

ESTIMATION OF AUTOCOVARANCE MATRICES FOR INFINITE DIMENSIONAL VECTOR LINEAR PROCESS

MONIKA BHATTACHARJEE
Statistics and Mathematics Unit
Indian Statistical Institute
203 B.T. Road, Kolkata 700108
INDIA
monaiidexp.gamma@gmail.com

ARUP BOSE *
Statistics and Mathematics Unit
Indian Statistical Institute
203 B.T. Road, Kolkata 700108
INDIA
bosearu@gmail.com

This version August 31, 2013

Abstract

Consider an Infinite Dimensional Vector Linear Process. Under suitable assumptions on the parameter space, we provide consistent estimators of the autocovariance matrices. In particular, under causality, this includes the IVAR process. In that case, we obtain consistent estimators for the parameter matrices. Explicit expression for the estimators are obtained for IVAR(1), under a fairly realistic parameter space. We also show that under some mild restrictions, the consistent estimator of the marginal large dimensional variance-covariance matrix has same convergence rate as that in case of i.i.d. samples.

Key words and phrases. High-dimensional data, IVAR, spatial variable, cross-sectional variables, variance-covariance matrix, marginal variance-covariance matrix, coefficient matrix, parameter matrix, k -th order autocovariance matrix, banding, consistency, convergence rate, operator norm.

AMS 2010 Subject Classifications. Primary 62H12, 62M10; Secondary 62F12, 65F35

*Research supported by J.C. Bose National Fellowship, Dept. of Science and Technology, Govt. of India.

1 Introduction

Vector ARMA processes are most common models in the theory of Econometrics and Finance. The estimates of the autocovariance matrices of different order under these models are well-known when the dimension of the vector is fixed (see [Hannan, 1970] and [Brockwell and Davis, 2002]).

However, in many situations the dimension grows along with the sample size or is large compared to the sample size. Examples include spatial data over time, fMRI datasets and datasets on spectroscopic imaging. In all these cases, the dimension is large compared to the sample size and also may increase as the next set of measurements are taken at the next time point. One of the key models in such situations is the infinite dimensional vector autoregressive (IVAR) process. Likewise there are infinite dimensional vector Autoregressive Moving Average (ARMA) models. The natural question which arises in these models is how to estimate the *sequence* of autocovariance matrices consistently. Note that these are all large dimensional matrices.

There are broadly two approaches to study the IVAR model in the literature. These are (i) data shrinkage and (ii) parameter shrinkage. Some of the works that use data shrinkage are [Forni et al., 2000], [Forni and Lippi, 2001], [Forni et al., 2004], [Chudik and Persaran, 2011], whereas [Anselin, 1988], [Doan et al., 1984] use parameter shrinkage.

Note that under an appropriate causality condition, a vector ARMA process can be expressed as a linear process (see [Brockwell and Davis, 2002] and [Hannan, 1970]). We will consider general Infinite Dimensional Vector Linear Processes with suitable assumptions on the variables and the parameter matrices. Our aim is to obtain consistent estimates of the sequence of autocovariance matrices using the parameter shrinkage approach and obtain bounds on the rate of consistency.

Let, $\{X_{t,p}^{(n)}, t = 1, 2, \dots, n\}$ be a sample of size n from the linear process model (2.1) given in Section 2. In this model n is the time dimension and p is the parameter dimension. We assume there is both time (column) and spatial (row) dependence and both dimensions grow, p growing much faster than n . We assume weak dependence in both time and space. Thus, $X_{t,p}$ will be less dependent on $X_{t-2,p}$ than on $X_{t-1,p}$. Also for any $1 \leq i \leq p$, let $X_{t,i,p}$ be the i -th component of the vector $X_{t,p}$. Then for any $t_1 < t$ and any $k > 0$, the dependence between $X_{t_1,(i \pm k),p}$ and $X_{t,i,p}$ decreases as the lag k increases. In Section 3, under appropriate quantification of this weak dependence, we define suitable classes for coefficient matrices for the model (2.1). These matrices are not assumed to be symmetric. The variance-covariance matrix Σ_p of the driving i.i.d. process is assumed to satisfy the same conditions as [Bickel and Levina, 2008] who discussed its consistent estimation in i.i.d. models.

In Section 4, we show that for each k , the banded version $B_{k_n}(\hat{\Gamma}_k)$ of the sample autocovariance matrices $\hat{\Gamma}_k$ are consistent for the corresponding population autocovariance matrices Γ_k of order k . Here $\{k_n\}$ is an appropriate sequence of banding coefficients whose value depends on n , p and a weak dependence parameter α . The convergence rate is also derived. In particular, if the row or column sums of Σ_p are bounded, then $B_{k_n}(\hat{\Gamma}_0)$ as an estimate of the dispersion matrix of X_t has the same convergence rate as in i.i.d. sample case given in [Bickel and Levina, 2008]. Using this consistency, we also provide consistent estimators for the parameter matrices of the IVAR(r) process under an appropriate class of parameter matrices.

Some simulation results are given in Section 5.

All the necessary lemmas and proofs are given in Section 6.

2 Infinite dimensional vector linear process

Suppose $\{X_{t,p}^{(n)}, t = 1, 2, 3, \dots, n\}$ is a sample of size n from the model

$$X_{t,p}^{(n)} = \sum_{j=0}^{\infty} \psi_{j,p}^{(n)} \varepsilon_{t-j,p}, t, n \geq 1 \text{ (almost surely)}. \quad (2.1)$$

For all t , $X_{t,p}^{(n)}$ and $\varepsilon_{t,p}$ are p -dimensional vectors. $\varepsilon_{t,p}$ are independently and identically distributed with mean 0 and variance-covariance matrix Σ_p . The $p \times p$ matrices $\psi_{j,p}^{(n)}$, $j \geq 1$ are called *coefficient matrices* of the model (2.1). The dimension $p = p(n)$ increases to ∞ along with the sample size n . Therefore, as p increases, the dimension of the coefficient matrices also increases. We are interested in the case where p is much larger than n . At the n -th stage the coefficient matrices are $\{\psi_{j,p}^{(n)}\}_{j=0}^{\infty}$ and the corresponding observations are $\{X_{t,p}^{(n)}, t = 1, 2, \dots, n\}$. Hence, we have the following triangular sequence:

$$\begin{array}{c} X_{1,p(1)}^{(1)} \\ X_{1,p(2)}^{(2)}, X_{2,p(2)}^{(2)} \\ X_{1,p(3)}^{(3)}, X_{2,p(3)}^{(3)}, X_{3,p(3)}^{(3)} \\ \vdots \\ X_{1,p(n)}^{(n)}, X_{2,p(n)}^{(n)}, X_{3,p(n)}^{(n)}, \dots, X_{n,p(n)}^{(n)} \\ \vdots \end{array}$$

and the sample at the n -th stage is the n -th row of this triangular sequence. For convenience of notation, we will usually omit the superscript (n) and also write p for $p(n)$. We shall often write ψ_j for $\psi_{j,p}^{(n)}$ etc.

Example 2.1. Consider the *finite order moving average process*,

$$X_{t,p} = \sum_{i=0}^r M_{i,p} \varepsilon_{t-i,p}, t \geq 1 \quad (2.2)$$

where $M_{i,p}$, $i = 0, 1, 2, \dots, r$ are square matrices of order p and $M_{0,p} = I_p$. In this case, $\psi_{i,p} = M_{i,p}$, $1 \leq i \leq r$ and $\psi_{i,p} = 0$, $i > r$.

Example 2.2. The *Infinite dimensional Vector Autoregressive Process* of order r (IVAR(r)) is given by

$$X_{t,p} = \sum_{i=1}^r A_{i,p} X_{t-i,p} + \varepsilon_{t,p}, t \geq 1 \quad (2.3)$$

where the $p \times p$ matrices $A_{i,p}$, $i = 1, 2, \dots, r$, will be called *parameter matrices*. Moreover, if for some $\epsilon > 0$,

$$\det(I - A_{1,p}z - A_{2,p}z^2 - \dots - A_{r,p}z^r) \neq 0, \text{ for all } z \in \mathbb{C} \text{ such that } |z| \leq 1 + \epsilon. \quad (2.4)$$

then we have the representation,

$$X_{t,p} = \sum_{j=0}^{\infty} \phi_{j,p} \varepsilon_{t-j}, \quad t \geq 1 \quad (\text{almost surely}), \quad (2.5)$$

where $\phi_{0,p} = I$ and $\phi_{j,p} = \sum_{i=1}^j A_{i,p} \phi_{j-i,p}$, $j \geq 1$. So, this fits into the model (2.1) with $\psi_{j,p} = \phi_{j,p}$, $j \geq 0$ ([Brockwell and Davis, 2002], [Hannan, 1970]).

In the general model (2.1), let $\Gamma_{k,p}$ be the k th *order autocovariance matrix*,

$$\Gamma_{k,p} = \text{Cov}(X_{t,p}, X_{(t-k),p}) = E(X_{t,p} X'_{(t-k),p}) = \sum_{j=k}^{\infty} \psi_j \Sigma_p \psi'_{j-k}. \quad (2.6)$$

Our primary goal is to estimate, for any fixed k , the autocovariance matrix sequence $\{\Gamma_{k,p}\}$ consistently from a sample of size n . Here p will be allowed to grow faster than n .

For each k , define the sample autocovariance matrices of order k as,

$$\hat{\Gamma}_{k,p} = \frac{1}{n-k} \sum_{t=k+1}^n X_{t,p} X'_{(t-k),p}, \quad 0 \leq k \leq n-1. \quad (2.7)$$

These are the naive method of moments estimators for the corresponding $\Gamma_{k,p}$. We often write Γ_k and $\hat{\Gamma}_k$ respectively for $\Gamma_{k,p}$ and $\hat{\Gamma}_{k,p}$. From the experience of covariance matrix estimation in high dimensional i.i.d. models (see [Bickel and Levina, 2008]), these estimators are not expected to be consistent in any reasonable norm.

To achieve consistency, two things are needed. First, we need to restrict the parameter space of the population autocovariance matrices. Second, we need to resort to some banding or tapering of the sample autocovariance matrices. We are able to show that once we take care of the above two issues, we achieve consistency. We also obtain an upper bound for the rate of consistency. The naive parameter matrices in the IVAR(r) model are obtained by least squares. Thus once the above consistency is achieved, we can easily obtain consistent estimates of the parameter matrices in these models.

3 Classes : parameter shrinkage

To impose restrictions on the parameter space, define for any matrix M of order p ,

$$\begin{aligned} \|M\|_{(1,1)} &= \max_j \sum_{i=1}^p |m_{ij}|, \quad \|M\|_2 = \sqrt{\lambda_{\max}(M'M)}, \\ \|M\|_{\infty} &= \max_{i,j} |m_{ij}|, \quad T(M, t) = \max_j \sum_{i:|i-j|>t} |m_{ij}|. \end{aligned}$$

To estimate the variance-covariance matrix from an *i.i.d.* sample, [Bickel and Levina, 2008] considered the following parameter space. This class includes many non-trivial variance-covariance matrices. We shall also work within this parameter space.

$$\mathcal{U}(\epsilon, \alpha, C) = \{\Sigma : 0 < \epsilon < \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) < \epsilon^{-1}, T(\Sigma, t) < Ct^{-\alpha} \forall t > 0\}$$

where $\epsilon, \alpha, C > 0$ and $\lambda_{\min}(\Sigma), \lambda_{\max}(\Sigma)$ are respectively the minimum and maximum eigenvalues of Σ .

To define an appropriate parameter space for the sequence of matrices $\{\psi_j\}_{j=0}^{\infty}$, we assume the following criteria.

- (i) **Time lag criterion:** We ensure that the dependence decreases appropriately with the lag. For this purpose, define

$$\max(\|\psi_j\|_{(1,1)}, \|\psi'_j\|_{(1,1)}) = r_j, j \geq 0.$$

We define the following class $\mathfrak{I}(\beta, \lambda)$ of sequence of matrices $\{\psi_j\}_{j=0}^{\infty}$ for some $0 < \beta < 1$ and $\lambda \geq 0$.

$$\mathfrak{I}(\beta, \lambda) = \left\{ \{\psi_j\}_{j=0}^{\infty} : \sum_{j=0}^{\infty} r_j^{\beta} < \infty, \sum_{j=0}^{\infty} r_j^{2(1-\beta)} j^{\lambda} < \infty \right\}.$$

Note that the summability above implies that the decay rate of r_j cannot be slower than a polynomial rate. In case of finite order moving average process, in Example 2.1, as we have a finite number of parameter matrices, $\{\psi_j\}$ will automatically belong to $\mathfrak{I}(\beta, \lambda)$ for all $0 < \beta < 1$ and $\lambda \geq 0$.

- (ii) **Spatial lag criterion:** For any $1 \leq i \leq p$, let $X_{t,i,p}$ be the i -th component of the vector $X_{t,p}$. Here we ensure that for any $t_1 < t$ and $k > 0$, the dependence between $X_{t_1, (i \pm k), p}$ and $X_{t, i, p}$ grows weaker as the lag k increases. We achieve this by putting restrictions over $\{T(\psi_j, t) : j = 0, 1, 2, \dots\}$ for all $t > 0$. Consider the following class $\mathcal{G}(C, \alpha, \eta, \nu)$ for some $C, \alpha, \nu > 0$ and $0 < \eta < 1$ as

$$\mathcal{G}(C, \alpha, \eta, \nu) = \left\{ \{\psi_j\} : T\left(\psi_j, t \sum_{u=0}^j \eta^u\right) < Ct^{-\alpha} r_j j^{\nu} \sum_{u=0}^j \eta^{-u\alpha}, \text{ and } \sum_{j=k}^{\infty} \frac{r_j r_{j-k} j^{\nu}}{\eta^{\alpha j}} < \infty \right\}.$$

Let us define the class of matrices $\mathcal{L}_T(C, \alpha)$,

$$\mathcal{L}_T(C, \alpha) = \{A : T(A, t) < Ct^{-\alpha} \forall t > 0\}$$

for some $C, \alpha > 0$. Observe that if $M_i \in \mathcal{L}_T(C, \alpha)$, $1 \leq i \leq r$, (see Example 2.1) then it will clearly be in $\mathcal{G}(C, \alpha, \eta, \nu)$ for any $0 < \eta < 1$ and $\nu > 0$.

In case of autoregressive process of order r , in Example 2.2, let

$$\|A_{i,p}\|_{(1,1)} = \theta_{i,p} \text{ and } \|A'_{i,p}\|_{(1,1)} = \theta'_{i,p}, 1 \leq i \leq r. \quad (3.1)$$

Also let $\{\alpha_{i,p} : i = 1, 2, \dots, r\}$ and $\{\alpha'_{i,p} : i = 1, 2, \dots, r\}$ respectively be the roots of the following polynomials.

$$\begin{aligned} 1 - \theta_{1,p}z - \theta_{2,p}z^2 \dots \theta_{r,p}z^r &= 0, \\ 1 - \theta'_{1,p}z - \theta'_{2,p}z^2 \dots \theta'_{r,p}z^r &= 0. \end{aligned}$$

Now, we consider the parameter space for $\{A_i\}_{i=1}^r$ as,

$$\mathcal{P}(C, \alpha, \epsilon) = \left\{ \{A_i\}_{i=1}^r : \inf_p \min_{1 \leq i \leq r} (|\alpha_{i,p}|, |\alpha'_{i,p}|) > 1 + \epsilon, \text{ and } A_i \in \mathcal{L}_T(C, \alpha) \forall i \right\}$$

for some $C, \epsilon, \alpha > 0$. It can be shown that if $\{A_i\}_{i=1}^r \in \mathcal{P}(C, \alpha, \epsilon)$, then the representation (2.5) is well-defined (in the almost sure sense). Also, $\{\phi_i\}_{i=0}^{\infty} \in \mathfrak{I}(\beta, 0)$ for any $0 < \beta < 1$ and $\{\phi_i\}_{i=0}^{\infty} \in \mathcal{G}(C, \alpha, \eta, 1)$ for some $0 < \eta < 1$ (see Lemma 6.2).

4 Main results

In this section we state all the main results. Their proofs are given in Section 6. For any matrix M of order p and $k > 0$, the *banded* version of M is given by

$$B_k(M) = ((m_{ij}I(|i - j| \leq k))).$$

The following theorem provides consistent estimates for the autocovariance matrix sequence $\{\Gamma_k\}_{k=0,1,2,\dots}$. Let $\varepsilon_{t,j,p}$ be the j -th component of $\varepsilon_{t,p}$. We often omit p for convenience and write ε_t for $\varepsilon_{t,p}$. Suppose for some $\lambda_0 > 0$,

$$\sup_{j \geq 1} E(e^{\lambda \varepsilon_{t,j}}) < \infty \text{ for all } |\lambda| < \lambda_0. \quad (4.1)$$

Theorem 4.1. *Suppose (4.1) holds, $\Sigma \in \mathcal{U}$ and $\{\psi_j\}_{j=0}^\infty \in \mathfrak{Y} \cap \mathcal{G}$. Then for $k_n = (n^{-1} \log p)^{-\frac{1}{2(\alpha+1)}}$,*

$$\|B_{k_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = O_p(k_n^{-\alpha} \|\Sigma_p\|_{(1,1)}).$$

Remark 4.1. *Note that the rate of convergence depends not only on the class parameters of the coefficient and variance-covariance matrix but also on $\|\Sigma_p\|_{(1,1)}$. This is to be expected since we are considering linear regression type models. Moreover, if $\|\Sigma_p\|_{(1,1)}$ is bounded, then the rate of convergence for the marginal variance-covariance matrix Γ_0 is same as given in [Bickel and Levina, 2008].*

The following Theorem provides consistency for the autocovariance matrices of the IVAR(r) process defined in Example 2.2 of Section 2. Its proof follows directly from Theorem 4.1 once we invoke Lemma 6.2 (see Section 6).

Theorem 4.2. *Suppose (4.1) holds, $\Sigma \in \mathcal{U}$ and $\{A_j\}_{j=1}^r \in \mathcal{P}$. Then for $k_n = (n^{-1} \log p)^{-\frac{1}{2(\alpha+1)}}$,*

$$\|B_{k_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = O_p(k_n^{-\alpha} \|\Sigma_p\|_{(1,1)}).$$

Now, right multiplying both sides of (2.3) with $X'_{t-k,p}$, $k = 1, 2, \dots, r$ successively and then taking expectation, we have

$$\begin{aligned} \Gamma_1 &= A_1 \Gamma_0 + A_2 \Gamma'_1 + \dots + A_r \Gamma'_{r-1} \\ \Gamma_2 &= A_1 \Gamma'_1 + A_2 \Gamma_0 + \dots + A_r \Gamma'_{r-2} \\ &\vdots \\ \Gamma_r &= A_1 \Gamma_{r-1} + A_2 \Gamma_{r-2} + \dots + A_r \Gamma_0. \end{aligned} \quad (4.2)$$

Let

$$\mathcal{Y}_r = (\Gamma'_1, \Gamma'_2, \dots, \Gamma'_r)', \quad \mathcal{A}_r = (A'_1, A'_2, \dots, A'_r)'$$

and let G_n be a block matrix with n^2 many $p \times p$ blocks

$$G_n(i, j) = \Gamma_{|i-j|} I(i < j) + \Gamma'_{|i-j|} I(i \geq j), \quad 1 \leq i, j \leq n. \quad (4.3)$$

Then from (4.2) we have,

$$\mathcal{Y}_r = G_n \mathcal{A}_r. \quad (4.4)$$

This is analogous to the Yule-Walker equation for finite dimensional AR process. Similar to the univariate case [Brockwell and Davis, 2002], it can be shown that if $a' \Gamma_0 a > 0$, then G_n is invertible and $a' \Gamma_h \Gamma_h' a \rightarrow 0$ as $h \rightarrow \infty \forall a \neq 0$ (see Lemma 6.9). Therefore, under this condition,

$$\mathcal{A}_r = G_n^{-1} \mathcal{Y}_r \quad (4.5)$$

i.e., each A_i is the finite sum of the finite products of $\{\Gamma_k, \Gamma_k^{-1}, 1 \leq k \leq r\}$. Hence, (after we invoke Lemma 6.1), equation (4.5) provides consistent estimates of A_i , once we replace the population autocovariance matrices by their consistent estimates. In the special case of IVAR(1) model this result takes the following shape.

Theorem 4.3. *Suppose (4.1) holds and $\Sigma \in \mathcal{U}(\epsilon, \alpha, C_\sigma)$. Let $k_n = (n^{-1} \log p)^{-\frac{1}{2(1+\alpha)}}$ and $A \in \{A_{\infty \times \infty} : \max(\|A\|_{(1,1)}, \|A'\|_{(1,1)}) < (1 - \delta), A, A' \in \mathcal{L}_T(C, \alpha)\}$. Then*

$$(i) \|B_{k_n}(\hat{\Gamma}_1)(B_{k_n}(\hat{\Gamma}_0))^{-1} - A\|_2 = O_p(\|\Sigma_p\|_{(1,1)} k_n^{-\alpha}) \text{ and}$$

$$(ii) \|\hat{\Sigma}_p - \Sigma_p\|_2 = O_p(\|\Sigma_p\|_{(1,1)} k_n^{-\alpha}),$$

where

$$\hat{\Sigma}_p = B_{k_n}(\hat{\Gamma}_0) - B_{k_n}(\hat{\Gamma}_1)(B_{k_n}(\hat{\Gamma}_0))^{-1} B_{k_n}(\hat{\Gamma}_1').$$

Remark 4.2. *We have stated all our results for banded estimators. Following the proofs given later and the arguments given in [Bickel and Levina, 2008] and [Bhattacharjee and Bose, 2013], one can obtain similar consistency results for tapered estimators. To illustrate we give below the version of Theorem 4.1 for tapered estimators.*

Let $g : \mathbb{R}^+ \cup \{0\} \rightarrow \mathbb{R}^+ \cup \{0\}$ be a continuous, non-increasing function such that $g(0) = 1$, $\int_0^\infty g(x) < \infty$ and $1 - g(x) = O(x^\nu)$ for some $\nu \geq 1$ in some neighborhood of zero. Define

$$r_{\sigma_n}(i, j) = g\left(\frac{|i - j|}{\sigma_n}\right) \quad \forall 1 \leq i, j \leq p \text{ and for some } \sigma_n > 0,$$

$$R_n = ((r_{\sigma_n}(i, j)))_{1 \leq i, j \leq p}, \quad \forall n \geq 1,$$

$$R_{\sigma_n}(A) = A * R_n, \quad \forall n \geq 1 \text{ and for any matrix } A.$$

Also, g is such that R_n is positive definite. One such choice is $g(x) = e^{-|x|}$. Then under the conditions of Theorem 4.1, we have

$$\|R_{\sigma_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = O_p(k_n^{-\frac{\alpha\gamma}{2(1+\gamma)}} \|\Sigma_p\|_{(1,1)})$$

and as $\gamma \rightarrow \infty$, we recover the rate given in Theorem 4.1.

5 Simulations

This section reports simulations for the IVAR(1) models with different choices of the parameter matrix A which have the Toeplitz structure. As we move away from the main diagonal, in one case the entries decrease exponentially and in the other case, they decrease polynomially.

Example 5.1. Exponentially decaying corners: Consider the IVAR(1) model with the non-symmetric parameter matrix $A = (((-0.5)^{|i-j|}))$. As the rows or columns of this matrix are summable, it belongs to the class defined in Theorem 4.3 with $\alpha = 1$.

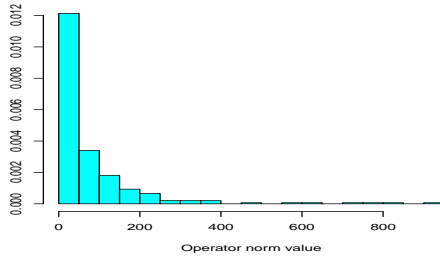
Example 5.2. Polynomially decaying corners: Consider the IVAR(1) model with the non-symmetric parameter matrix $A = (((-1)^{|i-j|} (|i-j| + 1)^{-\beta}))$ and we choose $\beta = 1.01, 1.1, 1.2$ and 1.5 . Again, since the rows and columns of the matrices are summable. They belongs to the class defined in Theorem 4.3, with $\alpha = \beta - 1$. This is clear from the fact

$$T(A, k) \leq \int_{k+1}^{\infty} \left(\frac{1}{x}\right)^{\beta} dx \leq C(k+1)^{\beta-1}$$

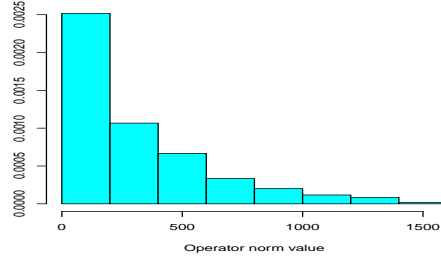
for some $C > 0$ and all $k \geq 1$.

We let $\varepsilon_t \sim N(0, I)$ where I is the identity matrix. In each case, we draw the histogram for $\|B_{k_n}(\hat{\Gamma}_1)(B_{k_n}(\hat{\Gamma}_0))^{-1} - A\|_2$ values for $R = 300$ replications. We consider two combinations of n and p , namely $n = 20$, $p = e^{\sqrt{n}} \sim 87$ and $n = 40$, $p = e^{\sqrt{n}} \sim 558$.

Note that most of the mass is concentrated near zero. Expectedly, the accuracy is sharper in the first example than in the second example. Moreover, as β increases, the histogram has more mass near zero and there is some mass in the high values of the tail. Some stray values beyond the range given in the figures were observed over the different sets of simulations but overall most of the mass was concentrated in the range $(0, 600)$. This indicates that the rates of convergence are probably quite sharp.



Example 5.1 $n = 20$



Example 5.1 $n = 40$

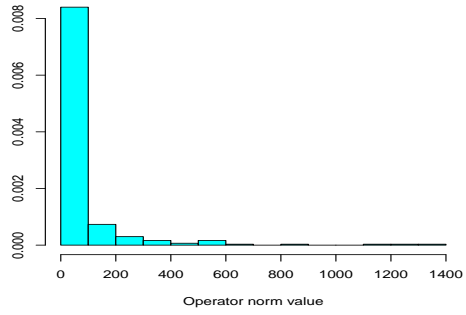
6 Proofs

The following inequalities for matrix norms will be useful in the subsequent proofs.

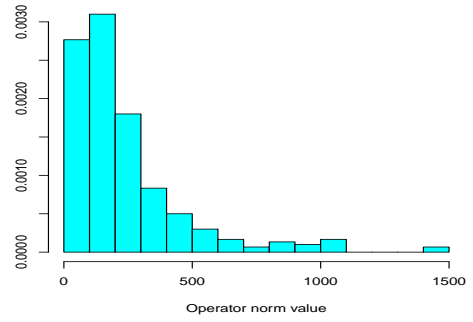
Lemma 6.1. *Let A and B be two $\infty \times \infty$ matrices. Then,*

(i) $\|AB\|_{(1,1)} \leq \|A\|_{(1,1)}\|B\|_{(1,1)}$,

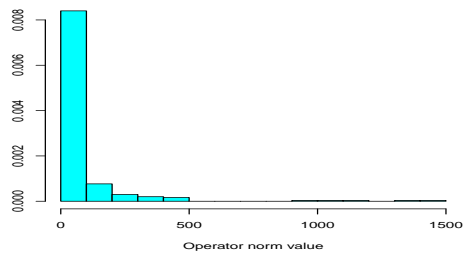
(ii) $T(AB, (\alpha + \beta)t) \leq \|A\|_{(1,1)} T(B, \alpha t) + \|B\|_{(1,1)} T(A, \beta t)$, for any $\alpha, \beta, t > 0$,



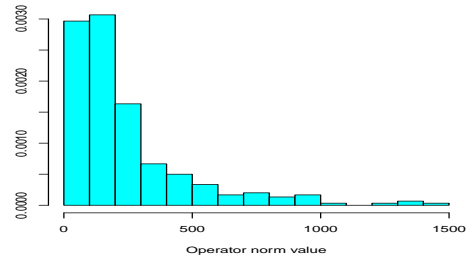
Example 5.2 $\beta = 1.01$ $n = 20$



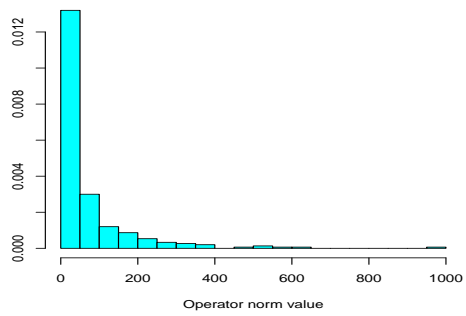
Example 5.2 $\beta = 1.01$ $n = 40$



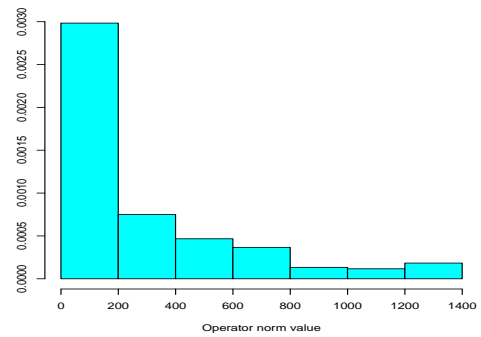
Example 5.2 $\beta = 1.1$ $n = 20$



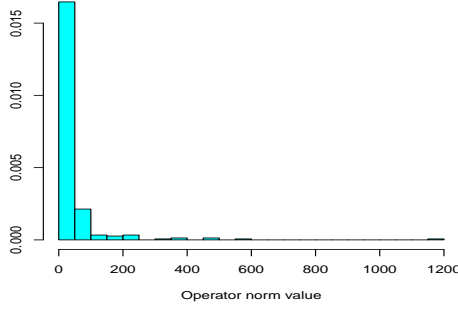
Example 5.2 $\beta = 1.1$ $n = 40$



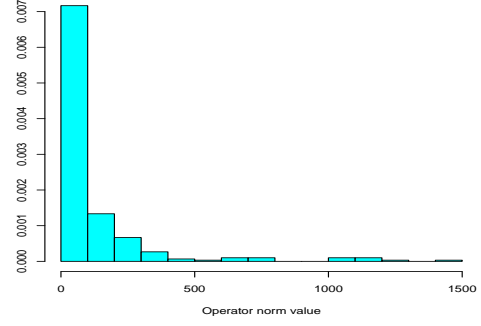
Example 5.2 $\beta = 1.2$ $n = 20$



Example 5.2 $\beta = 1.2$ $n = 40$



Example 5.2 $\beta = 1.5$ $n = 20$



Example 5.2 $\beta = 1.5$ $n = 40$

Proof. Proof of (i) is trivial. To prove (ii).

$$\begin{aligned}
T(AB, (\alpha + \beta)t) &\leq \max_k \sum_{j:|j-k|>(\alpha+\beta)t} \left(\sum_{l=1}^{\infty} |a_{jl}b_{lk}| \right) \\
&\leq \max_k \sum_{j:|j-k|>(\alpha+\beta)t} \left(\sum_{l:|l-k|\leq \alpha t} |a_{jl}b_{lk}| \right) + \max_k \sum_{j:|j-k|>(\alpha+\beta)t} \left(\sum_{l:|l-k|>\alpha t} |a_{jl}b_{lk}| \right) \\
&\leq \max_{j:|j-l|>\beta t, l:|l-k|\leq \alpha t} |a_{jl}b_{lk}| + \max_{j:|j-k|>(\alpha+\beta)t, l:|l-k|>\alpha t} |a_{jl}b_{lk}| \\
&\leq \left(\max_l \sum_{j:|j-l|>\beta t} |a_{jl}| \right) \left(\max_k \sum_{l=1}^{\infty} |b_{lk}| \right) + \left(\max_k \sum_{l:|l-k|>\alpha t} |b_{lk}| \right) \left(\max_l \sum_{j=1}^{\infty} |a_{jl}| \right) \\
&\leq \|B\|_{(1,1)} T(A, \beta t) + \|A\|_{(1,1)} T(B, \alpha t).
\end{aligned}$$

□

The following lemma proves the claim about $\mathcal{P}(C, \alpha, \varepsilon)$ made at the very end of Section 3.

Lemma 6.2. *If $\{A_i\}_{i=1}^r \in \mathcal{P}(C, \alpha, \varepsilon)$, then (2.5) holds. Also, $\{\phi_i\}_{i=0}^{\infty} \in \mathfrak{F}(\beta, 0) \cap \mathcal{G}(C, \alpha, \eta, 1)$ for any $0 < \beta < 1$ and some $0 < \eta < 1$.*

Proof. To prove (2.5), we need to show that if $\{A_i\}_{i=1}^r \in \mathcal{P}(C, \alpha, \varepsilon)$, then it will satisfy condition (2.4). Define the polynomials

$$\begin{aligned}
p_1(x) &= \theta_1 x + \theta_2 x^2 + \cdots + \theta_r x^r \\
p_2(x) &= \theta'_1 x + \theta'_2 x^2 + \cdots + \theta'_r x^r.
\end{aligned}$$

Here we write θ_i, θ'_i respectively for $\theta_{i,p}$ and $\theta'_{i,p}$ defined in (3.1). Note that $p_1(0) = p_2(0) = 0$ and both of them are increasing functions of x . Also, as $(1 - p_1(x))$ and $(1 - p_2(x))$ have all their roots strictly greater than $(1 + \varepsilon)$, $p_i(1 + \varepsilon) < 1 \forall i = 1, 2$. Now, for any $|z| \leq 1 + \varepsilon$ and any $x \neq 0$, using $\|A\|_2 \leq \sqrt{\|A\|_{(1,1)} \|A'\|_{(1,1)}}$ [Golub and Van Loan, 1996],

$$|x' A_1 x z + x' A_2 x z^2 + \cdots + x' A_r x z^r| \leq \sum_{i=1}^r \sqrt{\theta_i \theta'_i} |z|^i \leq \frac{1}{2} (p_1(1 + \varepsilon) + p_2(1 + \varepsilon)) < 1$$

Hence, there exists no $x \neq 0$ such that $(I - A_1z - A_2z^2 \cdots - A_rz^r)x = 0$. Therefore, (2.4) is satisfied. To show $\{\phi_i\}_{i=0}^\infty \in \mathfrak{F}(\beta, \lambda)$ for any $0 < \beta < 1$ and $\lambda = 0$, we consider the autoregressive processes

$$\begin{aligned} y_t &= \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_r y_{t-r} + e_t \\ z_t &= \theta'_1 z_{t-1} + \theta'_2 z_{t-2} + \cdots + \theta'_r z_{t-r} + e_t \end{aligned}$$

where e_t , $t = 1, 2, \dots$ are independently distributed with mean 0 and variance σ^2 . If $\{A_i\}_{i=1}^r \in \mathcal{P}(c, \alpha, \varepsilon)$, Then we have the representations,

$$y_t = \sum_{i=0}^{\infty} \alpha_i e_{t-i} \quad \text{and} \quad z_t = \sum_{i=0}^{\infty} \beta_i e_{t-i} \quad \forall t$$

where

$$\alpha_j = \sum_{i=1}^j \theta_i \alpha_{j-i} \quad \text{and} \quad \beta_j = \sum_{i=1}^j \theta'_i \beta_{j-i} \quad \forall j$$

and there exist $0 < \delta < 1$, $c > 0$, such that

$$\max(\alpha_i, \beta_i) < c\delta^i \quad \forall i.$$

Therefore, using Lemma 6.1(i) repeatedly, we have $\|\phi_i\|_{(1,1)} < \alpha_i$ and hence

$$\|\phi_i\|_{(1,1)} < c\delta^i \quad \forall i \quad \text{for some } c > 0, 0 < \delta < 1.$$

Therefore, $\{\phi_i\}_{i=0}^\infty \in \mathfrak{F}(\beta, \lambda)$ for any $0 < \beta < 1$ and $\lambda = 0$.

Now consider any $i \leq r$. Then

$$\|\phi_i\|_{(1,1)} \leq \sum_{j=1}^i \|A_j\|_{(1,1)} \|\phi_{i-j}\|_{(1,1)} < c\delta^i.$$

Hence,

$$\|A_i\|_{(1,1)} < c\delta^i, \quad 1 \leq i \leq r.$$

Since $A_i \in \mathfrak{F}_T(c, \alpha) \quad \forall i \leq r$, we have $T(A_i, t) < C_1 \delta^i t^{-\alpha}$, for some $C_1 > 0$. Note that

$$T(\phi_1, (1 + \eta)t) < ct^{-\alpha} \delta (1 + \eta^{-\alpha}).$$

We now apply induction. Suppose,

$$T(\phi_j, (1 + \eta + \cdots + \eta^j)t) < ct^{-\alpha} (1 + \eta^{-\alpha} + \cdots + \eta^{-j\alpha}) \delta^j j.$$

Then, by Lemma 6.1 (iii), for all $j > k$,

$$\begin{aligned} T(A_k \phi_{j-k}, (1 + \eta + \cdots + \eta^j)t) &\leq \delta^k ct^{-\alpha} (1 + \eta^{-\alpha} + \cdots + \eta^{-(j-k)\alpha}) \delta^{j-k} j + \delta^{j-k} ct^{-\alpha} \eta^{-(j-k+1)\alpha} \delta^k \\ &\leq ct^{-\alpha} \delta^j j (1 + \eta^{-\alpha} + \cdots + \eta^{-j\alpha}). \end{aligned}$$

Since, $\phi_{j+1} = \sum_{i=0}^{j+1} A_i \phi_{j-i+1}$, for some $C' > 0$,

$$T(\phi_{j+1}, (1 + \eta + \cdots + \eta^{j+1})t) \leq C' t^{-\alpha} \delta^j j (1 + \eta^{-\alpha} + \cdots + \eta^{-(j+1)\alpha}).$$

Hence the proof is complete. \square

To prove Theorem 4.1, we need the following Lemmas.

Lemma 6.3. Let $\{a_j\}_{j=0}^{\infty}$ be any sequence of positive real members such that $\sum_{j=0}^{\infty} a_j^{\beta} < \infty$ and

$\sum_{j=0}^{\infty} a_j^{2(1-\beta)} j^{\lambda} < \infty$, for some $\lambda > 0$ and $0 < \beta < 1$. Then

$$p^2 \sum_{1 \leq i, j < \infty} \left(\frac{1}{p} \right)^{\frac{M}{(a_i a_j)^{2(1-\beta)}}} \rightarrow 0$$

as $p \rightarrow \infty$ for an appropriately chosen $M > 0$.

Proof. Let $\nu = 2(1 - \beta)$. As $\sum_{j=1}^{\infty} a_j^{\nu} j^{\lambda} < \infty$, we have $\frac{1}{a_j^{\nu}} > j^{\lambda}$ for all $j > N$ and some $N \geq 1$. Now

$$p^2 \sum_{i, j} p^{-\frac{M}{(a_i a_j)^{\nu}}} \leq \sum_{1 \leq i, j \leq N} p^{-\frac{M}{(a_i a_j)^{\nu} + 2}} + p^2 \sum_{\{1 \leq i, j \leq N\}^c} p^{-M(ij)^{\lambda}}.$$

As we have finitely many terms, for an appropriately chosen large M , the first term tends to 0 as $p \rightarrow \infty$. Now,

$$\begin{aligned} p^2 \sum_{\{1 \leq i, j \leq N\}^c} p^{-M(ij)^{\lambda}} &\leq p^2 \sum_{k=N}^{\infty} (k - N + 1) p^{-Mk^{\lambda}} \leq C_1 p^2 \sum_{r=R}^{\infty} r^{\frac{1}{\lambda}} p^{-Mr} \\ &\leq C_1 p^{2-MR} \sum_{r=0}^{\infty} (r + R)^{\frac{1}{\lambda}} p^{-Mr} \leq C_2 p^{2-MR} \end{aligned}$$

for an C_1, C_2 and $R > 0$. This tends to 0 for an appropriately chosen large $M > 0$. Hence the proof is complete. \square

Lemma 6.4. If $\{\psi_j\}_{j=0}^{\infty} \in \mathcal{G}$, then for all $t > 0$ and some $c' > 0$,

$$T(\Gamma_k, t) < c' t^{-\alpha} \|\Sigma_p\|_{(1,1)}.$$

Proof. Let $\delta_p = \|\Sigma_p\|_{(1,1)}$. From the properties of \mathcal{G} and Lemma 6.1, for some $C_1 > 0$,

$$\begin{aligned} T(\psi_j \Sigma_p, (1 + \eta + \dots + \eta^{j+1})t) &\leq \delta_p c t^{-\alpha} (1 + \eta^{-\alpha} + \dots + \eta^{-j\alpha}) r_j j^{\nu} + r_j c t^{-\alpha} \eta^{-(j+1)\alpha} \\ &\leq C_1 t^{-\alpha} \delta_p (1 + \eta^{-\alpha} + \dots + \eta^{-(j+1)\alpha}) r_j j^{\nu}. \end{aligned}$$

Again, from Lemma 6.1 and for some $C_2 > 0$,

$$\begin{aligned} T(\psi_j \Sigma_p \psi_{j-k}, 2(1 + \eta + \dots + \eta^{j+1})t) &\leq r_{j-k} C_1 t^{-\alpha} \delta_p (1 + \eta^{-\alpha} + \dots + \eta^{-(j+1)\alpha}) r_j j^{\nu} \\ &\quad + \delta_p r_j c t^{-\alpha} (1 + \eta^{-\alpha} + \dots + \eta^{-(j+1)\alpha}) r_{j-k} j^{\nu} \\ &\leq C_2 t^{-\alpha} \delta_p (1 + \eta^{-\alpha} + \dots + \eta^{-(j+1)\alpha}) r_j r_{j-k} j^{\nu}. \end{aligned}$$

Hence, for some $C_3 > 0$, we have

$$T(\Gamma_k, \frac{2}{1-\eta}t) < C_3 t^{-\alpha} \delta_p \sum_{j=k}^{\infty} \frac{r_j r_{j-k}}{\eta^{\alpha j}} j^{\nu}.$$

Hence the proof is complete. \square

For any random variable ξ , define

$$\Gamma_k(\xi) = \frac{d^k}{dt^k} \left(\log E \left(e^{it\xi} \right) \right)_{t=0}$$

whenever $E|\xi|^k < \infty$.

Lemma 6.5. [Saulis and Statulevičius, 1991] Suppose $E\xi = 0$ and there exist $\gamma \geq 0$, $H > 0$ and $\bar{\Delta} > 0$ such that

$$|\Gamma_k(\xi)| \leq \left(\frac{k!}{2} \right)^{1+\gamma} \frac{H}{\bar{\Delta}^{k-2}}, \quad k = 2, 3, 4, \dots$$

Then for all $x \geq 0$,

$$P[\pm \xi \geq x] \leq e^{-\frac{x^2}{2} \left(H + x \bar{\Delta}^{-\frac{1}{2\gamma+1}} \right)^{-\frac{2\gamma+1}{\gamma+1}}}.$$

Lemma 6.6. Suppose, $\{\varepsilon_t\}$ are i.i.d. $N(0, \Sigma_p)$. Then

$$(i) \quad \left\| \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon'_t - \Sigma_p \right\|_\infty = O_p \left(\sqrt{n^{-1} \log p} \right),$$

$$(ii) \quad \left\| \frac{1}{n-k} \sum_{t=1}^n \varepsilon_t \varepsilon'_{t-k} \right\|_\infty = O_p \left(\sqrt{n^{-1} \log p} \right).$$

Proof. (i) follows from [Bickel and Levina, 2008]. For (ii) Let, $z_{t,i} = \frac{\varepsilon_{t,i}}{\sqrt{\Gamma_{0,ii}}}$ $\forall i, t$, where $\Gamma_{0,ii}$ is the (i, i) th entry of Γ_0 . Then for some $c_1 > 0$,

$$P \left[\left\| \frac{1}{n-k} \sum_{t=k+1}^n \varepsilon_t \varepsilon'_{t-k} \right\|_\infty > t \right] \leq \sum_{l,m} \sum_{j=1}^k p \left[\left| \sum_{i \geq 0} \left\{ \frac{(z_{j+(k+1)i,l} + z_{j+(k+1)i,m})^2}{2} - 1 \right\} > c_1 n t \right| \right. \\ \left. + \sum_{l,m} \sum_{j=1}^k p \left[\left| \sum_{i \geq 0} \left\{ \frac{(z_{j+(k+1)i,l} + z_{j+(k+1)i,m})^2}{2} - 1 \right\} > c_2 n t \right| \right]. \right.$$

Since, $\frac{(z_{j+(k+1)i,l} + z_{j+(k+1)i,m})^2}{2}$, $i \geq 0$ are all independent χ_1^2 variables, by Lemma 6.5, for some $c_2, c_3 > 0$

$$P \left[\left\| \frac{1}{n-k} \sum_{t=k+1}^n \varepsilon_t \varepsilon'_{t-k} \right\|_\infty > t \right] \leq c_3 p^2 k e^{-c_2 n t^2}$$

which tends to 0 as $n \rightarrow \infty$ for $t = M \sqrt{n^{-1} \log p}$ and appropriately chosen $M > 0$. Hence, (ii) is proved. \square

The Gaussianity assumption in Lemma 6.6 is removed in Lemma 6.8. As a prelude to that we need the following lemma. For $n \geq 1$, let U_1, U_2, \dots, U_n be independent random variables with

$$EU_j = 0 \quad \text{and} \quad \sigma_j^2 = \text{Var}(U_j) > 0, \quad j = 1, 2, \dots$$

Set

$$S_n = \sum_{j=1}^n U_j \quad \text{and} \quad B_n^2 = \sum_{j=1}^n \sigma_j^2, \quad Z_n = \frac{S_n}{B_n}.$$

We say that $\{U_j\}$ ($j \in N$) satisfies condition (P), if there exist positive constants A, C, C_1, C_2, \dots such that for all $|z| < A$ and $j \in N$,

$$\left| \frac{\ln E(e^{zU_j})}{z^2} \right| \leq C_j^2 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{j=1}^n C_j^2 \leq C. \quad (6.1)$$

Lemma 6.7. *Suppose a sequence of random variables $\{U_j\}$ with $EU_j = 0$ and $\sigma_j^2 = \text{Var}(U_j) > 0$ satisfies condition (P). Then there exist some $A, C > 0$ such that*

$$|\Gamma_k(Z_n)| \leq \frac{k!C}{(AB_n)^{k-2}} \quad \forall k \geq 3$$

that is, for the random variable $\xi = Z_n$, the conclusion of Lemma 6.5 holds with $v = 0, H = 2C, \bar{\Delta} = AB_n$. In particular, if U_i are i.i.d. then, (6.1) holds if

$$\left| \frac{\ln E(e^{zU_1})}{z^2} \right| \leq C, \quad \text{for all } |z| < A, \quad \text{for some } A, C > 0. \quad (6.2)$$

Also for a random variable U with $EU = 0$, if there exists $A', C' > 0$ such that $E(e^{\lambda U}) \leq C$ for all $|\lambda| < A$, then (6.2) holds.

Proof. The first part of the lemma is easy to show. The cumulants $\{\kappa_n\}$ of a random variable U are define by the cumulant generating function

$$g(z) = \log(E(e^{zU})) = \sum_{n=1}^{\infty} \kappa_n \frac{z^n}{n!}$$

The cumulants are related to the moments by the following recursion formula

$$\kappa_n = \mu'_n - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \kappa_m \mu'_{n-m}$$

As all the moments of U exist, $\kappa_n, n = 1, 2, \dots$ also exist. Moreover, $\kappa_1 = \mu'_1 = 0$ and $\kappa_2 = \mu'_2 - \mu'_1 = \mu'_2$. Hence,

$$\left| \frac{g(z)}{z^2} \right| \leq \frac{\mu'_2}{2!} + \sum_{n=3}^{\infty} \kappa_n \frac{|A|^{n-2}}{n!}.$$

□

Lemma 6.8. *Let $\{\varepsilon_i\}$ be i.i.d. with mean 0 and variance-covariance matrix Σ_p . If for some $|\lambda| < \lambda_0, \sup_{j \geq 1} E(e^{\lambda \varepsilon_{i,j}}) < \infty$, then (i) and (ii) of Lemma 6.6 hold.*

Proof. (i) follows from [Bickel and Levina, 2008]. For (ii), using Lemma 6.7, we need the existence of Moment generating function of $\frac{(Z_{t,j} + Z_{(t-k),l})^2}{2} - 1 \quad \forall j, l$ in some neighborhood of zero. This existence follows from the fact that $(x + y)^2 < 2(x^2 + y^2)$. □

The following lemma implies the invertibility of the matrix G_n defined by (4.3) in Section 4.

Lemma 6.9. *If $a' \Gamma_0 a > 0$ and $a' \Gamma_h \Gamma'_h a \rightarrow 0$ as $h \rightarrow \infty \quad \forall a \neq 0$, then G_n defined by (4.3) is non-singular for all n .*

Proof. Suppose that G_r is non-singular but G_{r+1} is singular. Then there exist a, a_1, a_2, \dots, a_r such that

$$a'X_{r+1} = \sum_{j=1}^r a'_j X_j.$$

By stationarity, $a'X_{r+1} = \sum_{j=1}^r a'_j X_{j+n-1} \forall h \geq 1$. So, $\forall n \geq r+1, \exists a_1^{(n)}, a_2^{(n)}, \dots, a_r^{(n)}$ such that $A^{(n)} = (a_1^{(n)}, a_2^{(n)}, \dots, a_r^{(n)})$, $X_r = (X_1, X_2, \dots, X_r)'$ and $a'X_n = A^{(n)}X_r$. Hence,

$$a'\Gamma_0 a = A^{(n)}G_r A^{(n)'} = A^{(n)}P \wedge P'A^{(n)} \geq \lambda_1 A^{(n)}PP'A^{(n)'} \geq \lambda_1 A^{(n)}A^{(n)'} = \lambda_1 \sum_{i=1}^r \|a_i^{(n)}\|_2^2$$

where, as G_r is non-singular, $G_r = P \wedge P'$ with $P'P = PP' = I$ and λ_1 is the smallest eigenvalue of G_r . Therefore, $\|a_i^{(n)}\|_2$ are bounded function of n for all i . Again,

$$\begin{aligned} a'X_n &= A^{(n)}X_r \Rightarrow a'X_n X'_n a = A^{(n)}X_r X'_n a \\ \Rightarrow a'\Gamma_0 a &= \sum_{j=1}^r a_j^{(n)'} \Gamma'(n-j)a. \end{aligned}$$

Hence,

$$|a'\Gamma_0 a| \leq \sum_{j=1}^n \|a_j^{(n)}\|_2 \|\Gamma'(n-j)\|_2 \|a\|_2 \leq C \sum_{j=1}^r \|\Gamma'(n-j)\|_2$$

for some $c > 0$ and tends to zero as $n \rightarrow \infty$. So, $a'\Gamma_0 a = 0$ for some $a \neq 0$. This is a contradiction. Hence, the result holds. \square

6.1 Proof of Theorem 4.1

From [Golub and Van Loan, 1996], pp. 83 and 84), for any square matrix A ,

$$\|A\|_2 \leq \sqrt{\|A\|_{(1,1)} \|A'\|_{(1,1)}}.$$

Using this,

$$\|B_{k_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 \leq \sqrt{\|B_{k_n}(\hat{\Gamma}_k) - \Gamma_k\|_{(1,1)} \|B_{k_n}(\hat{\Gamma}_k) - \Gamma_k\|_{(1,1)'}}.$$

Also, using triangle inequality,

$$\begin{aligned} \|B_{k_n}(\hat{\Gamma}_k) - \Gamma_k\|_{(1,1)} &\leq \|B_{k_n}(\hat{\Gamma}_k) - B_{k_n}(\Gamma_k)\|_{(1,1)} + T(\Gamma_k, k_n) \\ &\leq (2k_n + 1) \|\hat{\Gamma}_k - \Gamma_k\|_\infty + T(\Gamma_k, k_n). \end{aligned}$$

From Lemma 6.4,

$$T(\Gamma_k, k_n) = O(k_n^{-\alpha} \|\Sigma_p\|_{(1,1)}).$$

Using Model (2.1),

$$\|\hat{\Gamma}_k - \Gamma_k\|_\infty \leq \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} r_i r_j \left\| \frac{1}{n-k} \sum_{t=k}^n \varepsilon_{t,j} \varepsilon'_{(t-k),i} - E_{ij} \right\|_\infty$$

where $E_{ij} = E\varepsilon_{t,j}\varepsilon'_{(t-k)i} \forall i, j$. Hence for some $C_1 > 0$,

$$\begin{aligned} P[\|\hat{\Gamma}_k - \Gamma_k\|_\infty > t] &\leq P\left[\sum_j \sum_i r_i r_j \left\| \frac{1}{n-k} \sum_{t=k}^n \varepsilon_{t,j} \varepsilon'_{(t-k)i} - E_{ij} \right\|_\infty > \sum_j \sum_i \frac{C_1 t}{r_i^{-\beta} r_j^{-\beta}}\right] \\ &\leq \sum_j \sum_i P\left[\left\| \frac{1}{n-k} \sum_{t=k}^n \varepsilon_{t,j} \varepsilon'_{(t-k)i} - E_{ij} \right\|_\infty > \frac{C_1 t}{r_i^{1-\beta} r_j^{1-\beta}}\right]. \end{aligned}$$

Now, by Lemma 6.6 and 6.8, for $t = M \sqrt{n^{-1} \log p}$,

$$P[\|\hat{\Gamma}_k - \Gamma_k\|_\infty > t] \leq p^2 \sum p^{-\frac{M}{(r_i r_j)^{2(1-\beta)}}}.$$

By Lemma 6.3, this tends to zero as $n \rightarrow \infty$. Hence, $\|\hat{\Gamma}_k - \Gamma_k\|_\infty = O_p(\sqrt{n^{-1} \log p})$ and the proof follows.

6.2 Proof of Theorem 4.3

Using [Bhatia, 2009], for large n ,

$$\|(B_{k_n}(\hat{\Gamma}_0))^{-1} - \Gamma_0^{-1}\| \leq \frac{\|\Gamma_0^{-1}\|_2^2 \|B_{k_n}(\hat{\Gamma}_0) - \Gamma_0\|_2}{1 - \|\Gamma_0^{-1}\|_2 \|B_{k_n}(\hat{\Gamma}_0) - \Gamma_0\|_2}.$$

If $n^{-1} \log p \rightarrow 0$, then for some $C > 0$ and for sufficiently large n ,

$$\|(B_{k_n}(\hat{\Gamma}_0))^{-1} - \Gamma_0^{-1}\|_2 \leq C \|B_{k_n}(\hat{\Gamma}_0) - \Gamma_0\|_2.$$

So, $\|(B_{k_n}(\hat{\Gamma}_0))^{-1} - \Gamma_0^{-1}\|_2 = O_p(\|\Sigma_p\|_{(1,1)} k_n^{-\alpha})$. Again, by the fact

$$\|AB - CD\|_2 \leq \|A - C\|_2 \|B - D\|_2 + \|A - C\|_2 \|D\|_2 + \|C\|_2 \|B - D\|_2$$

and using $A = \Gamma_1 \Gamma_0^{-1}$, (i) follows. Result (ii) is immediate from the relation

$$\Gamma_0 - A\Gamma_1' - \Gamma_1 A' + A\Gamma_0 A' = \Sigma.$$

References

- L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers., Dordrecht, The Netherlands, 1988.
- R. Bhatia. *Notes on Functional Analysis*. Hindustan Book Agency, India, 2009.
- M. Bhattacharjee and A. Bose. Consistency of large dimensional sample covariance matrices under weak dependence. *Statistical Methodology*, to appear, 2013. doi: 10.1016/j.stamet.2013.08.005.

- P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008.
- P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting, Second Edition*. Springer text in Statistics, New York, 2002.
- A. Chudik and M. H. Pesaran. Infinite-dimensional vars and factor models. *Journal of Econometrics*, 163(1):4–22, 2011.
- T. Doan, R. Litterman, and C. Sims. Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews*, 3:1–100, 1984.
- M. Forni and M. Lippi. The generalized factor model: Representation theory. *Econometric Theory*, 17:1113–1141, 2001.
- M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, 82:540–554, 2000.
- M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics*, 119:231–235, 2004.
- G. H. Golub and C. F. Van Loan. *Matrix Computations, 3rd ed.* The Johns Hopkins University Press, Baltimore, 1996.
- E. Hannan. *Multiple Time Series*. John Wiley and Sons, Inc., New York, 1970.
- L. Saulis and V. A. Statulevičius. *Limit Theorems for Large Deviations*. Kluwer Academic Publishers, Dordrecht, 1991.