')'$

where the $\{(\dim \mathbf{U}+p)\times \dim \mathbf{U}\}$-matrix $C_\perp$ has full column rank then a $\{(\dim \mathbf{U}+p)\times p\}$-matrix $C$ can be chosen so $(C, C_\perp)$ is regular and $\mathsf{Cov}(\hat U_{t-1}, \tilde U_{t-1}) = 0$ where $\hat U_{t-1} = C'(\tilde U'_{t-2}, \varepsilon'_{t-1})'$. Then

$$Q(\varepsilon'_{t-1}, \tilde U'_{t-2}, Z'_{t-1}w_\perp) = Q(\hat U'_{t-1}, \tilde U'_{t-1}, Z'_{t-1}w_\perp).$$

By partitioned inversion, see Nielsen (2006, equation 3.4), then

$$T(\hat\Omega_{k-1} - \hat\Omega_k) = \{Q(Y_{t-1}|w'_\perp Z_{t-1}, \tilde U_{t-1}) + R_V\}\{1 + \mathrm{o}_{\mathsf{P}}(1)\} + \mathrm{o}_{\mathsf{P}}(1). \tag{6.7}$$

*Fifthly,* note $T^{-1}\sum \varepsilon_t \tilde U'_{t-1}$ and $T^{-1}\sum Z_{t-1}\tilde U'_{t-1}$ are $\mathrm{o}(T^{-1/4})$ *a.s.* by Theorem 4.7 and (4.10), assuming A, C with $\gamma > 2$, while $T^{-1}\sum Z^{\otimes}_{t-1} - \Omega_{ZZ}$ and $T^{-1}\sum Z_{t-1}\varepsilon'_t - \mathsf{E}(Z_1\varepsilon'_2)$ are $\mathrm{o}_{\mathsf{P}}(1)$ by Theorem 4.6, Lemma 4.8 assuming A, C, while Nielsen (2005a, Theorem 6.2) gives laws of large numbers for sums of $\{(\varepsilon'_t, \tilde U'_{t-1})'\}^{\otimes 2}$. Combine these to see $\widehat{\mathsf{Cov}}(\varepsilon_t, \tilde U_{t-1}|w'_\perp Z_{t-1}) = \mathrm{o}_{\mathsf{P}}(T^{3/4})$. Due to Theorem 4.7 then $\mathsf{c}_{\varepsilon u|z} = \mathrm{o}_{\mathsf{P}}(T^{-1/4})$.

*Sixthly,* due to Nielsen (2005a, Theorem 2.4, 2006, Lemma 3.9), assuming A, C, then

$$T^{-1}\sum_{t=1}^T \begin{pmatrix} \varepsilon_{t-1} \\ \tilde U_{t-2} \end{pmatrix}^{\otimes 2} \overset{a.s.}{=} \begin{pmatrix} \Omega & 0 \\ 0 & \sum_{j=0}^\infty \mathbf{U}^j \Omega_{UU}(\mathbf{U}^j)' \end{pmatrix} + \mathrm{o}\left(T^{-\zeta}\right), \tag{6.8}$$

for $\Omega_{UU} = \mathsf{Var}(e^{\otimes 2}_{U,t})$ and $\zeta < \min\{\gamma/(2+\gamma), 1/2\}$. Now, construct $\hat U_{t-1}$ from $\varepsilon_{t-1}, \tilde U_{t-1}$ so $\mathsf{E}\hat U_{t-1}\tilde U'_{t-1} = 0$. It follows that $T^{-1}\sum \hat U_{t-1}\tilde U'_{t-1} = \mathrm{o}(T^{-1/4})$ *a.s.* for $\gamma > 2/3$. Then follow step 5 to show $\mathsf{c}_{yu|z} = \mathrm{o}_{\mathsf{P}}(T^{-1/4})$.

*Seventhly,* due to step 5 and 6 then Lemma 6.6$(ii)$ implies

$$T(\hat\Omega_{k-1} - \hat\Omega_k) = \{Q(\hat U_{t-1}|w'_\perp Z_{t-1}) + R_V\}\{1 + \mathrm{o}_{\mathsf{P}}(1)\} + \mathrm{o}_{\mathsf{P}}(1).$$

*Finally,* since $\tilde U_{t-1}$ is $\mathcal{F}_{t-1}$-measurable then taking iterated expectations shows $\mathsf{Cov}(\hat U_{t-1}, Z_{t-1}) = 0$. Together with laws of large numbers for $\hat U_{t-1}$ and $Z_{t-1}$ this implies $Q(\hat U_{t-1}|w'_\perp Z_{t-1})$ has the desired form.

$(ii)$ Use $(iii, a)$ instead of $(iii, b)$ in Theorem 4.7 assuming E in addition. ■

**Proof of Theorem 6.1.** Theorem 5.1$(i, iv)$ assuming A, C, D with $\gamma > 1$ shows that $\hat\Omega_{k-1} \to \tilde\Omega$ in probability and $\tilde\Omega > 0$ *a.s.* Lemma 6.5$(i)$ assuming A, C, D with $\gamma > 2$ describes the limit of $T(\hat\Omega_{k-1} - \hat\Omega_k)$. Nielsen (2006, Lemma 3.5,$ii$) assuming A, C, D shows $R_V = \mathrm{o}_{\mathsf{P}}(1)$. Insert these results in (6.1). Due to the Taylor expansion $LR = -T\log\det(1 - T^{-1}F_T) = \mathrm{tr}(F_T) + \mathrm{o}(F_T)$ the test statistic has leading term

$$\mathrm{tr}\{\tilde\Omega^{-1}\sum_{t=1}^T(\varepsilon_t|w'_\perp Z_{t-1})\hat U'_{t-1}(\sum_{t=1}^T \hat U^{\otimes 2}_{t-1})^{-1}\sum_{t=1}^T \hat U_{t-1}(\varepsilon_t|w'_\perp Z_{t-1})'\},$$

with $\hat U_t$ defined in Lemma 6.5. By Theorem 4.5 assuming A, D with $\gamma > 2$ and defining $Y_t = (\varepsilon'_t, Z'_{t-1}\mathbf{W})$ then $\sum_{t=1}^T Y_t\hat U'_{t-1} = \sum_{t=1}^T\sum_{j=0}^{t-1}\mathbf{Y}^j e_{Y,t}\hat U'_{t-1-j} + \mathrm{o}(T^{1/2})$. The leading term is a martingale. Then apply the Central Limit Theorem of Brown and Eagleson (1971). ■

**Proof of Theorem 6.2.** Consider $j < k_0$. The condition $f(T) = \mathrm{o}(T)$ implies

$$\Phi_j - \Phi_{k_0} = \log\det\{I_p + (\hat{\Omega}_j - \hat{\Omega}_{k_0})\hat{\Omega}_{k_0}^{-1}\} + \mathrm{o}(1).$$

$(i)$ Theorem 5.1$(i, iv)$ assuming A, C, D with $\gamma > 1$ shows that $\hat{\Omega}_{k-1} \to \tilde{\Omega}$ in probability and $\tilde{\Omega} > 0$ *a.s.* So it suffices to show that $\lambda_{\max}(\hat{\Omega}_j - \hat{\Omega}_{k_0})$ has a positive limit. Due to the successive inclusion of regressors it holds that $\hat{\Omega}_0 \geq \cdots \geq \hat{\Omega}_{k-1} \geq \hat{\Omega}_k$ using the ordering of positive semidefinite matrices. Thus it suffices to consider $j = k_0 - 1$.

Define the residuals $R_t = (X_{t-k_0}|X_{t-1}, \ldots, X_{t-k_0-1}, D_t)$. In the vector autoregression of order $k_0$ the least squares estimator for $A_{k_0}$ is $\hat{A}_{k_0} = \sum_{t=1}^T X_t R_t' (\sum_{t=1}^T R_t^{\otimes 2})^{-1}$, which implies

$$\hat{\Omega}_j - \hat{\Omega}_{k_0} = T^{-1} \sum_{t=1}^T X_t R_t' (\sum_{t=1}^T R_t^{\otimes 2})^{-1} \sum_{t=1}^T R_t X_t' = \hat{A}_{k_0}(T^{-1} \sum_{t=1}^T R_t^{\otimes 2}) \hat{A}_{k_0}'.$$

Due to Theorem 5.1 $(iii, v)$ then $\hat{A}_{k_0} = \tilde{A}_{k_0} + \mathrm{o}_{\mathsf{P}}(1)$ with $\tilde{A}_{k_0} = A_{k_0} + A_{bias}$. Here $A_{k_0} \neq 0$ by the definition of $k_0$ while the random bias $A_{bias}$ has the property that $\mathsf{P}(A_{k_0} + A_{bias} = 0) = 0$. Using Theorems 4.7, (4.10) and arguments as in Chan and Wei (1988) then $\lambda_{\min}(T^{-1} \sum_{t=1}^T R_t^{\otimes 2})$ either diverges or has a positive limit. The desired result then follows.

$(ii)$ Use Theorem 5.2 instead of Theorem 5.1 assuming E in addition. By Theorem 4.9 assuming A, C, D, E then $\liminf \lambda_{\min}(T^{-1} \sum_{t=1}^T R_t^{\otimes 2}) > 0$ *a.s.* ■

**Proof of Theorem 6.3.** Consider $k_0 < j \leq K$. Then

$$\Phi_{j+1} - \Phi_j = \log\det\{I_p - (\hat{\Omega}_j - \hat{\Omega}_{j+1})\hat{\Omega}_j^{-1}\} + T^{-1} f(T).$$

A Taylor expansion shows that if $\mathcal{I}_T = T(\hat{\Omega}_j - \hat{\Omega}_{j+1})\hat{\Omega}_j^{-1} = \mathrm{o}\{g(T)\}$ for some function $g$ so $f(T)/g(T) \to \infty$ the desired result holds. As in the proof of Theorem 6.1 assuming A, C, D with $\gamma > 2$ then $\mathcal{I}_T = \mathrm{O}_{\mathsf{P}}(1)$. ■

**Proof of Theorem 6.4.** As in the proof of Theorem 6.3 it has to be argued that $\mathcal{I} = T(\hat{\Omega}_j - \hat{\Omega}_{j+1})\hat{\Omega}_j^{-1} = \mathrm{o}\{g(T)\}$ where $g(T) = \mathrm{o}\{f(T)\}$.

Theorem 5.2 assuming A, C, D, E with $\gamma > 1$ shows that $\hat{\Omega}_{k-1} \to \tilde{\Omega}$ *a.s.*

Lemma 6.5$(ii)$ assuming A, C, D with $\gamma > 2$ shows that $T(\hat{\Omega}_j - \hat{\Omega}_{j+1})$ is decomposed in terms of quantities $L$ and $R_V$, defined in (6.2), (6.3).

If $\mathbf{V} = 1$ and $\dim \mathbf{D} = 0$ then $R_V = \mathrm{o}(1)$ *a.s.* by Nielsen (2006, Lemma 3.5) assuming A, C with $\gamma > 0$. If $\dim \mathbf{V} = 0$ then $R_V = 0$ by construction.

It is left to estimate the order of magnitude of $L$ defined in (6.2). As argue in connection with Theorem 5.2 and (6.8) then $T^{-1} \sum_{t=1}^T \{(Z_{t-1}' w_\perp, \varepsilon_t')'\}^{\otimes 2} \to \Omega_1 = \operatorname{diag}(w_\perp' \mathbf{W}, I_p)\Omega_Y \operatorname{diag}(\mathbf{W}' w_\perp, I_p)$ and $T^{-1} \sum_{t=1}^T \hat{U}_{t-1}^{\otimes 2} \to \Omega_2$ *a.s.* for some positive definite, random $\Omega_1$, $\Omega_2$. Thus, consider the coordinates of the matrix

$$\mathcal{I}_T = \Omega_1^{-1/2} \sum_{t=1}^T \begin{pmatrix} w_\perp' Z_{t-1} \\ \varepsilon_t \end{pmatrix} \hat{U}_{t-1}' \Omega_2^{-1/2}.$$

Theorem 4.5, assuming $\gamma > 2$, shows $\mathcal{I}_T \overset{a.s.}{=} \sum_{t=1}^{T} m_{eU,t} + \mathrm{o}(1)$, where $m_{eU,t} = \Omega_Y^{-1/2} \sum_{j=0}^{t-1} \mathbf{Y}^j e_{Y,t} \hat{U}'_{t-1-j} \Omega_2^{-1/2}$ is an $\mathcal{F}_t$-martingale difference.

It holds $T^{-1} \sum_{t=1}^{T} m_{eU,t}^{\otimes 2} \to I_{2p-n}$ $a.s.$ assuming A, C with $\gamma > 2$, which can be proved along the lines of Theorem 4.6. Now, apply the law of iterated logarithms by Stout (1974, Theorem 5.4.1) for each $(i,j)$-coordinate of $\mathcal{I}_T$. A conditional version of the truncation argument in the proof of Theorem 4.5 may be needed again assuming A with $\gamma > 2$. It then holds $\limsup_{T\to\infty} (2T \log\log T)^{-1/2} |\mathcal{I}_{T,ij}| \leq 1$.

By partitioned inversion then

$$\begin{aligned} &\sum_{t=1}^{T} \hat{U}_{t-1} (\varepsilon_t | w'_\perp Z_{t-1})' \{\sum_{t=1}^{T} (\varepsilon_t | w'_\perp Z_{t-1})^{\otimes 2}\}^{-1} \sum_{t=1}^{T} (\varepsilon_t | w'_\perp Z_{t-1}) \hat{U}'_{t-1} \\ = &\sum_{t=1}^{T} \hat{U}_{t-1} \begin{pmatrix} w'_\perp Z_{t-1} \\ \varepsilon_t \end{pmatrix}' \{\sum_{t=1}^{T} \begin{pmatrix} w'_\perp Z_{t-1} \\ \varepsilon_t \end{pmatrix}^{\otimes 2}\}^{-1} \sum_{t=1}^{T} \begin{pmatrix} w'_\perp Z_{t-1} \\ \varepsilon_t \end{pmatrix} \hat{U}'_{t-1} \\ &- \sum_{t=1}^{T} \hat{U}_{t-1} (w'_\perp Z_{t-1})' \{\sum_{t=1}^{T} (w'_\perp Z_{t-1})^{\otimes 2}\}^{-1} \sum_{t=1}^{T} (w'_\perp Z_{t-1}) \hat{U}'_{t-1}. \end{aligned}$$

Using the triangle inequality and the above bound gives the desired result. ■


## Acknowledgements

The author received financial support from ESRC grant RES-000-27-0179.


## References


Akaike, h. (1973) Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Caski, eds., *Proceeding of the Second International Symposium on Information Theory*, 267-281. Budapest: Akademiai Kiado.

Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676-687.

Breiman, L. (1968) *Probability.* Reading, MA: Addison-Wesley.

Brown, B.M. and Eagleson, G.K. (1971). Martingale convergence to infinitely divisible laws with finite variance. *Transactions of the American Mathematical Society* 162, 449-453.

Chan, N.H. and Wei, C.Z. (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

Chen, L.H.Y. (1978) A short note on the conditional Borel-Cantelli lemma. *Annals of Probability* 6, 699-700.

Chow, Y.S. (1965) Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics* 36, 552-558.

Donsker, M.D. and Varadhan, S.R.S. (1977) On Laws of Iterated Logarithms for local times. *Communications on Pure and Applied Mathematics* 30, 707-753.

Duflo, M. (1997) *Random iterative methods.* Berlin: Springer.

Duflo, M., R. Senoussi & R. Touati (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.

Engsted, T. (2006) Explosive bubbles in the cointegrated VAR model. *Finance Research Letters* 3, 154-162.

Hannan, E.J. and Quinn, B.G. (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B, 41, 190-195.

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

Lai, T.L. & C.Z. Wei (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. & C.Z. Wei (1983) 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. & C.Z. Wei (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P.R. Krishnaiah (ed.), *Multivariate Analysis VI*, pp. 375-393. Amsterdam: Elsevier Science Publishers.

Nielsen, B. (2005a) Strong consistency results for least squares statistics in general vector autoregressive models. *Econometric Theory* 21, 534-561.

Nielsen, B. (2005b) Analysis of co-explosive processes. Nuffield College Discussion Paper.

Nielsen, B. (2006) Order determination in general vector autoregressions. In Ho, H.-C., Ing, C.-K., and Lai, T.L. (eds): *Time Series and Related Topics: In Memory of Ching-Zong Wei.* IMS Lecture Notes and Monograph Series, 52, 93-112. doi:10.1214/074921706000000978.

Nielsen, B. (2008) On the Explosive Nature of Hyper-Inflation Data. *Economics: The Open-Access, Open-Assessment E-Journal* 2, 2008-21. http://www.economics-ejournal.org/economics/journalarticles/2008-21

Paulsen, J. (1984) Order determination of multivariate autoregressive time series with unit roots. *Journal of Time Series Analysis*, **5**, 115-127.

Phillips, P.C.B. and Magdalinos, T. (2008) Limit theory for explosively cointegrated systems. *Econometric Theory* 24, 865-887.

Phillips, P.C.B. and Solo, V. (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971-1001

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

Quinn, B. G. (1980) Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society* B, 42, 182-185.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

Stout, W.F. (1974) *Almost sure convergence.* New York: Academic Press

Tsay, R.S. (1984) Order selection in nonstationary autoregressive models. *Annals of Statistics*, **12**, 1425-1433.

Varga, R.S. (2000) *Matrix iterative analysis.* 2nd edition. Berlin: Springer.