Cotrending: testing for common deterministic trends in varying means model *†‡

Marie-Christine Düker
Ruhr-Universität Bochum

Vladas Pipiras
University of North Carolina

Raanju R. Sundararajan
Southern Methodist University
King Abdullah University of Science and Technology

May 13, 2019

Abstract

In a varying means model, the temporary evolution of a \( p \)-vector system is determined by \( p \) deterministic nonparametric functions superimposed by error terms, possibly dependent cross sectionally. The basic interest is in linear combinations across the \( p \) dimensions that make the deterministic functions constant over time. The number of such linearly independent linear combinations is referred to as a cotrending dimension, and their spanned space as a cotrending space. This work puts forward a framework to test statistically for cotrending dimension and space. Connections to principal component analysis and cointegration are also considered. Finally, a simulation study to assess the finite-sample performance of the proposed tests, and applications to several real data sets are also provided.

1 Introduction

The topic and the results of this work can be viewed from several interesting angles. We shall first describe the problem in more technical terms and then discuss its connections to other lines of investigation. We are interested here in a statistical model of the form

\[
X_t = \mu \left( \frac{t}{T} \right) + Y_t, \quad t = 1, \ldots, T.
\]  

(1.1)

Here, \( t \) is thought as time, the observations \( X_t \) are \( p \)-vectors, \( \mu : [0, 1] \to \mathbb{R}^p \) is a \( p \)-vector deterministic function with component functions \( (\mu_1(u), \ldots, \mu_p(u))^\prime \) and \( Y_t \) are \( p \)-vector i.i.d. error terms with \( E Y_t = 0 \). We shall further assume that the covariance matrix of the error terms \( E Y_t Y_t' \) may vary with time, and also treat the simpler special case when it does not, that is, \( E Y_t Y_t' = \Sigma \), separately – the reader may have this case in mind for the rest of this section. We think of (1.1) as modeling


†Keywords and phrases: varying means, cotrending dimension and space, matrix rank and nullity, asymptotic normality, testing, principal component analysis, cointegration.

‡This work was carried out during a stay of the first author in the Department of Statistics and Operation Research at the University of North Carolina, Chapel Hill. The first author thanks the department for its hospitality, and also the DFG (RTG 2131) for financial support. The second author was supported in part by the NSF grant DMS-1712966.
varying means across $p$ dimensions and shall refer to (1.1) as a *varying means* or *VM model*. The mean vector function $\mu(u)$ is nonparametric and will be assumed to be piecewise continuous. The focus here is on the "fixed $p$, large $T$" asymptotics.

The question we ask here for the VM model is whether there are (linearly independent) linear combinations of the components of $X_t$ that are stationary across time $t$ at the mean level. That is, we look for a $p \times d_1$ matrix $B_1$ with linearly independent columns (which are not identically zero) such that $\mathbb{E}B_1'X_t = B_1'\mu(t/T) \equiv \mu_1$ for a constant $d_1 \times 1$ vector $\mu_1$ or, at the model level,

$$B_1'\mu(u) = \mu_1, \ u \in (0, 1]. \quad (1.2)$$

**Definition 1.1.** The largest $d_1$ for which (1.2) holds is called the *cotrending dimension* of the corresponding *cotrending subspace* $B_1$, spanned by the columns of $B_1$. Similarly, $d_2 = p - d_1$ is called the *noncotrending dimension* of the corresponding *noncotrending subspace* $B_2$, with $B_2 \perp B_1$.

The dimension $d_2$ indicates how many non-constant deterministic functions drive the system (1.1). We are interested here in inference about $d_1$ (and hence $d_2$), and that of the corresponding subspace.

To make inference about $d_1$ and $d_2$, we relate (1.2) to a problem involving matrix nullity (the number of zero eigenvalues) and rank. The matrix in question is defined based on the following observation. Under mild assumptions on $\mu$ (see Section 3 below), the relation (1.2) is equivalent to

$$B_1'MB_1 = 0, \quad (1.3)$$

where

$$M = \int_0^1 (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})'du \quad \text{with} \quad \bar{\mu} = \int_0^1 \mu(u)du. \quad (1.4)$$

The matrix $M$ is positive semidefinite. Then, according to (1.3),

$$d_1 = \text{nl}\{M\}, \ d_2 = \text{rk}\{M\}, \quad (1.5)$$

where $\text{nl}\{M\}$ and $\text{rk}\{M\}$ denote the nullity and the rank of the matrix $M$. Inference about $d_1, d_2$ is then that about the nullity or the rank of the matrix $M$. Similarly, the cotrending subspace $B_1$ is spanned by the eigenvectors associated with the zero eigenvalues of $M$.

A number of tests are available for the rank of a matrix, given an asymptotically normal estimator $\hat{M}$ of $M$, especially in the econometrics literature (Gill and Lewbel (1992), Cragg and Donald (1996), Kleibergen and Paap (2006), Robin and Smith (2000)), and slightly less so in the statistics literature (Anderson (1951), Eaton and Tyler (1991), Camba-Mendez and Kapetanios (2001)). Furthermore, there are technical reasons for $\hat{M}$ to be nondefinite, when $M$ is positive semidefinite itself, as it is the case here (Donald et al. (2007) and Section 3 below). Much of the technical contribution of this work consists of introducing such an estimator for $M$ in (1.4), proving its asymptotic normality result and then applying the available matrix rank tests. We shall also discuss what can be said about the convergence of the sample eigenvectors corresponding to the cotrending subspace $B_1$. Another less technical contribution is to relate the considered problem to a number of other lines of work, as outlined next and investigated in greater depth below.

The problem described above is related to stationary subspace analysis (SSA), which was one motivating starting point. In SSA, one similarly seeks linear combinations of vector observations collected over time that are stationary, possibly not just at the mean but also the covariance level. The SSA was introduced by Von Bünau et al. (2009), and studied further by Blythe et al. (2012), Sundararajan and Pourahmadi (2018). In the work somewhat parallel to this (Sundararajan et al. (2019)), we use similarly matrix constructs and their eigenstructure to study the SSA at the
covariance level, supposing the mean is zero, though the overall approach turns out to be much more involved than the one presented here.

This work, probably unsurprisingly to the reader, also has connections to principal component analysis (PCA). Two aspects of this connection should be highlighted here and kept in mind. Unlike in the standard PCA, to estimate $M$ in (1.4), we shall not work with the sample covariance matrix of the data but rather effectively with the autocovariance matrix at lag 1. Using such covariance matrices in PCA though is not completely new; see e.g. Lam and Yao (2012). Furthermore, from the PCA perspective, this work provides a new framework where the number of principal components can be tested for in a theoretically justified approach.

Lastly, we shall also draw connections to cointegration. Cointegration is a, if not the, approach of choice in modern time series analysis that also seeks linear combination of nonstationary time series that are stationary. Nonstationarity though is understood in the form of random walks, whereas the formulation (1.1) takes the view of deterministic trends. Similar time series realizations are nevertheless expected to be captured by either formulation. In our real data applications, we shall also contrast our approach to cointegration. The term “cotrending” used in this work is inspired by “cointegrating”. While “integrated” refers to random walks, “trended” alludes here to deterministic trends.

The rest of the paper is organized as follows. In Section 2, we introduce an estimator of $M$ in (1.4) and state its asymptotic normality results, whose proofs can be found in Appendix A. The application of some of the available matrix rank tests is discussed in Section 3. Section 4 concerns inference of the cotrending subspace $B_1$. Section 5 details connections to PCA and cointegration. A simulation study and applications are considered in Sections 6 and 7. Section 8 concludes.

2 Matrix estimator and its asymptotic normality

We introduce here an asymptotically normal estimator $\hat{M}_S$ for $M$ in (1.4) and a consistent estimator for its limiting covariance matrix. As noted in the introduction, we shall allow the covariance matrix of the error terms $Y_t$ in the VM model (1.1) to vary with time. More specifically, we set

$$Y_t = \sigma(t/T)Z_t$$

with i.i.d. vectors $Z_t$ satisfying $E Z_t = 0$, $E Z_t Z_t' = I_p$, so that the VM model (1.1) becomes

$$X_t = \mu \left( \frac{t}{T} \right) + \sigma \left( \frac{t}{T} \right) Z_t, \quad t = 1, \ldots, T,$$

where $\sigma : [0, 1] \to \mathbb{R}^{p \times p}$. In order to use the available matrix rank tests, we seek a symmetric estimator $\hat{M}_S = \hat{M}_S(T)$ of $M$ such that

$$\sqrt{T} \text{vech}(\hat{M}_S - M) \xrightarrow{d} N(0, C).$$

Furthermore, we need a consistent estimator $\hat{C} = \hat{C}(T)$ for the resulting covariance matrix $C$, that is, $\hat{C} \converges P \to C$. Throughout this work, $\xrightarrow{d}$ and $\converges P$ stand for the convergence in distribution and probability, respectively. For a matrix (or vector) $A$, we also set $A^2 = AA'$.

By Proposition 2.1 in Donald et al. (2007), a positive semidefinite estimator $\hat{M}$ of $M$ satisfying (2.2) would necessarily have a singular limiting covariance matrix $C$. Most asymptotically valid and “nondegenerate” matrix rank tests found in the literature assume that $C$ is nonsingular. Nonsingularity can often be achieved by working with an estimator of $M$ which is nondefinite. For this reason, as an estimator of $M$, we suggest the symmetrized sample autocovariance matrix at lag 1 given by

$$\hat{M}_S = \frac{1}{2}(\hat{M} + \hat{M}')$$

(2.3)
Remark 2.1. The intuition behind the estimator (2.3) is quite simple. Replacing $X_T$ by $\bar{\mu}$ for simplicity, note that
\[
E \left( \frac{1}{T} \sum_{t=1}^{T-1} (X_t - \bar{\mu})(X_{t+1} - \bar{\mu})' \right) = \frac{1}{T} \sum_{t=1}^{T-1} (\mu(t/T) - \bar{\mu})(\mu((t+1)/T) - \bar{\mu})',
\]
where we used the fact that $E Z_t Z_{t+1}' = 0$. The latter expression approximates $M$ for piecewise continuous $\mu$.

The following proposition gives a consistent estimator for the limiting covariance matrix $C$ in (2.5) and (2.6). It is proved in Appendix A.

Proposition 2.2. Under the assumptions of Proposition 2.1, the estimator
\[
\hat{C} = D_p^+ \frac{1}{T} \sum_{t=1}^{T-3} \left( \frac{1}{4} (\Delta X_{t+1})^2 \otimes (\Delta X_{t+3})^3 + 2((\Delta X_{t+3})^2 \otimes (X_t - \bar{X}_T)(X_{t+1} - \bar{X}_T)') \right) D_p^{+'} \tag{2.7}
\]
with $\Delta X_t = X_t - X_{t-1}$, is an asymptotically unbiased and consistent estimator for $C$ in (2.5).

The following corollary gives a simpler estimator for the limiting covariance matrix in Corollary 2.1.

Corollary 2.2. Under the assumptions of Proposition 2.1 and when $\sigma^2(u) \equiv \Sigma$, the estimator
\[
\hat{C} = D_p^+ (\Sigma \otimes \hat{\Sigma}) + 4(\hat{M}_S \otimes \hat{\Sigma}) D_p^{+'} \quad \text{with} \quad \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T-1} (\Delta X_{t+1})^2
\]
with $\hat{M}_S$ as in (2.4), is an asymptotically unbiased and consistent estimator for $C$ in (2.6).
As noted above, we want to have nonsingular limiting covariance matrix \( C \) of the estimator \( \hat{M}_S \) to ensure the applicability of available matrix rank tests. This motivated the choice of a nondefinite estimator \( \hat{M}_S \), yielding the limiting covariance matrix \( C \) in (2.5). Since \((\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})'\) in (2.5) might be nondefinite, nonsingularity can be achieved through \( \int_0^1 \sigma^2(u) \otimes \sigma^2(u) du \). For this reason, we require hereafter \( \sigma^2(u) \) to be nonsingular for all \( u \in (0, 1] \).

3 Inference of (non)cotrending dimension

In this section, we give more details about the application of available matrix rank tests to infer \( \text{rk}\{M\} \) (or \( \text{nI}\{M\} \)) having a matrix estimator \( \hat{M}_S \) satisfying (2.2). We start by justifying the statements in (1.5).

**Lemma 3.1.** Suppose that \( \mu \) is piecewise continuous and let \( d_1, d_2 \) be the cotrending and non-cotrending dimensions defined for \( \mu \) in Section 1. Then, the relations (1.5) hold.

**Proof:** The cotrending and noncotrending dimensions \( d_1, d_2 \) are defined in terms of the relation (1.2). Since \( \mu_1 = B'_1 \bar{\mu} \), (1.2) can be written as

\[
B'_1(\mu(u) - \bar{\mu}) = 0,
\]

which implies (1.3). The converse is a consequence of writing (1.3) as

\[
\int_0^1 (x'B'_1(\mu(u) - \bar{\mu}))^2 du = 0
\]

for any \( x \in \mathbb{R}^{d_1} \). This implies that \( x'B'_1(\mu(u) - \bar{\mu}) = 0 \) a.e. \( du \), for all \( x \). By Fubini’s theorem, \( x'B'_1(\mu(u) - \bar{\mu}) = 0 \) a.e. \( dx du \) and because of continuity in \( x \), \( x'B'_1(\mu(u) - \bar{\mu}) = 0 \) for all \( x \), a.e. \( du \). The latter implies that \( B'_1(\mu(u) - \bar{\mu}) = 0 \) a.e. \( du \). Since \( \mu \) is piecewise continuous, we also have \( B'_1(\mu(u) - \bar{\mu}) = 0 \) for all \( u \), which implies (1.2). \[\square\]

To test for the rank of the matrix \( M \), we use the so-called SVD matrix rank test proposed by Kleibergen and Paap (2006), and more precisely, its analogue for symmetric matrices found in Donald et al. (2007). The test has some advantages over other matrix rank tests, for example, the limiting covariance matrix \( C \) in (2.2) is not required to have a Kronecker product structure.

Consider the following hypothesis testing problem,

\[
H_0 : \text{rk}\{M\} = r \quad \text{vs.} \quad H_1 : \text{rk}\{M\} > r,
\]

(3.1)

where \( r = 0, \ldots, p - 1 \) is fixed. The SVD matrix rank test is based on the singular value decomposition of \( M \) as

\[
M = USU' = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} U'_{11} & U'_{12} \\ U'_{21} & U'_{22} \end{pmatrix},
\]

where \( U \) is orthogonal and the diagonal matrix \( S \) consists of the singular values of \( M \) in decreasing order. The matrices \( S_1 \) and \( U_{11} \) are of dimension \( r \times r \) and the other matrices in the above partition have corresponding dimensions. Furthermore, as in Donald et al. (2007), \( M \) can be written as

\[
M = A_r B_r + A_{r,\perp} A_r B_{r,\perp}
\]

with

\[
A_r = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 U'_{11}, \quad B_r = (I_r - (U'_{11})^{-1} U_{21}), \quad A_{r,\perp} = B_{r,\perp} = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1} (U_{22} U'_{22})^{1/2}
\]

\[\]
and
\[
\Lambda_r = (U_{22}U_{22}')^{-\frac{1}{2}} U_{22} S_2 U_{22}' (U_{22}U_{22}')^{-\frac{1}{2}}.
\]
Then, the null hypothesis \(H_0: \text{rk}\{M\} = r\) is equivalent to \(H_0: \Lambda_r = 0\). In the SVD matrix rank test, the latter hypothesis is tested having the symmetric matrix estimator \(\hat{M}_S\) in (2.3). Let \(\hat{\Lambda}_r\) be the quantity analogous to \(\Lambda_r\) but defined through \(\hat{M}_S\). By Proposition 2.1, the estimator \(\hat{M}_S\) satisfies (2.2) with the limiting covariance matrix (2.5). Then, by Donald et al. (2007), Proposition 4.1, \(\sqrt{T} \text{vech}(\hat{\Lambda}_r) \overset{d}{\to} \mathcal{N}(0, \Omega_r)\) under the hypothesis \(H_0: \Lambda_r = 0\), where \(\Omega_r = D^+_r \text{vech}(B_{r,\perp} \otimes A_{r,\perp}) D^+_r C D^+_p (B_{r,\perp} \otimes A_{r,\perp}) D^+_p\) and the matrix \(C\) is defined in (2.5). The suggested SVD test statistic is then
\[
\hat{\xi}_{\text{svd}}(r) = T \text{vech}(\hat{\Lambda}_r)' \hat{\Omega}_r^{-1} \text{vech}(\hat{\Lambda}_r).
\]
Here, \(\hat{\Omega}_r\) is defined by replacing the component matrices of \(\Omega_r\) by their sample counterparts, including \(\hat{C}\) defined by (2.7). By Theorem 4.1 in Donald et al. (2007), under \(H_0: \text{rk}\{M\} = r\) and if the matrix \(\Omega_r\) is non-singular,
\[
\hat{\xi}_{\text{svd}}(r) \overset{d}{\to} \chi^2((p - r)(p - r + 1)/2),\tag{3.2}
\]
where \(\chi^2(K)\) denotes the chi-square distribution with \(K\) degrees of freedom. Furthermore, Theorem 4.1 in Donald et al. (2007) gives \(\hat{\xi}_{\text{svd}}(r) \overset{p}{\to} \infty\) under \(H_1: \text{rk}\{M\} > r\).

The estimator of the matrix rank itself is defined as the first \(r\), starting with \(r = 0\), then \(r = 1\) and so on till \(r = p - 1\), for which the null hypothesis \(H_0: \text{rk}\{M\} = r\) is not rejected. By using the aforementioned asymptotic results, the resulting estimator can be shown to be consistent for \(\text{rk}\{M\}\) in a standard way when the significance level suitably depends on the sample size.

## 4 Inference of (non)cotrending subspace

We are interested here in inference about the cotrending subspace \(B_1\) spanned by the columns of \(B_1\) satisfying (1.2) and hence also about the noncotrending subspace \(B_2\) characterized by \(B_2 \perp B_1\). The cotrending subspace is spanned by the eigenvectors of \(M\) in (1.4) associated with the zero eigenvalues. We shall establish a consistency result for the estimated eigenvectors when using \(\hat{M}_S\) in (2.3) for \(M\), and also discuss an available method to test whether a certain set of vectors lies in \(B_1\).

Let \(\lambda_1 \geq \cdots \geq \lambda_{p-d_1} > \lambda_{p-d_1+1} = \cdots = \lambda_p = 0\) and \(\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p\) denote the eigenvalues of the symmetric matrices \(M\) and \(\hat{M}_S\), respectively. Let also \(v_j\) and \(\hat{v}_j\) be the corresponding orthonormal eigenvectors satisfying \(Mv_j = \lambda_j v_j\) and \(\hat{M}_S\hat{v}_j = \hat{\lambda}_j \hat{v}_j\). Furthermore, define the \(p \times d\) matrices
\[
V = (v_i, v_{i+1}, \ldots, v_{i+d-1}) \quad \text{and} \quad \hat{V} = (\hat{v}_i, \hat{v}_{i+1}, \ldots, \hat{v}_{i+d-1}).\tag{4.1}
\]
Observe that \(V = B_1\) when \(i = p - d_1 + 1\) and \(d = d_1\). The next result proves consistency of the eigenvectors in \(\hat{V}\). A short proof can be found in Appendix A and is based on the Davis-Kahan theorem.

**Proposition 4.1.** Suppose \(\min\{\lambda_{i-1} - \lambda_i, \lambda_{i+d-1} - \lambda_{i+d}\} > 0\), where \(\lambda_0 := \infty\) and \(\lambda_{p+1} := -\infty\). Then, there is an orthogonal matrix \(\hat{O} \in \mathbb{R}^{d \times d}\) such that
\[
||\hat{V} \hat{O} - V||_F = O_p \left(\frac{1}{\sqrt{T}}\right),
\]
where \(\hat{V}\) and \(V\) are as in (4.1).
The condition in the proposition is naturally satisfied for \( i = p - d_1 + 1 \) and \( d = d_1 \), yielding consistency of the estimated cotrending vectors \( \hat{V} = \hat{B}_1 \).

Suppose that \( \hat{r} \) is the estimated rank of the matrix \( M \) following the proposed testing procedure in Section 3. Then, the estimated cotrending dimension \( \hat{d}_1 = p - \hat{r} \) refers to the number of eigenvectors which generate the estimated cotrending subspace \( \hat{B}_1 \). We expect that \( \hat{d}_1 = d_1 \) in the asymptotic sense and that in some situations, the respective estimated eigenvectors in \( \hat{B}_1 \) will have some of the entries close to zero, suggesting that the corresponding component series are not involved in cotrending relations. For this reason, one might be interested to test if the vectors that one gets by setting these small entries to zero are still part of the cotrending subspace. To test whether a certain set of vectors lies in the cotrending subspace \( B_1 \), we use a test proposed by Tyler (1981).

Let \( \Lambda = (\lambda_1, \lambda_{i+1}, \ldots, \lambda_{i+d-1}) \) be the eigenvalues associated with the eigenvectors from \( V \) in (4.1). The total eigenprojection of \( M \) associated with \( \Lambda \) is defined as

\[
P_0 = \sum_{\lambda \in \Lambda} P_\lambda \quad \text{with} \quad P_\lambda = v_j v_j',
\]

where \( Q \) denotes a \( p \times q \) matrix with \( \text{rk}(Q) = q \leq d \). In other words, we want to test if the columns of the matrix \( Q \) lie in the subspace generated by the eigenvectors of \( M \) in \( V \) associated with the eigenvalues \( \lambda_i, \ldots, \lambda_{i+d-1} \). When \( i = p - d_1 + 1 \) and \( d = d_1 \), \( H_0 \) states that the columns of \( Q \) are in \( B_1 \).

By Proposition 2.1, the estimator \( \hat{M}_S \) satisfies (2.2) with a limiting covariance matrix \( C \) in (2.5). Using the relation \( \text{vec}(M) = D_p \text{vech}(M) \), one may also infer that \( \sqrt{T} \text{vec}(\hat{M}_S - M) \overset{d}{\rightarrow} N(0, D_p CD_p' \Sigma Q) \) holds, which coincides with the required assumptions in Tyler (1981). Furthermore, by Proposition 2.2, there is a consistent estimator \( \hat{C} \) for \( C \) and \( \Sigma \) is nonsingular. Then, by Theorem 4.1 in Tyler (1981),

\[
\text{vec}(\sqrt{T}(I_p - \hat{P}_0)Q) \overset{d}{\rightarrow} N(0, \Sigma Q)
\]

under the hypothesis \( H_0 : P_0 Q = Q \), where

\[
\Sigma Q = (Q' \otimes I_p) R_\Lambda D_p CD_p' R_\Lambda (Q \otimes I_p)
\]

with \( C \) as in (2.5) and a \( p^2 \times p^2 \) matrix \( R_\Lambda \) defined as

\[
R_\Lambda = \sum_{\lambda \in \Lambda} \sum_{\mu \notin \Lambda} \frac{1}{(\lambda - \mu)} P_\lambda \otimes P_\mu.
\]

The suggested test statistic is then defined as

\[
\hat{\gamma}_{evd}(Q) = T(\text{vec}(Q))' \hat{\Sigma}_Q^+ \text{vec}(Q),
\]

where \( \hat{\Sigma}_Q^+ \) denotes the Moore-Penrose inverse of \( \hat{\Sigma}_Q \). The matrix \( \hat{\Sigma}_Q \) is defined by replacing the component matrices of \( \Sigma Q \) by their sample counterparts, i.e. \( R_\Lambda \) is written in terms of \( \hat{V} \) and \( C \) is replaced by its consistent estimator \( \hat{C} \) given in (2.7). Then, by Theorem 5.3 in Tyler (1981), under \( H_0 : P_0 Q = Q \),

\[
\hat{\gamma}_{evd}(Q) \overset{d}{\rightarrow} \chi^2(q(p - d)).
\]
As shown in Section 6 in Tyler (1981), based on the preceding test, one can define an asymptotic 100(1 - α)% confidence region as

\[ \{ Q \mid Q \in \mathbb{R}^{p \times q}, \text{rk}(Q) = q, \hat{\gamma}_{\text{eov}}(Q) < \chi^2_{1-\alpha}(q(p - d)) \}, \]  

(4.3)

where \( \chi^2_{1-\alpha}(K) \) denotes the \( (1-\alpha) \)-quantile of a chi-square distribution with \( K \) degrees of freedom, and \( Q = \{ v \in \mathbb{R}^p \mid v = Qw \text{ for some } w \in \mathbb{R}^q \} \) is the subspace generated by \( Q \). Since the matrix \( \Sigma_Q \) depends on the matrix \( Q \) one is testing for, it has to be recalculated for each \( Q \). In fact, the confidence region in (4.3) can be expressed in terms of the estimated eigenvectors \( \hat{V} \) in (4.1) as

\[ \{ Q \mid \hat{V}'Q = I_d \text{ and } T(\text{vec}(Q - \hat{V}))'\hat{\Sigma}_V^+ \text{vec}(Q - \hat{V}) < \chi^2_{1-\alpha}(r(p - d)) \}. \]  

(4.4)


5 Connections to other approaches

We discuss here connections of the VM model and the introduced estimation framework to principal component analysis (PCA) (Section 5.1) and cointegration (Section 5.2).

5.1 Connections to PCA

It is instructive to contrast our model and approach to PCA. In PCA, one typically works with the sample covariance matrix

\[ \hat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}_T)(X_t - \bar{X}_T)', \]

which is the autocovariance function at lag 0, rather than this function at lag 1 as in (2.3) and (2.4). This has the following implications. For the VM model, we get by replacing \( \bar{X}_T \) by \( \hat{\mu} \) for simplicity,

\[
E\hat{\Gamma} = E\left( \frac{1}{T} \sum_{t=1}^{T} (X_t - \hat{\mu})(X_t - \hat{\mu})' \right) \\
= \frac{1}{T} \sum_{t=1}^{T} \left( \mu\left( \frac{t}{T} \right) - \hat{\mu} \right) \left( \mu\left( \frac{t}{T} \right) - \hat{\mu} \right)' + \frac{1}{T} \sum_{t=1}^{T} \sigma\left( \frac{t}{T} \right) \sigma\left( \frac{t}{T} \right)' \\
= M + \int_0^1 \sigma^2(u)du + O\left( \frac{1}{T} \right),
\]

that is, \( \hat{\Gamma} \) is not expected to be a consistent estimator for \( M \). Another important difference between \( \hat{\Gamma} \) and our estimator \( \hat{M}_S \), as noted above, is that \( \hat{\Gamma} \) is positive semidefinite whereas \( \hat{M}_S \) is nondefinite.

Vice versa, we also note that our estimator \( \hat{M}_S \) would not be of much interest in the many PCA scenarios that work with independent copies of the vectors \( X_t \) (e.g. Jolliffe (1986)). Indeed, for such vectors, the estimator \( \hat{M}_S \) based on the autocovariances at lag 1 would be zero asymptotically.

5.2 Connections to cointegration

In this section, we establish interesting connections of our approach to cointegration (Granger (1981), Engle and Granger (1987), Johansen (1991)). In cointegration, one similarly seeks linear
combinations of nonstationary time series that become stationary but this is for stochastic random walks (rather than deterministic trends) and with stationarity understood in a stronger sense (than just that at the mean level). We focus below on a popular class of vector error correction (VEC) models that allow for cointegration, and shall examine the behavior of our matrix estimator \( \hat{M} \) for this class of models. The obtained results will shed light on how our approach and cointegration relate.

Suppose that a \( p \)-vector time series \( X_t, \ t \in \mathbb{Z} \), follows a VAR(\( \ell \)) model

\[
X_t = \sum_{i=1}^{\ell} \Pi_i X_{t-i} + \varepsilon_t
\]

with

\[
E \varepsilon_0 = 0, \ E \varepsilon_t \varepsilon'_s = \Sigma_{\varepsilon}, \ E \varepsilon_t \varepsilon'_{s,t} = 0, \ t \neq s.
\]

It can be written in the form of a VEC model,

\[
\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{\ell-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t
\]

with

\[
\Pi = \Pi_1 + \cdots + \Pi_\ell - I_p, \ \Gamma_i = -(\Pi_{i+1} + \cdots + \Pi_\ell).
\]

Assume that \( X_t \) is at most integrated of order one, denoted as I(1), so that the first difference \( \Delta X_t = X_t - X_{t-1} \) is stationary, denoted as I(0). In this setting, Engle and Granger (1987) defined the cointegrating rank in terms of the matrix \( \Pi \) in (5.3) as

\[
r^* = \text{rk}(\Pi).
\]

The following three cases are distinguished. The case \( r^* = p \) when \( \Pi \) has full rank, corresponds to a stationary VAR series \( X_t \). When \( r^* = 0 \) or \( \Pi = 0 \), the VAR(\( \ell \)) model reduces to a VAR(\( \ell - 1 \)) model in first differences. When \( 0 < r^* < p \), the series is said to be cointegrated of order \( r^* \). In this case, the series has \( r^* \) linearly independent cointegrating relationships, that is, linearly independent vectors \( \beta_i, i = 1, \ldots, r^* \), such that \( \beta'_i X_t \) is stationary. Furthermore, the matrix \( \Pi \) in (5.3) can be written as

\[
\Pi = \alpha \beta'
\]

with matrices \( \alpha, \beta \) of dimension \( p \times r^* \) and full rank. The matrix \( \beta \) consists of \( r^* \) linearly independent columns, the cointegrating vectors \( \beta_i \). These vectors also form a basis for the cointegrating subspace.

**Remark 5.1.** Note the following curious difference between our cotrending approach and cointegration. While both cotrending dimension and cointegrating rank aim to measure similar quantities, the former is related to the nullity of a certain matrix in our approach whereas the latter is related to the rank of a matrix in cointegration. This has certainly been quite confusing to us, especially when comparing the two approaches, and should be kept in mind for the rest of this work. On the other hand, this disparity should perhaps not be surprising from the following observation: it is well known that the cointegrating rank is also the nullity of the spectral density matrix at the zero frequency of the differenced series \( \Delta X_t \) (e.g. Hayashi (2011), Maddala and Kim (1999)).

To analyze our estimator in the context of cointegration, we need some technical assumptions. The so-called Granger representation, introduced in Johansen (1991), Theorem 4.1, enables one to separate the cointegrated process into stationary and nonstationary components. Define

\[
C(z) = \Pi + \sum_{i=0}^{\ell} \Gamma_i (1-z) z^i,
\]
Proposition 5.1. Suppose that $\Gamma_0 = -I_p$. Let also $\alpha_\perp, \beta_\perp$ be the orthogonal complements of $\alpha, \beta$ in (5.5). Then, if $\det(C(z)) = 0$ has roots on or outside the unit circle and if the matrix

$$\alpha_\perp' \left( I_p - \sum_{i=1}^{\ell-1} \Gamma_i \right) \beta_\perp$$

is invertible, the time series $X_t$ has the representation

$$X_t = L \sum_{i=1}^{t} \varepsilon_i + \sum_{j=0}^{\infty} \tilde{L}_j \varepsilon_{t-j} + \tilde{X}_0,$$

(5.6)

with $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ as in (5.1) and $\tilde{X}_0$ as an initial value. Furthermore,

$$L = \beta_\perp \left( \alpha_\perp' \left( I_p - \sum_{i=1}^{\ell-1} \Gamma_i \right) \beta_\perp \right)^{-1} \alpha_\perp'$$

(5.7)

and the series $\sum_{j=0}^{\infty} \|\tilde{L}_j\|_F$ is finite, where $\| \cdot \|_F$ denotes the Frobenius norm. The first term in (5.6) is I(1) and the second one is I(0). The matrix $L$ has rank $p - r^*$ and determines the number of noncointegrating stochastic random walks.

We suppose hereafter that the assumptions on the process to admit the Granger representation are satisfied. Furthermore, we decompose the matrix $L$ into its non-zero and zero rows. Without loss of generality, we write $L = (L'_n, 0_{p-n})'$, where the subscript $n$ refers to the $n$ non-zero rows, and $p-n$ to the $p-n$ zero rows. Similarly, decompose the identity matrix into its first $n$ rows and the remaining $p-n$ rows as $I_{p \times p} = (I'_{1,n}, I'_{2,p-n})'$. The following result investigates the asymptotic behavior of the estimator $\hat{M}_S$ defined in (2.3) for a cointegrated system in (5.1). It is proved in Appendix A.

**Proposition 5.1.** Suppose that $X_t$ follows a VAR($\ell$) model (5.1) with cointegrating rank $0 \leq r^* \leq p$. Assume also that $\Sigma_\varepsilon$ in (5.2) is positive definite and $\mathbb{E} \| \varepsilon_0 \|^4 < \infty$. Then, the symmetric estimator $\hat{M}_S$ in (2.3) satisfies

$$\Delta_{2,T}^{-\frac{1}{2}} \hat{M}_S \Delta_{2,T}^{-\frac{1}{2}} \xrightarrow{d} \left( L_n \Sigma_\varepsilon^{\frac{1}{2}} Z \Sigma_\varepsilon^{\frac{1}{2}} L'_n \begin{array}{c} 0 \\ \frac{1}{2}(\Upsilon_{p-n}(1) + \Upsilon'_{p-n}(1)) \end{array} \right),$$

(5.8)

where $\Delta_{2,T} = \text{diag} (TI_n, I_{p-n})$, $\Upsilon_{p-n}(1) = I_{2,p-n} \sum_{j=0}^{\infty} \tilde{L}_j \tilde{L}_j' + I_{2,p-n}$, and

$$Z = \int_0^1 (Z(u) - \bar{Z})(Z(u) - \bar{Z})' du \quad \text{with} \quad \bar{Z} = \int_0^1 Z(u) du,$$

(5.9)

and a $p$-vector standard Brownian motion $\{Z(t)\}_{t \in [0,1]}$.

By construction, $\text{rk}\{L_n\} = \text{rk}\{L\} = p - r^*$. For this reason and since $Z$ in (5.9) and $\Sigma_\varepsilon^{\frac{1}{2}}$ are positive definite,

$$\text{rk}\{L_n \Sigma_\varepsilon^{\frac{1}{2}} Z \Sigma_\varepsilon^{\frac{1}{2}} L'_n \} = \text{rk}\{L_n\} = p - r^*.$$ 

Then,

$$\text{rk}\left( \begin{array}{cc} L_n \Sigma_\varepsilon^{\frac{1}{2}} Z \Sigma_\varepsilon^{\frac{1}{2}} L'_n & 0 \\ 0 & \frac{1}{2}(\Upsilon_{p-n}(1) + \Upsilon'_{p-n}(1)) \end{array} \right) \geq p - r^*.$$
In general, this suggests that testing for nullity with $\hat{M}_S$ in the cotrending approach will yield estimates smaller than the true cointegrating rank. However, when the matrix $L$ is assumed to have only non-zero rows, the convergence result (5.8) for $\hat{M}_S$ reduces to

$$\frac{1}{T} \hat{M}_S \xrightarrow{d} L\Sigma_{\varepsilon}^{\frac{1}{2}} Z\Sigma_{\varepsilon}^{\frac{1}{2}} L'.$$

Then, the corresponding rank is given by

$$\text{rk}\{L\Sigma_{\varepsilon}^{\frac{1}{2}} Z\Sigma_{\varepsilon}^{\frac{1}{2}} L'\} = \text{rk}\{L\} = p - r^*,$$

that is, the cotrending approach will tend to produce the same estimates as the cointegrating rank. In practice, one can ensure this condition on $L$ by multiplying the VAR model with a random matrix $R \in \mathbb{R}^{p \times p}$ with full rank, since $\text{rk}\{RL\} = \text{rk}\{L\}$. As long as $L$ is “not too sparse,” this ensures that the matrix $RL$ has only non-zero rows.

While the above discussion (and subsequent numerical results) argues that the cotrending approach will tend to give the cointegrating rank for cointegrated system, the converse is not necessarily expected as we illustrate in Section 6 below. We also note that the discussion above also holds for $r^* = 0$, that is, the situation associated with a spurious regression of independent random walks. Thus, in this case, the cotrending approach will tend to estimate the cotrending dimension $r^* = 0$ as well.

### 6 Simulation study

We use here Monte Carlo simulations to assess the performance of our cotrending test and to compare it to a cointegrating test. For the cotrending test, we formulate the hypothesis testing problem (3.1) as

$$H_0 : d_1 = d \quad \text{vs.} \quad H_1 : d_1 < d,$$

where $d = 1, \ldots, p$. The sequential testing here starts with $d = p$, then $d = p - 1$ and so on, till the null hypothesis is not rejected. To test for the cointegrating rank $r^*$ in (5.4), we apply the widely used Johansen test (Johansen (1991)). The corresponding hypothesis testing problem can be written as

$$H_0 : r^* = r \quad \text{vs.} \quad H_1 : r^* > r,$$

where $r = 0, \ldots, p - 1$. The sequential testing is carried out for $r = 0, r = 1$, etc. We present the simulation results in PP-plots as follows. Due to the different hypothesis testing problems, we present $p + 1$ plots for different values $d = 1, \ldots, p$ and $r = 0, \ldots, p - 1$. The probability $\alpha \in (0, 1)$ on the vertical axis is plotted versus $p_l(\alpha) = P(\xi_l(l) > q_l(\alpha))$, $l = d, r$ on the horizontal axis. The values $q_l(\alpha)$ are such that $P(\xi_l(l) > q_l(\alpha)) = \alpha$. The respective test statistic $\hat{\xi}_l(l)$ either coincides with $\hat{\xi}_\text{sed}(l)$ in (3.2) or the Johansen test statistic. The probability $p_l(\alpha)$ is estimated with 500 Monte Carlo replications of the corresponding test statistics. The critical values for the Johansen test are approximated as proposed in Johansen (1988), p. 239.

As the first numerical example, we consider $X_t$ from the VM model (1.1) with $p = 5$, $T = 500$ and

$$\mu(u) = (0, 7, 14, \sin(7u), \sin(7(u + 0.2)))'.$$

The errors $Y_t$ are multivariate Gaussian i.i.d. with $E Y_t Y_t' = I_5$. The true cotrending dimension for (6.1) is $d_1 = 3$. Observe from Figure 1 that the Johansen test rejects the considered hypotheses all the time, thus settling on $r^* = p = 5$ and suggesting that the series is stationary. The cotrending
Figure 1: PP-plots for a simulated VM model with true cotrending dimension $d_1 = 3$. 
test rejects the hypothesis for $d_1 = 4, 5$ and detects the cotrending dimension $d_1 = 3$ with the size matching the nominal value quite well, since dashed line lies close to the 45° line for smaller $\alpha$.

For a second example, we simulate a three dimensional VAR(2) model with true cointegrating rank $r^* = 2$ and sample size $T = 500$. The model in (5.1) reduces to

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \varepsilon_t.$$  \hfill (6.2)

The series $\varepsilon_t$ is simulated as a multivariate Gaussian i.i.d. series and the coefficient matrices are chosen as

$$\Pi_1 = \begin{pmatrix} 0.5 & 0.2 & 0 \\ -0.2 & -0.5 & 0.7 \\ 0.3 & 0 & -0.1 \end{pmatrix}, \quad \Pi_2 = \begin{pmatrix} 0.5 & -0.2 & 0 \\ -0.1 & 0.3 & -0.2 \\ 0.7 & 0.1 & -0.5 \end{pmatrix}.$$  \hfill (6.3)

The true cointegrating rank is $r^* = 2$, since

$$\Pi = -(I_3 - \Pi_1 - \Pi_2) = \begin{pmatrix} 0 & 0 & 0 \\ -0.3 & -1.2 & 0.5 \\ 1 & 0.1 & -1.4 \end{pmatrix}.$$  \hfill (6.4)

has rank 2. Observe from Figure 2 that both tests detect the cointegrating rank mostly correctly, though the Johansen test is quite undersized for this example.

In summary, the proposed cotrending test works in both examples as expected, while the cointegrating test certainly does not detect the cotrending dimension. The latter result was expected, since the simulated data in the first example appear stationary.

7 Applications

In this section, we apply the testing procedures proposed in Sections 3 and 4 to estimate the cotrending dimension and to make inference about the cotrending space in two real data sets. For comparison, we also apply the Johansen test to estimate the cointegrating rank.

The first data set concerns consumption in the United Kingdom. Three different variables are considered: the real consumption expenditure, the real income and the real wealth. The three series make part of the Raotbl3 data set of the R package urca (Pfaff (2008)). Previous works on cointegrated time series have used this data set; see Holden and Perman (1994) and Pfaff (2008). The data are quarterly, from the fourth quarter in 1966 to the second quarter in 1991. The time plot of the three series is given in the left plot of Figure 3. From bottom to top, the time series represent consumption expenditure, income and wealth. Due to their similar temporal patterns, one might expect a relationship between consumption and income. Our testing procedure estimates the cotrending dimension and the cotrending space vector as

$$\hat{d}_1 = 1 \quad \text{and} \quad \hat{B}_1 = (0.7349 \quad -0.6758 \quad -0.0571)'.$$  \hfill (7.1)

(We used a 5% significance level in sequential testing for $\hat{d}_1$.) As expected, the weights 0.7349 and $-0.6758$ are larger for the first two series (consumption and income). Since the third component of the vector $\hat{B}_1$ in (7.1) is close to zero, one might suspect that the vector

$$Q = (0.7349 \quad -0.6758 \quad 0)'$$

is an element of the underlying true cotrending subspace. In terms of the notation and procedure in Section 4, however, the hypothesis $H_0 : P_0 Q = Q$ in (4.2) is rejected at a 5% significance level.
Figure 2: PP-plots of a simulated VAR(2) model with true cointegrating rank $r^* = 2$.

Figure 3: The time plots of the quarterly consumption series in the United Kingdom from 1967 to 1991 (left hand), and the daily closing price series of three different ETF baskets in 2015 (right plot).
Figure 4: Visualization of the asymptotic 0.95 confidence region with respect to $\hat{B}_1$ in (7.1) for the first data example (Consumption data of the United Kingdom).

Figure 4 shows the asymptotic 95% confidence region (4.4) computed from the estimated vector $\hat{V} = \hat{B}_1$. Observe that the first and second components (represented by the $x$- and $y$-axes) of a cotrending vector $Q$ in the confidence region are non-zero. Even though the third component (represented by the $z$-axis) takes values close to zero, it cannot be set to zero either. For this reason, all three time series are part of the cotrending relation.

The cointegrating rank $r^*$ estimated by the Johansen test, at a 5% significance level, is $\hat{r}^* = 1$, and coincides with the estimated cotrending dimension. The corresponding cointegrating vector, normalized to the first component of the cotrending vector in (7.1), is

$$(0.7349 \ -0.6837 \ -0.0460)^\prime.$$ 

The second and third components are also similar to those of the cotrending vector in (7.1).

The second data set concerns five different ETF baskets, namely SPY, IVV, VOO, VBK and QQQ, available from finance.yahoo.com through the R package quantmod (Ryan and Ulrich (2018)). The considered ETFs track the US S&P 500 stock market index. The time plot of the five series is given in the right plot of Figure 3. The data set consists of the daily closing prices for the period Jan 01, 2015 to Dec 31, 2015. Proceeding as for the first data set above, we estimate the cotrending dimension and the cotrending space vectors as

$$\hat{d}_1 = 2 \text{ and } \hat{B}_1 = \begin{pmatrix} 0.8061 \ -0.4368 \ -0.3993 \ 0.0002 \ 0.0002 \\ -0.0026 \ 0.6721 \ -0.7404 \ 0.0008 \ 0.0069 \end{pmatrix}. \quad (7.2)$$

Replacing the small entries of the matrix $\hat{B}_1$ in (7.2) with zero, leads to

$$Q = \begin{pmatrix} 0.8061 \ -0.4368 \ -0.3993 \ 0 \ 0 \\ 0 \ 0.6721 \ -0.7404 \ 0 \ 0.0069 \end{pmatrix},$$

which could naturally be tested to lie in the cotrending subspace $B_1$. Testing for this through the hypothesis in (4.2) at a 5% significance level, the null is not rejected. As a result, one gets two cotrending relations, each involving three data series.
Applying the Johansen test to estimate the cointegrating rank $r^*$ yields $\hat{r}^* = 2$, which again coincides with the estimated cotrending dimension. The corresponding cointegrating vectors are estimated as

$$
\begin{pmatrix}
0.8061 & -0.4327 & -0.4042 & 0.0005 & 0.0005 \\
0.0507 & 0.6721 & -0.7992 & 0.0012 & 0.0078 \\
0.0507 & 0.6721 & -0.7992 & 0.0012 & 0.0078
\end{pmatrix},
$$

which are normalized with respect to the estimated cotrending space vectors in (7.2).

8 Conclusions

In this work, we proposed a modeling framework for $p$-vector time series exhibiting deterministic trends that allows testing about linear combinations across the $p$ series which have constant means over time. The methodology could be viewed as an alternative to cointegration analysis that concerns stochastic trends.

Related to the last point, in particular, several other general remarks should be made. Possible advantages of the cotrending approach over cointegration are its relative simplicity and nonparametric nature, with deterministic trends even allowed to be discontinuous. A possible present disadvantage of the cotrending approach is its perhaps oversimplified model. But we view this model as foundational in considering more elaborate models. For example, one could try to incorporate temporal dependence in errors $Y_t$ and we expect that in this case, a suitable estimator of $M$ should involve the average of autocovariance functions over more lags than just one.

Yet another difference of the cotrending and the cointegration approaches is that the former makes no implications about existing long-term equilibria in a system, though the VM model could in principal be used in short-term forecasting as well. Whether the lack of long-term equilibria is viewed as disadvantage is perhaps up for debate.

A Proofs

Proof of Proposition 2.1: To prove the asymptotic normality of the estimator $\hat{M}_S$ in (2.3), we first consider $\hat{M}$ in (2.4) and prove its asymptotic normality using a result on possibly nonstationary $m$-dependent random variables in Sen (1968).

The estimator $\hat{M}$ in (2.4) can be written as

$$
\hat{M} = \frac{1}{T} \sum_{t=1}^{T-1} (X_t - \bar{X}_T)(X_{t+1} - \bar{X}_T)' = R_1 - R_2 - R_3 + R_4 \tag{A.1}
$$

with

$$
R_1 = \frac{1}{T} \sum_{t=1}^{T-1} \left( \mu \left( \frac{t}{T} \right) - \bar{\mu}_T \right) \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right)',
$$

$$
R_2 = \frac{1}{T} \sum_{t=1}^{T-1} (Y_t \bar{Y}_T' + \bar{Y}_T Y_{t+1}') - \bar{Y}_T \bar{Y}_T',
$$

$$
R_3 = \frac{1}{T} \sum_{t=1}^{T-1} \left[ \bar{Y}_T \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right)' + \left( \mu \left( \frac{t}{T} \right) - \bar{\mu}_T \right) \bar{Y}_T' \right],
$$

$$
R_4 = \frac{1}{T} \sum_{t=1}^{T-1} \left[ Y_t Y_{t+1}' + Y_t (\mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T)' + (\mu \left( \frac{t}{T} \right) - \bar{\mu}_T) Y_{t+1}' \right],
$$
where $\bar{\mu}_T = \frac{1}{T} \sum_{t=1}^{T} \mu \left( \frac{t}{T} \right)$. The term $R_1$ is the deterministic part in the decomposition (A.1), $R_2$, $R_3$ will not contribute to the limit and $R_4$ will determine the normal limit. The deterministic term $R_1$ satisfies

$$R_1 = M + O \left( \frac{1}{T} \right)$$  \hspace{1cm} (A.2)

and the second term $R_2$ is such that

$$\sqrt{T} R_2 = \sqrt{T} \bar{Y}_T Y_T' + o_p(1) = o_p(1).$$ \hspace{1cm} (A.3)

The term $R_3$ satisfies

$$\sqrt{T} R_3 = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t \frac{1}{T} \sum_{t=1}^{T-1} \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right)' + \frac{1}{T} \sum_{t=1}^{T-1} \left( \mu \left( \frac{t}{T} \right) - \bar{\mu}_T \right) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t' = o_p(1)$$ \hspace{1cm} (A.4)

by the central limit theorem and since $\frac{1}{T} \sum_{t=1}^{T-1} \left( \mu \left( \frac{t}{T} \right) - \bar{\mu}_T \right) = O(\frac{1}{T})$. It thus remains to prove the asymptotic normality of $R_4$.

Consider a real matrix $\Lambda = (\lambda_{ij})_{i,j=1,...,p}$. By the Cramér-Wold theorem, it is enough to prove the asymptotic normality of $\frac{1}{T} \sum_{t=1}^{T} w_t$, where

$$w_t = \text{vec}(\Lambda)' \text{vec}(W_t), \quad W_t = Y_t Y_{t+1} + Y_t \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right)' + \left( \mu \left( \frac{t}{T} \right) - \bar{\mu}_T \right) Y_{t+1}' \cdot \frac{1}{T} \sum_{t=1}^{T} w_t.$$

The multivariate sequence $\{ W_t \}$ is $1$-dependent and so is the univariate sequence $\{ w_t \}$. Lemma 2.2 in Sen (1968) requires the moment condition $E |w_t|^{2+\delta} < \infty$ for some $\delta > 0$ and for all $t$.$^1$ Let $k = 2 + \delta$ and $c$ be a generic constant that depends on $p$ and can change from line to line. Set $W_t = (W_{ij,t})_{i,j=1,...,p}$ and let a single subscript $i$ refer to the $i$th component of the respective vector. Then,

$$E |w_t|^k = E |\text{tr}(\Lambda' W_t)|^k \leq \sum_{i,j=1}^{p} p^{2(k-1)} E |\lambda_{ij} W_{ij,t}|^k \leq c \max_{1 \leq i,j \leq p} |\lambda_{ij}|^k \sum_{i,j=1}^{p} E |W_{ij,t}|^k \leq 3^{k-1} \sum_{i,j=1}^{p} (E |Y_{i,t} Y_{j,t+1}|^k + E |Y_{i,t} (\mu_j \left( \frac{t+1}{T} \right) - \bar{\mu}_{j,T})|^k + E |(\mu_i \left( \frac{t+1}{T} \right) - \bar{\mu}_{i,T}) Y_{j,t}|^k),$$ \hspace{1cm} (A.5)

where we used Hölder’s inequality. The conclusion that the last expression is finite follows by using

$$E |Y_{i,t}|^k \leq p^{k-1} \sum_{l=1}^{p} \sup_{1 \leq i \leq T} |\sigma_{il} \left( \frac{t}{T} \right)|^k E |Z_{l,0}|^k < \infty,$$

since $E \|Z_0\|^k < \infty$, and the piecewise continuity of $\mu$ and $\sigma^2$. Combining (A.2), (A.3), (A.4), (A.5) and Lemma A.1 below, Lemma 2.2 in Sen (1968) gives

$$\sqrt{T} \text{vec}(\hat{M} - M) \overset{d}{\to} \mathcal{N}(0, \hat{C})$$

with $\hat{C}$ as in (A.7). Note that $D_{1,\frac{1}{T}} \text{vec}(A) = \text{vech}(A)$ and $D_{p,\frac{1}{T}} N_p = D_{p,\frac{1}{T}}$ with $N_p = \frac{1}{T} (I_{p^2} + K_p)$, where $K_p$ denotes the so-called commutation matrix, which transforms $\text{vec}(A)$ into $\text{vec}(A')$ for a matrix $A \in \mathbb{R}^{p \times p}$. Furthermore, $N_p \text{vec}(\hat{M}) = \text{vec}(\hat{M}_S)$. These observations yield

$$\sqrt{T} \text{vech}(\hat{M}_S - M) \overset{d}{\to} \mathcal{N}(0, C)$$

with $C$ as in (2.5).

\[\square\]

$^1$The moment condition in Sen (1968) is stated with $\delta = 1$ but the proof also works for $\delta > 0$. 

17
The next auxiliary result was used in the proof of Proposition 2.1 above.

**Lemma A.1.** Suppose that the assumptions of Proposition 2.1 hold. Let $\widehat{M}$ be the estimator in (2.4) and $R_4$ be the last term in the decomposition (A.1). Then, the covariance matrices of $\widehat{M} - M$ and $R_4$ satisfy

$$E(\text{vec}(\widehat{M} - M)(\text{vec}(\widehat{M} - M))^t) = E R_4 R_4' + o\left(\frac{1}{T}\right) = \frac{1}{T} \tilde{C} + o\left(\frac{1}{T}\right)$$  \hspace{1cm} (A.6)

with

$$\tilde{C} = \int_0^1 \sigma^2(u) \otimes \sigma^2(u) du + 2 \int_0^1 (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})' \otimes \sigma^2(u) du N_p$$

$$+ 2 \int_0^1 \sigma^2(u) \otimes (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})' du N_p.$$  \hspace{1cm} (A.7)

**Proof:** The first relation in (A.6) follows from (A.2), (A.3) and (A.4). It is thus enough to show the second relation in (A.6) concerning the covariance of $R_4$. Decompose $R_4$ into $R_4 = R_{41} + R_{42}$ with

$$R_{41} = \frac{1}{T} \sum_{t=1}^T Y_t Y_{t+1}',$$

$$R_{42} = \frac{1}{T} \sum_{t=1}^T \left[ Y_t \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right)' + \left( \mu \left( \frac{t}{T} \right) - \bar{\mu}_T \right) Y_{t+1}' \right].$$

We consider these terms separately to calculate the limiting covariance matrix.

For $R_{41}$, we write

$$E(\text{vec}(R_{41})(\text{vec}(R_{41}))') = \frac{1}{T^2} \sum_{t,r=1}^T E \left( \text{vec}(Y_t Y_{t+1}'(\text{vec}(Y_{t+1}' Y_{r+1})))' \right)$$

$$= \frac{1}{T^2} \sum_{t,r=1}^T E \left( (I_p \otimes Y_t)Y_{t+1}'Y_{r+1}'(I_p \otimes Y_t)' \right)$$

$$= \frac{1}{T^2} \sum_{t,r=1}^T E \left( (I_p \otimes Y_t)(Y_{t+1}'Y_{r+1}' \otimes 1)(I_p \otimes Y_t)' \right)$$

$$= \frac{1}{T^2} \sum_{t,r=1}^T E(Y_{t+1}'Y_{r+1}' \otimes Y_{r+1}'Y_{t+1}) = \frac{1}{T} \int_0^1 \sigma^2(u) \otimes \sigma^2(u) du + O\left(\frac{1}{T^2}\right),$$  \hspace{1cm} (A.8)

where the second equality follows by $\text{vec}(AB) = (I_q \otimes A) \text{vec}(B) = (B' \otimes I_m) \text{vec}(A)$ for an $m \times n$ matrix $A$ and an $n \times q$ matrix $B$; see Theorem 2 in Magnus and Neudecker (1999), p. 35. In the fourth equality, the relation $AB \otimes CD = (A \otimes C)(B \otimes D)$ is used.

The covariance of the term $R_{42}$ can be written as

$$E(\text{vec}(R_{42})(\text{vec}(R_{42}))') = \frac{1}{T^2} \int_0^1 (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})' \otimes \sigma^2(u) du N_p$$

$$+ \frac{1}{T^2} \int_0^1 \sigma^2(u) \otimes (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})' du N_p + O\left(\frac{1}{T^2}\right).$$  \hspace{1cm} (A.9)
since for example

\[
\frac{1}{T^2} \sum_{t,r=1}^{T} E \left[ \text{vec} \left( Y_t \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right) \right)^\prime \left( \text{vec} \left( Y_r \left( \mu \left( \frac{r+1}{T} \right) - \bar{\mu}_T \right) \right) \right)^\prime \right] \\
= \frac{1}{T^2} \sum_{t,r=1}^{T} \left( \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right) \otimes I_p \right) E(Y_tY_r^\prime) \left( \left( \mu \left( \frac{r+1}{T} \right) - \bar{\mu}_T \right)^\prime \otimes I_p \right) \\
= \frac{1}{T^2} \sum_{t=1}^{T} \left( \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right) \otimes I_p \right) \left( 1 \otimes \sigma^2 \left( \frac{t}{T} \right) \right) \left( \left( \mu \left( \frac{t+1}{T} \right) - \bar{\mu}_T \right)^\prime \otimes I_p \right) \\
= \frac{1}{T} \int_0^1 (\mu(\mu) - \bar{\mu})(\mu(\mu) - \bar{\mu})^\prime \otimes \sigma^2(\mu)du + O\left( \frac{1}{T^2} \right),
\]

where we used the same arguments as in (A.8).

**Proof of Proposition 2.2:** To prove the consistency of \( \hat{C} \) in (2.7), we use Theorem 2 in Andrews (1988), which gives sufficient conditions for the law of large numbers for \( L^1 \)-mixingales.

For simplicity, we replace \( \bar{X}_T \) with \( \bar{\mu} \) by the weak law of large numbers and decompose the resulting estimator \( \hat{C} \) as

\[
\frac{1}{T} \sum_{t=1}^{T-3} \left( \frac{1}{4}(\Delta X_{t+1})^2 \otimes (\Delta X_{t+3})^2 + 2((\Delta X_{t+3})^2 \otimes (X_t - \bar{\mu})(X_{t+1} - \bar{\mu})^\prime) \right) = A_1 + A_2 + B_1 + B_2,
\]

where setting \( \tilde{M}_t = \mu \left( \frac{t}{T} \right) \),

\[
A_1 = \frac{1}{T} \sum_{t=1}^{T-3} \frac{1}{4}(\Delta \tilde{M}_{t+1})^2 \otimes (\Delta \tilde{M}_{t+3})^2,
\]

\[
A_2 = \frac{1}{T} \sum_{t=1}^{T-3} \frac{1}{4} \left[ (\Delta \tilde{M}_{t+1})^2 \otimes \left( (\Delta Y_{t+3})^2 + \Delta Y_{t+3}(\Delta \tilde{M}_{t+3})^\prime + \Delta \tilde{M}_{t+3}(\Delta Y_{t+3})^\prime \right) \right. \\
+ \left. \left( (\Delta Y_{t+1})^2 + \Delta Y_{t+1}(\Delta \tilde{M}_{t+1})^\prime + \Delta \tilde{M}_{t+1}(\Delta Y_{t+1})^\prime \right) \otimes (\Delta \tilde{M}_{t+3})^2 \right. \\
+ \left. \left( (\Delta Y_{t+1})^2 + \Delta Y_{t+1}(\Delta \tilde{M}_{t+1})^\prime + \Delta \tilde{M}_{t+1}(\Delta Y_{t+1})^\prime \right) \otimes \left( (\Delta Y_{t+3})^2 + \Delta Y_{t+3}(\Delta \tilde{M}_{t+3})^\prime + \Delta \tilde{M}_{t+3}(\Delta Y_{t+3})^\prime \right) \right]
= \frac{1}{T} \sum_{t=1}^{T-3} W_{t,1},
\]

\[
B_1 = \frac{1}{T} \sum_{t=1}^{T-3} 2(\Delta \tilde{M}_{t+3})^2 \otimes (\tilde{M}_t - \bar{\mu})(\tilde{M}_{t+1} - \bar{\mu})^\prime,
\]

\[
B_2 = \frac{1}{T} \sum_{t=1}^{T-3} 2 \left[ \left( (\Delta Y_{t+3})^2 + \Delta Y_{t+3}(\Delta \tilde{M}_{t+3})^\prime + \Delta \tilde{M}_{t+3}(\Delta Y_{t+3})^\prime \right) \otimes (X_t - \bar{\mu})(X_{t+1} - \bar{\mu})^\prime \right. \\
+ \left. (\Delta \tilde{M}_{t+3})^2 \otimes \left( Y_tY_{t+1} + Y_t(\tilde{M}_{t+1} - \bar{\mu})^\prime + (\tilde{M}_t - \bar{\mu})Y_{t+1}^\prime \right) \right] = \frac{1}{T} \sum_{t=1}^{T-3} W_{t,2}.
\]

The deterministic terms \( A_1 \) and \( B_1 \) are asymptotically negligible, since \( A_1 = O\left( \frac{1}{T} \right) \) and \( B_1 = \)
proof. For the terms $A_2$ and $B_2$, note that

$$
E A_2 = \frac{1}{T} \sum_{t=1}^{T-3} \frac{1}{4} E \left( (\Delta \tilde{M}_{t+1})^2 \otimes (\Delta Y_{t+3})^2 + (\Delta Y_{t+1})^2 \otimes (\Delta \tilde{M}_{t+3})^2 + (\Delta Y_{t+1})^2 \otimes (\Delta Y_{t+3})^2 \right)
= \frac{1}{T} \sum_{t=1}^{T-3} \frac{1}{4} E \left( (\Delta Y_{t+1})^2 \otimes (\Delta Y_{t+3})^2 \right) + O \left( \frac{1}{T} \right) = \int_0^1 \sigma^2(u) \otimes \sigma^2(u) du + O \left( \frac{1}{T} \right),
$$

$$
E B_2 = \frac{1}{T} \sum_{t=1}^{T-3} 2E \left( (\Delta Y_{t+3})^2 \otimes (X_t - \bar{\mu})(X_{t+1} - \bar{\mu})' \right)
= 4 \int_0^1 \sigma^2(u) \otimes (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})' du + O \left( \frac{1}{T} \right).
$$

Then,

$$
\frac{1}{T} \sum_{t=1}^{T-3} (W_{t,1} - EW_{t,1}) + \frac{1}{T} \sum_{t=1}^{T-3} EW_{t,1} - \int_0^1 \sigma^2(u) \otimes \sigma^2(u) du = \frac{1}{T} \sum_{t=1}^{T-3} (W_{t,1} - EW_{t,1}) + O \left( \frac{1}{T} \right),
$$

$$
\frac{1}{T} \sum_{t=1}^{T-3} (W_{t,2} - EW_{t,2}) + \frac{1}{T} \sum_{t=1}^{T-3} EW_{t,2} - 4 \int_0^1 \sigma^2(u) \otimes (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})' du
= \frac{1}{T} \sum_{t=1}^{T-3} (W_{t,2} - EW_{t,2}) + O \left( \frac{1}{T} \right),
$$

where we subtracted the respective summands of $C$ and included the expected values of $W_{t,1}$ and $W_{t,2}$. To prove the convergence in probability, it is enough to consider $W_{t,1} - EW_{t,1}$ and $W_{t,2} - EW_{t,2}$ componentwise. We write

$$
R_{1,t} = (W_{t,1} - EW_{t,1})_{ij}, \quad R_{2,t} = (W_{t,2} - EW_{t,2})_{ij},
$$

where the subscript denotes the $ij$th component for $i, j = 1, \ldots, p^2$. The sequences $\{R_{1,t}\}$ and $\{R_{2,t}\}$ are 3-dependent and hence $L^1$-mixingales. By Theorem 2 in Andrews (1988), the uniformly integrability of $R_{1,t}$ and $R_{2,t}$ implies convergence to zero in probability of the corresponding sample means. Since,

$$
E|R_{1,t}|^{2+\delta} < \infty \quad \text{and} \quad E|R_{2,t}|^{2+\delta} < \infty \quad \text{for all} \quad t = 1, \ldots, T,
$$

by using the same arguments as in (A.5), the piecewise continuity of $\mu$ and $\sigma^2$ and the moment condition $E\|Z_0\|^{2+\delta}$ suffice to prove the uniformly integrability of $R_{1,t}$ and $R_{2,t}$. \hfill \Box

Proof of Proposition 5.1: Following the notation in Section 5.2, we decompose $X_t$ given by its Granger representation (5.6) in accordance to the zero and nonzero rows of the matrix $L$ in (5.7) into

$$
X_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} L_n Z_{1,t} + I_{1,n} Z_{2,t} + I_{1,n} \bar{X}_0 \\ I_{2,p-n} Z_{2,t} + I_{2,p-n} \bar{X}_0 \end{pmatrix},
$$

(A.10)

where

$$
Z_{1,t} = \sum_{i=1}^{t} \varepsilon_i \quad \text{and} \quad Z_{2,t} = \sum_{j=0}^{\infty} \bar{L}_j \varepsilon_{t-j}.
$$

Note that

$$
\frac{1}{T^{1/2}} \bar{X}_0 = o_p(1), \quad \frac{1}{T} \sum_{t=1}^{T} Z_{2,t} = o_p(1), \quad \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t = o_p(1),
$$

(A.11)
where the second relation follows by Proposition 6.3.10 in Brockwell and Davis (1991). As in the proof of Proposition 2.1 we first investigate the convergence result for \( \hat{M} \) in (2.4). The normalized estimator \( \hat{M} \) can be written as

\[
\Delta_{2,T}^{-1/2} \hat{M} \Delta_{2,T}^{-1/2} = \left( \frac{1}{T} R_{11} \frac{1}{T^{1/2}} R_{12} \right)
\]

where

\[
R_{ij} = \frac{1}{T} \sum_{t=1}^{T-1} (X_{i,t} - \bar{X}_i)(X_{j,t+1} - \bar{X}_j)'
\]

and \( \bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{i,t} \) for \( i, j = 1, 2 \)

with \( X_{i,t} \) for \( i = 1, 2 \) as in (A.10). We consider the terms \( R_{11}, R_{22}, R_{12} \) and \( R_{21} \) separately.

Set \( \bar{Z}_1 = \frac{1}{T} \sum_{t=1}^{T} Z_{1,t} \). Then, \( R_{11} \) can be written as

\[
\frac{1}{T} R_{11} = \frac{1}{T^2} \sum_{t=1}^{T-1} (X_{1,t} - L_n \bar{Z}_1) (X_{1,t+1} - L_n \bar{Z}_1)' + o_p(1)
\]

\[
= \frac{1}{T^2} \sum_{t=1}^{T-1} (L_n Z_{1,t} + I_{1,n} Z_{2,t} - L_n \bar{Z}_1) (L_n (Z_{1,t} + \varepsilon_{t+1}) + I_{1,n} Z_{2,t+1} - L_n \bar{Z}_1)' + o_p(1) \quad \text{(A.12)}
\]

\[
= \frac{1}{T^2} \sum_{t=1}^{T-1} L_n (Z_{1,t} - \bar{Z}_1) (Z_{1,t} - \bar{Z}_1)' \epsilon_n' + o_p(1),
\]

where the first, second and third equalities follow by (A.11) and Lemma A.2, (i), (ii) and (iv), below. Then, (A.12) and Lemma 3.1, (c) in Phillips and Durlauf (1986) yield

\[
\frac{1}{T} R_{11} = \frac{1}{T^2} \sum_{t=1}^{T} L_n (Z_{1,t} - \bar{Z}_1) (Z_{1,t} - \bar{Z}_1)' \epsilon_n' + o_p(1) \quad \xrightarrow{d} \quad L_n \Sigma_\varepsilon^{1/2} Z \Sigma_\varepsilon^{1/2} \epsilon_n'.
\]

The matrix \( R_{22} \) contains only stationary components. Its convergence

\[
R_{22} = \frac{1}{T} \sum_{t=1}^{T-1} I_{2,p-n} (Z_{2,t} - \bar{Z}_2) (Z_{2,t+1} - \bar{Z}_2)' \epsilon_{2,p-n} + o_p(1) \quad \xrightarrow{p} \quad \Upsilon_{p-n}(1)
\]

is a consequence of Lemma A.2, (iv), where \( \bar{Z}_2 = \frac{1}{T} \sum_{t=1}^{T} Z_{2,t} \) and

\[
\Upsilon_{p-n}(1) = I_{2,p-n} \text{Cov}(Z_{2,0}, Z_{2,1}) I_{2,p-n}' = I_{2,p-n} \sum_{j=0}^{\infty} \hat{L}_j \hat{L}_{j+1} I_{2,p-n}'
\]

denotes the autocovariances of order one.

The third term \( R_{12} \) satisfies

\[
\frac{1}{T^{1/2}} R_{12} = \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} (L_n Z_{1,t} + I_{1,n} Z_{2,t} - L_n \bar{Z}_1) (I_{2,p-n} Z_{2,t+1} - I_{2,p-n} \bar{Z}_2)' + o_p(1)
\]

\[
= \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} L_n (Z_{1,t} - \bar{Z}_1) (Z_{2,t+1} - \bar{Z}_2)' \epsilon_{2,p-n} + o_p(1) \quad \text{(A.13)}
\]

\[
= \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} L_n (Z_{1,t} Z_{2,t+1} - \frac{1}{T^{1/2}} \bar{Z}_1 \bar{Z}_2) I_{2,p-n} + o_p(1),
\]

21
where the first equality follows by (A.11) and the second equality by Lemma A.2, (iv). Finally, \( \frac{1}{T^{1/2}} R_{12} = o_p(1) \), since the product of the normalized sample means \( \frac{1}{T^{1/2}} \tilde{Z}_1 \tilde{Z}_2 = o_p(1) \) by Lemma A.2, (iii) and (A.11). The remaining term in the last line of (A.13) satisfies \( \frac{1}{T^{1/2}} \sum_{t=1}^{T-1} Z_{1,t} Z'_{2,t+1} = o_p(1) \) by Lemma A.2 (ii).

The result for the fourth term \( R_{21} \) follows by the same arguments as in (A.13). The convergence in distribution of the symmetric estimator \( \hat{M} \) is a consequence of the established convergence of \( \hat{M} \).

\[ \square \]

The following lemma was used in the preceding proof.

**Lemma A.2.** Set \( Z_{1,t} = \sum_{i=1}^{t} \varepsilon_i \) and \( Z_{2,t} = \sum_{j=0}^{\infty} L_j \varepsilon_{t-j} \), where \( \{ \varepsilon_t \}_{t \in \mathbb{Z}} \) is a sequence of i.i.d. random vectors satisfying (5.2) with positive definite \( \Sigma_\varepsilon \) and \( \mathbb{E} \| \varepsilon_0 \|^2 < \infty \), and \( \sum_{j=0}^{\infty} \| \tilde{L}_j \| = \infty \). Then,

\[
\begin{align*}
(i) \quad & \frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t-1} \varepsilon'_t \to_d \frac{1}{2} (\Sigma_\varepsilon^{1/2} \mathbb{E}(1) \Sigma_\varepsilon^{1/2} - \Sigma_\varepsilon), \\
(ii) \quad & \frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t} \varepsilon'_{t+1} \to_d \frac{1}{2} (\Sigma_\varepsilon^{1/2} \mathbb{E}(1) \Sigma_\varepsilon^{1/2} + \Sigma_\varepsilon) \sum_{j=0}^{\infty} L_j, \\
(iii) \quad & \frac{1}{T} \sum_{t=1}^{T} Z_{1,t} \to \Sigma_\varepsilon^{1/2} \int_0^1 \mathbb{E}(t) dt, \\
(iv) \quad & \frac{1}{T} \sum_{t=1}^{T-1} Z_{2,t} \varepsilon'_{2,t+1} - \text{Cov}(Z_{2,0}, Z_{2,1}) \to 0,
\end{align*}
\]

where \( \mathbb{E}(t) \) is a \( p \)-dimensional standard Brownian motion.

**Proof:** The statement (i) is the same as in Lemma 3.1, (d) in Phillips and Durlauf (1986).

For the convergence in (ii), set

\[
\frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t} Z'_{2,t+1} = \sum_{j=0}^{\infty} \frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t} \varepsilon'_{t+1-k} \tilde{L}_j =: \sum_{j=0}^{\infty} \bar{Y}_j(T). 
\]

Then, by Theorem 4.2 in Billingsley (1986), it is enough to prove

\[
\sum_{j=0}^{\infty} Y_j(T) \to \frac{1}{2} (\Sigma_\varepsilon^{1/2} \mathbb{E}(1) \Sigma_\varepsilon^{1/2} + \Sigma_\varepsilon) \sum_{j=0}^{\infty} \bar{L}_j 
\]

(A.14)

for each \( k \geq 1 \) and

\[
\sum_{j=k+1}^{\infty} Y_j(T) = o_p(1), \quad \text{as} \ T \to \infty, \ k \to \infty. \tag{A.15}
\]

The convergence in (A.14) is a consequence of

\[
\begin{align*}
\sum_{j=0}^{k} Y_j(T) &= \frac{k}{T} \sum_{j=0}^{k} \left( \frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t} \varepsilon'_{t+1-k} \tilde{L}_j + \frac{1}{T} \sum_{t=1}^{j-1} \varepsilon_i \varepsilon'_{t-k} \tilde{L}_j \right) \\
&= \frac{k}{T} \sum_{j=0}^{k} \left( \frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t} \varepsilon'_{t+1-k} \tilde{L}_j + \frac{1}{T} \sum_{t=1}^{j-1} \varepsilon_i \varepsilon'_{t-k} \tilde{L}_j \right) \\
&= \frac{k}{T} \sum_{j=0}^{k} \left( \frac{1}{T} \sum_{t=1}^{T-1} Z_{1,t} \varepsilon'_{t+1-k} + \Sigma_\varepsilon \right) \tilde{L}_j + o_p(1) \tag{A.16}
\end{align*}
\]
The equality (A.16) follows since \( \{\varepsilon_j\}_{j \in \mathbb{Z}} \) is stationary and ergodic, and so is any transformation of \( \varepsilon_j \). Indeed, by the ergodic theorem and since \( E \varepsilon_{t-i} \varepsilon_{t-i-j} = \Sigma_\varepsilon \) for \( l = j - 1 \) and 0 otherwise,\[ \begin{align*}
\frac{1}{T} \sum_{t=1}^{T-1} \sum_{l=0}^{j-1} \varepsilon_{t-l} \varepsilon_{t-l-j} \tilde{L}^j_t &= E \left( \sum_{l=0}^{j-1} \varepsilon_{t-l} \varepsilon_{t-l-j} \tilde{L}^j_t \right) + o_p(1) = \Sigma_\varepsilon \tilde{L}^j_t + o_p(1),
\end{align*} \]
see Theorem 2 in Hannan (1970), p. 203. For the equality (A.17) note that
\[ \begin{align*}
\frac{1}{T} \sum_{t=1}^{T-1} \sum_{l=0}^{j-1} \varepsilon_{t-l} \varepsilon_{t-l-j} \tilde{L}^j_t &= E \left( \sum_{l=0}^{j-1} \varepsilon_{t-l} \varepsilon_{t-l-j} \tilde{L}^j_t \right) + o_p(1) = \Sigma_\varepsilon \tilde{L}^j_t + o_p(1),
\end{align*} \]
(A.19)

see Theorem 2 in Hannan (1970), p. 203. For the equality (A.17) note that
\[ \begin{align*}
\frac{1}{T} \sum_{t=1}^{T-1} \sum_{l=0}^{j-1} \varepsilon_{t-l} \varepsilon_{t-l-j} \tilde{L}^j_t &= E \left( \sum_{l=0}^{j-1} \varepsilon_{t-l} \varepsilon_{t-l-j} \tilde{L}^j_t \right) + o_p(1) = \Sigma_\varepsilon \tilde{L}^j_t + o_p(1),
\end{align*} \]
(A.20)

since
\[ \begin{align*}
E \left\| \frac{1}{T} \sum_{t=r}^{s} Z_{t, t \varepsilon_{t+1}} \right\|^2 &= \frac{1}{T^2} \sum_{t_1, t_2=r}^{s} \sum_{i_1=1}^{t_1} \sum_{i_2=1}^{t_2} E \text{tr}(\varepsilon_{t_1+1, i_1} \varepsilon_{i_2} \varepsilon_{t_2+1, i_2}^\prime) \\
&= \frac{1}{T^2} \sum_{t=r}^{s} t(E \| \varepsilon_0 \|^2) = o(1),
\end{align*} \]
(A.21)

where either \( r = 1 - j \) and \( s = 0 \) or \( r = T - j \) and \( s = T - 1 \). The convergence in (A.18) is a consequence of (i).

The equality (A.15) can be proven by
\[ \begin{align*}
E \left\| \sum_{j=k+1}^{\infty} Y_j(T) \right\|^2 &= \sum_{j_1, j_2=k+1}^{\infty} \frac{1}{T^2} \sum_{t_1, t_2=1}^{T-1} \sum_{i_1=1}^{t_1} \sum_{i_2=1}^{t_2} E \text{tr}(\tilde{L}_{j_1}^j \varepsilon_{t_1-j_1} \varepsilon_{t_2-j_2}^\prime \tilde{L}_{j_2}^j) \\
&= \sum_{j_1, j_2=k+1}^{\infty} \frac{1}{T^2} \sum_{t_1, t_2=2}^{T-j_1} \sum_{i_1=1}^{t_1} \sum_{i_2=1}^{t_2} E \text{tr}(\Sigma \varepsilon \tilde{L}_{j_1}^j \tilde{L}_{j_2}^j) + 2 \sum_{j_1, j_2=k+1}^{T-j_1} \frac{1}{T^2} \sum_{k=1}^{T-j_2} \sum_{i_1=1}^{T-j_2} \sum_{i_2=1}^{T-j_2} \text{tr}(\Sigma \varepsilon \tilde{L}_{j_2}^j \tilde{L}_{j_1}^j) \\
&+ \sum_{j_1, j_2=k+1}^{\infty} \frac{1}{T^2} \sum_{l=2}^{T-j_1} \sum_{i=1}^{T-j_1} E \| \varepsilon_0 \|^2 \text{tr}(\Sigma \varepsilon \tilde{L}_{j_2}^j \tilde{L}_{j_1}^j) \\
&= \sum_{j_1, j_2=k+1}^{T+1} \frac{1}{T^2} (T - 1 + m - m)(\text{tr}(\Sigma \varepsilon \tilde{L}_{j_2}^j \tilde{L}_{j_1}^j) + E \| \varepsilon_0 \|^2 \text{tr}(\Sigma \varepsilon \tilde{L}_{j_2}^j \tilde{L}_{j_1}^j)) \\
&+ 2 \sum_{j_1, j_2=k+1}^{T+1} \frac{1}{T^2} (T - 1)^2 \text{tr}(\Sigma \varepsilon \tilde{L}_{j_2}^j \tilde{L}_{j_1}^j) \\&+ 2 \sum_{j_1, j_2=k+1}^{T+1} \frac{1}{T^2} (T - 1)^2 \text{tr}(\Sigma \varepsilon \tilde{L}_{j_2}^j \tilde{L}_{j_1}^j)
\end{align*} \]

23
\[
\leq 2 \sum_{j_1,j_2=k+1}^{T+1} \frac{1}{T} |\text{tr}(\Sigma^* \tilde{L}_{j_2} \tilde{L}_{j_1}) + E \|\varepsilon_0\|^2 \text{tr}(\Sigma_\varepsilon \tilde{L}_{j_2} \tilde{L}_{j_1})| + 2 \sum_{j_1,j_2=k+1}^{T+1} |\text{tr}(\Sigma_\varepsilon^2 \tilde{L}_{j_2} \tilde{L}_{j_1})| \to 0,
\]

as \( T \to \infty \) and \( k \to \infty \), since \( \sum_{j=0}^{\infty} \|\bar{L}_j\|_F < \infty \). Thereby, we used the notation \( \overline{m} = \min\{j_1,j_2\} \), \( \overline{m} = \max\{j_1,j_2\} \), \( \Sigma^* := E(\varepsilon_0 \varepsilon_0^T \varepsilon_0^T) \) and the fact that

\[
E(\varepsilon_{i_1} \varepsilon_{i_2}^T \varepsilon_{i_2} \varepsilon_{i_2}^T) =
\begin{cases}
\Sigma^*, & i_1 = i_2 = l_1 = l_2, \\
\Sigma_\varepsilon^2, & i_1 = l_1 \neq i_2 = l_2, \\
\Sigma_\varepsilon^2, & i_1 = l_2 \neq i_2 = l_1, \\
E(\varepsilon_0 \varepsilon_0^T \Sigma_\varepsilon), & i_1 = i_2 \neq l_1 = l_2.
\end{cases}
\]


Proof of Proposition 4.1: By Theorem 2 in Samworth et al. (2014), there is an orthogonal matrix \( \tilde{O} \in \mathbb{R}^{d \times d} \), such that

\[
\|\tilde{V} \tilde{O} - V\|_F \leq 2 \frac{\|\tilde{M}_S - M\|_F}{\min\{\lambda_{i-1} - \lambda_i, \lambda_{i+d-1} - \lambda_{i+d}\}}. \tag{A.22}
\]

Set \( \tau = \varepsilon \min\{\lambda_{i-1} - \lambda_i, \lambda_{i+d-1} - \lambda_{i+d}\} \). Then, by applying (A.22) and Chebyshev’s inequality, we get for all \( \varepsilon > 0 \),

\[
P(\|\tilde{V} \tilde{O} - V\|_F \geq \varepsilon) \leq P(\|\tilde{M}_S - M\|_F \geq \tau) = P \left( \sum_{i,j=1}^{p} |\tilde{e}_i'(\tilde{M}_S - M)e_j|^2 \right)^{\frac{1}{2}} \geq \tau \)
\[
\leq \frac{1}{\tau^2} E \sum_{i,j=1}^{p} |\tilde{e}_i'(\tilde{M}_S - M)e_j|^2
\]
\[
= \frac{1}{\tau^2} \sum_{i,j=1}^{p} (\text{vec}(\tilde{e}_i \tilde{e}_j))'(\frac{1}{T} \tilde{N}_p \tilde{C} \tilde{N}_p + o(\frac{1}{T})) \text{vec}(\tilde{e}_i \tilde{e}_j)
\]
\[
= \frac{1}{\tau^2} \frac{1}{T} (\text{tr}(\tilde{N}_p \tilde{C}) + p^2 o(1)),
\]

where \( \{e_i\}_{i=1}^{p} \) are \( p \)-dimensional unit vectors and the second to last equality is a consequence of Lemma A.1 with \( \tilde{C} \) as in (A.7).

References


Marie-Christine Düker
Faculty of Mathematics
Ruhr-Universität Bochum
Universitätsstr. 150, IB
44780 Bochum, Germany
Marie-Christine.Dueker@rub.de

Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Hanes Hall
Chapel Hill, NC 27599, USA
pipiras@email.unc.edu

Raanju R. Sundararajan
Statistics Program
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
raanju.sundararajan@kaust.edu.sa