# Cotrending: testing for common deterministic trends in varying means model

Marie-Christine Düker[a], Vladas Pipiras[b,*], Raanju Sundararajan[c]

[a]Faculty of Mathematics, Ruhr-Universität Bochum, Universitätsstr. 150, IB, 44780 Bochum, Germany
[b]Dept. of Statistics and Operations Research, CB# 3260, Hanes Hall, University of North Carolina, Chapel Hill, NC 27599-3260, USA
[c]Dept. of Statistical Science, Southern Methodist University, Dallas, TX 75205, USA

**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.

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

*2020 MSC:* Primary 62M10, Secondary 62G10.

## 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

$$X_t = \mu\Big(\frac{t}{T}\Big) + Y_t, \quad t \in \{1, \dots, T\}. \tag{1}$$

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 $(\mu_1(u), \dots, \mu_p(u))^\top$ and $Y_t$ are $p$-vector i.i.d. error terms with $\mathrm{E}\, Y_t = 0$. We shall further assume that the covariance matrix of the error terms $\mathrm{E}\, Y_t Y_t^\top$ may vary with time, and also treat the simpler special case when it does not, that is, $\mathrm{E}\, Y_t Y_t^\top = \Sigma$, separately – the reader may have this case in mind for the rest of this section. The case when $Y_t$ are dependent in time will also be discussed but the focus below will be on the case when $Y_t$ are independent across time. We think of (1) as modeling varying means across $p$ dimensions and shall refer to (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 $\mathrm{E}\, B_1^\top X_t = B_1^\top \mu(t/T) \equiv \mu_1$ for a constant $d_1 \times 1$ vector $\mu_1$ or, at the model level,

$$B_1^\top \mu(u) = \mu_1, \quad u \in (0, 1]. \tag{2}$$

*Corresponding author.
*Email address:* pipiras@email.unc.edu (Vladas Pipiras)

**Definition 1.** The largest $d_1$ for which (2) holds is called the cotrending dimension of the corresponding cotrending subspace $\mathcal{B}_1$, spanned by the columns of $B_1$. Similarly, $d_2 = p - d_1$ is called the noncotrending dimension of the corresponding noncotrending subspace $\mathcal{B}_2$, with $\mathcal{B}_2 \perp \mathcal{B}_1$.

The dimension $d_2$ indicates how many non-constant deterministic functions drive the system (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 (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 (2) is equivalent to

$$B_1^\top M B_1 = 0, \tag{3}$$

where

$$M = \int_0^1 (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})^\top du \quad \text{with} \quad \bar{\mu} = \int_0^1 \mu(u) du. \tag{4}$$

The matrix $M$ is positive semidefinite. Then, according to (3),

$$d_1 = \mathrm{nl}\{M\}, \quad d_2 = \mathrm{rk}\{M\}, \tag{5}$$

where $\mathrm{nl}\{M\}$ and $\mathrm{rk}\{M\}$ 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 $\mathcal{B}_1$ 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 $\widehat{M}$ of $M$, especially in the econometrics literature (Cragg and Donald [8], Gill and Lewbel [14], Kleibergen and Paap [21], Robin and Smith [29]), and slightly less so in the statistics literature (Anderson [1], Eaton and Tyler [12], Camba-Mendez and Kapetanios [6]). Furthermore, there are technical reasons for $\widehat{M}$ to be nondefinite, when $M$ is positive semidefinite itself, as it is the case here (Donald et al. [11] and Section 3 below). Nondefiniteness refers to matrices which are neither positive semi-definite nor negative semi-definite. Much of the technical contribution of this work consists of introducing such an estimator for $M$ in (4), proving its asymptotic normality result and then applying the available matrix rank tests. The considered estimator turns out to have a surprisingly simple form. We shall also discuss what can be said about the convergence of the sample eigenvectors corresponding to the cotrending subspace $\mathcal{B}_1$. 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. [5], and studied further by Blythe et al. [3], Sundararajan and Pourahmadi [33]. See also [7] in the context of locally stationary processes (Dahlhaus [9], Dahlhaus and Polonik [10], Ombao et al. [25]). In the work somewhat parallel to this (Sundararajan et al. [32]), 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 (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 [22]. 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) 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 (4) and state its asymptotic normality results, whose proofs can be found in Appendix A. Complementary results can be found in Appendix B which got moved to the Supplementary material. The application of some of the available matrix rank tests is discussed in Section 3. Section 4 concerns inference of the cotrending subspace $\mathcal{B}_1$. In Section 5, we discuss possible extensions of the underlying model and give an alternative approach to estimate the cotrending dimension. Section 6 details connections to PCA and cointegration. A simulation study and applications are considered in Sections 7 and 8. Section 9 concludes.

## 2. Matrix estimator and its asymptotic normality

We introduce here an asymptotically normal estimator $\widehat{M}_S$ for $M$ in (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) to vary with time. More specifically, we set $Y_t = \sigma(t/T)Z_t$ with i.i.d. vectors $Z_t$ satisfying $\mathrm{E}\, Z_t = 0$, $\mathrm{E}\, Z_t Z_t^\top = I_p$, so that the VM model (1) becomes for $t \in \{1, \dots, T\}$

$$X_t = \mu\Big(\frac{t}{T}\Big) + \sigma\Big(\frac{t}{T}\Big)Z_t, \tag{6}$$

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

$$\sqrt{T}\, \mathrm{vech}(\widehat{M}_S - M) \xrightarrow{\mathrm{d}} \mathcal{N}(0, C). \tag{7}$$

Furthermore, we need a consistent estimator $\widehat{C} = \widehat{C}(T)$ for the resulting covariance matrix $C$, that is, $\widehat{C} \xrightarrow{\mathrm{p}} C$. Throughout this work, $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{p}}$ stand for the convergence in distribution and probability, respectively. We also use the notation $\mathrm{vec}(M)$ for a matrix $M$. The vec operator transforms a matrix $M$ into a vector by stacking the columns of $M$ in a vector. The vech operator is obtained by eliminating all elements above the diagonal of $M$ from $\mathrm{vec}(M)$. The two operators relate to each other through the duplication matrix $D_p$ as $\mathrm{vec}(M) = D_p\, \mathrm{vech}(M)$. For a matrix (or vector) $A$, we also set $A^2 = AA^\top$.

By Proposition 2.1 in Donald et al. [11], a positive semidefinite estimator $\widehat{M}$ of $M$ satisfying (7) 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

$$\widehat{M}_S = \frac{1}{2}(\widehat{M} + \widehat{M}^\top), \tag{8}$$

with

$$\widehat{M} = \frac{1}{T}\sum_{t=1}^{T-1}(X_t - \bar{X}_T)(X_{t+1} - \bar{X}_T)^\top, \quad \bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t. \tag{9}$$

(See also Remark 1 for some insight into (8).) Note that though we use the sample autocovariance matrix, the considered model (6) is generally nonstationary. The following proposition formulates the asymptotic normality result for $\widehat{M}_S$. The matrix $D_p^+ = (D_p^\top D_p)^{-1} D_p^\top$ appearing below is the Moore-Penrose inverse of the duplication matrix $D_p$, and is known as the elimination matrix; see Magnus and Neudecker [24]. We call a function on $[0, 1]$ piecewise continuous if $[0, 1]$ can be partitioned into a finite union of subintervals where the function is continuous, has left and right limits at the subinterval endpoints, and is also right-continuous.

**Proposition 1.** *Let $X_t$ follow the model* (6) *with piecewise continuous $\mu, \sigma^2$ and i.i.d. $Z_t$ satisfying* $\mathrm{E}\, Z_t = 0$, $\mathrm{E}\, Z_t Z_t^\top = I_p$ *and* $\mathrm{E}\, \|Z_t\|^{2+\delta} < \infty$ *for some* $\delta > 0$. *Then, the estimator $\widehat{M}_S$ of $M$ satisfies* (7) *with the limiting covariance matrix*

$$C = D_p^+\Big(\int_0^1 \sigma^2(u) \otimes \sigma^2(u)du + 4\int_0^1 (\mu(u) - \bar{\mu})(\mu(u) - \bar{\mu})^\top \otimes \sigma^2(u)du\Big)(D_p^+)^\top. \tag{10}$$

The proof of Proposition 1 can be found in Appendix A. The following corollary formulates the result when $\sigma^2(u) \equiv \Sigma$ does not depend on $u$.

**Corollary 1.** *When $\sigma^2(u) \equiv \Sigma$, Proposition 1 holds with the limiting covariance matrix*

$$C = D_p^+((\Sigma \otimes \Sigma) + 4(M \otimes \Sigma))(D_p^+)^\top. \tag{11}$$

**Remark 1.** The intuition behind the estimator $\widehat{M}_S$ in (8) is quite simple. Replacing $\bar{X}_T$ by $\bar{\mu}$ for simplicity, note that

$$\mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T-1}(X_t-\bar{\mu})(X_{t+1}-\bar{\mu})^\top\right) = \frac{1}{T}\sum_{t=1}^{T-1}\left(\mu\Big(\frac{t}{T}\Big)-\bar{\mu}\right)\left(\mu\Big(\frac{t+1}{T}\Big)-\bar{\mu}\right)^\top,$$

where we used the fact that $\mathrm{E}\, Z_t Z_{t+1}^\top = 0$. The latter expression approximates $M$ for piecewise continuous $\mu$. The estimation error made by replacing $\overline{\mu}$ with $\overline{X}_T$ is asymptoticly negligible as further investigated in the proof of Proposition 2.1.

The following proposition gives a consistent estimator for the limiting covariance matrix $C$ in (10) and (11). It is proved in Appendix A.

**Proposition 2.** *Under the assumptions of Proposition 1, the estimator*

$$\widehat{C} = D_p^+\frac{1}{T}\sum_{t=1}^{T-3}\Big(\frac{1}{4}(\Delta X_{t+1})^2 \otimes (\Delta X_{t+3})^2 + 2((\Delta X_{t+3})^2 \otimes (X_t-\bar{X}_T)(X_{t+1}-\bar{X}_T)^\top)\Big)(D_p^+)^\top \tag{12}$$

*with $\Delta X_t = X_t - X_{t-1}$, is an asymptotically unbiased and consistent estimator for $C$ in* (10)*.*

The following corollary gives a simpler estimator for the limiting covariance matrix in Corollary 1.

**Corollary 2.** *Under the assumptions of Proposition 1 and when $\sigma^2(u) \equiv \Sigma$, the estimator*

$$\widehat{C} = D_p^+\Big(\frac{1}{4}(\widehat{\Sigma}\otimes\widehat{\Sigma}) + 2(\widehat{M}_S \otimes \widehat{\Sigma})\Big)(D_p^+)^\top, \ \ \widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T-1}(\Delta X_{t+1})^2$$

*with $\widehat{M}_S$ as in* (9)*, is an asymptotically unbiased and consistent estimator for $C$ in* (11)*.*

As noted above, we want to have nonsingular limiting covariance matrix $C$ of the estimator $\widehat{M}_S$ to ensure the applicability of available matrix rank tests. This motivated the choice of a nondefinite estimator $\widehat{M}_S$, yielding the limiting covariance matrix $C$ in (10). Since $(\mu(u)-\bar{\mu})(\mu(u)-\bar{\mu})^\top$ in (10) might be non-positive definite, 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 $\mathrm{rk}\{M\}$ (or $\mathrm{nl}\{M\}$) having a matrix estimator $\widehat{M}_S$ satisfying (7). We start by justifying the statements in (5).

**Lemma 1.** *Suppose that $\mu$ is piecewise continuous and let $d_1$, $d_2$ be the cotrending and noncotrending dimensions defined for $\mu$ in Definition 1. Then, the relations* (5) *hold.*

**Proof:** The cotrending and noncotrending dimensions $d_1, d_2$ are defined in terms of the relation (2). Since $\mu_1 = B_1^\top\bar{\mu}$, (2) can be written as

$$B_1^\top(\mu(u)-\bar{\mu}) = 0,$$

which implies (3). The converse is a consequence of writing (3) as

$$\int_0^1 (x^\top B_1^\top(\mu(u)-\bar{\mu}))^2 du = 0$$

for any $x \in \mathbb{R}^{d_1}$. This implies that $x^\top B_1^\top(\mu(u)-\bar{\mu}) = 0$ a.e. $du$, for all $x$. By Fubini's theorem, $x^\top B_1^\top(\mu(u)-\bar{\mu}) = 0$ a.e. $dxdu$ and because of continuity in $x$, $x^\top B_1^\top(\mu(u)-\bar{\mu}) = 0$ for all $x$, a.e. $du$. The latter implies that $B_1^\top(\mu(u)-\bar{\mu}) = 0$ a.e. $du$. Since $\mu$ is piecewise continuous, we also have $B_1^\top(\mu(u)-\bar{\mu}) = 0$ for all $u$, which implies (2). □

To test for the rank of the matrix $M$, we use the so-called SVD matrix rank test proposed by Kleibergen and Paap [21] and, more precisely, its analogue for symmetric matrices found in Donald et al. [11]. The test has some advantages over other matrix rank tests considered in Donald et al. [11], for example, the limiting covariance matrix $C$ in (7) is not required to have a Kronecker product structure. Furthermore, in a simulation study of Donald et al. [11], the SVD matrix rank test outperforms other matrix rank tests in terms of its power. Consider the following hypothesis testing problem,

$$H_0 : \mathrm{rk}\{M\} = r \quad \text{vs.} \quad H_1 : \mathrm{rk}\{M\} > r, \tag{13}$$

where $r \in \{0, \dots, p-1\}$ is fixed. The SVD matrix rank test is based on the singular value decomposition of $M$ as

$$M = USU^\top = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} U_{11}^\top & U_{21}^\top \\ U_{12}^\top & U_{22}^\top \end{pmatrix},$$

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. [11], $M$ can be written as

$$M = A_r B_r + A_{r,\perp} \Lambda_r B_{r,\perp}$$

with

$$A_r = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 U_{11}^\top, \quad B_r = \begin{pmatrix} I_r & (U_{11}^\top)^{-1} U_{21} \end{pmatrix}, \quad A_{r,\perp} = B_{r,\perp}^\top = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1} (U_{22} U_{22}^\top)^{\frac{1}{2}}$$

and

$$\Lambda_r = (U_{22} U_{22}^\top)^{-\frac{1}{2}} U_{22} S_2 U_{22}^\top (U_{22} U_{22}^\top)^{-\frac{1}{2}}.$$

Note, that for a positive definite real matrix $A$, $A^{\frac{1}{2}} = BD^{\frac{1}{2}} B^\top$, where $D$ contains the eigenvalues, and $B$ the orthonormal eigenvectors of $A$. Similarly, $A^{-1} = BD^{-1} B^\top$. Then, the null hypothesis $H_0 : \mathrm{rk}\{M\} = r$ is equivalent to $H_0 : \Lambda_r = 0$. In the SVD matrix rank test, the latter hypothesis is tested having the symmetric matrix estimator $\widehat{M}_S$ in (8). Let $\widehat{\Lambda}_r$ be the quantity analogous to $\Lambda_r$ but defined through $\widehat{M}_S$. By Proposition 1, the estimator $\widehat{M}_S$ satisfies (7) with the limiting covariance matrix (10). Then, by Donald et al. [11], Proposition 4.1, $\sqrt{T}\,\mathrm{vech}(\widehat{\Lambda}_r) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Omega_r)$ under the hypothesis $H_0 : \Lambda_r = 0$, where $\Omega_r = D_{p-r}^+ (B_{r,\perp} \otimes A_{r,\perp}^\top) D_p C D_p^\top (B_{r,\perp}^\top \otimes A_{r,\perp}) (D_{p-r}^+)^\top$ and the matrix $C$ is defined in (10). The suggested SVD test statistic is then

$$\widehat{\xi}_{svd}(r) = T\,\mathrm{vech}(\widehat{\Lambda}_r)^\top \widehat{\Omega}_r^{-1}\,\mathrm{vech}(\widehat{\Lambda}_r).$$

Here, $\widehat{\Omega}_r$ is defined by replacing the component matrices of $\Omega_r$ by their sample counterparts, including $\widehat{C}$ defined by (12). By Theorem 4.1 in Donald et al. [11], under $H_0 : \mathrm{rk}\{M\} = r$ and if the matrix $\Omega_r$ is non-singular,

$$\widehat{\xi}_{svd}(r) \xrightarrow{\mathrm{d}} \chi^2((p-r)(p-r+1)/2), \tag{14}$$

where $\chi^2(K)$ denotes the chi-square distribution with $K$ degrees of freedom. Furthermore, Theorem 4.1 in Donald et al. [11] gives $\widehat{\xi}_{svd}(r) \xrightarrow{\mathrm{p}} \infty$ under $H_1 : \mathrm{rk}\{M\} > r$.

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 $H_0 : \mathrm{rk}\{M\} = r$ is not rejected. By using the aforementioned asymptotic results, the resulting estimator can be shown to be consistent for $\mathrm{rk}\{M\}$ in a standard way when the significance level suitably depends on the sample size (e.g., Robin and Smith [29], Theorem 5.2).

## 4. Inference of (non)cotrending subspace

We are interested here in inference about the cotrending subspace $\mathcal{B}_1$ spanned by the columns of $B_1$ satisfying (2) and hence also about the noncotrending subspace $\mathcal{B}_2$ characterized by $\mathcal{B}_2 \perp \mathcal{B}_1$. The cotrending subspace is spanned by the eigenvectors of $M$ in (4) associated with the zero eigenvalues. We shall establish a consistency result for the estimated eigenvectors when using $\widehat{M}_S$ in (8) for $M$, and also discuss an available method to test whether a certain set of vectors lies in $\mathcal{B}_1$.

Let $\lambda_1 \geqslant \cdots \geqslant \lambda_{p-d_1} > \lambda_{p-d_1+1} = \cdots = \lambda_p = 0$ and $\widehat{\lambda}_1 \geqslant \cdots \geqslant \widehat{\lambda}_p$ denote the eigenvalues of the symmetric matrices $M$ and $\widehat{M}_S$, respectively. Let also $v_j$ and $\widehat{v}_j$ be the corresponding orthonormal eigenvectors satisfying $Mv_j = \lambda_j v_j$ and $\widehat{M}_S \widehat{v}_j = \widehat{\lambda}_j \widehat{v}_j$. Furthermore, define the $p \times d$ matrices

$$V = (v_i, v_{i+1}, \ldots, v_{i+d-1}), \quad \widehat{V} = (\widehat{v}_i, \widehat{v}_{i+1}, \ldots, \widehat{v}_{i+d-1}). \tag{15}$$

Observe that $V = B_1$ when $i = p - d_1 + 1$ and $d = d_1$. The next result proves consistency of the eigenvectors in $\widehat{V}$. A short proof can be found in Appendix A and is based on the Davis-Kahan theorem. For convenience, the Davis-Kahan theorem is stated in the appendix above the proof of the following proposition.

**Proposition 3.** *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* $\widehat{O} \in \mathbb{R}^{d \times d}$ *such that*

$$\|\widehat{V}\widehat{O} - V\|_F = O_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big),$$

*where* $\widehat{V}$ *and* $V$ *are as in* (15) *and* $\|\cdot\|_F$ *denotes the Frobenius norm.*

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 $\widehat{V} = \widehat{B}_1$.

Suppose that $\widehat{r}$ is the estimated rank of the matrix $M$ following the proposed testing procedure in Section 3. Then, the estimated cotrending dimension $\widehat{d}_1 = p - \widehat{r}$ refers to the number of eigenvectors which generate the estimated cotrending subspace $\widehat{\mathcal{B}}_1$. We expect that $\widehat{d}_1 = d_1$ in the asymptotic sense and that in some situations, the respective estimated eigenvectors in $\widehat{\mathcal{B}}_1$ 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 $\mathcal{B}_1$, we use a test proposed by Tyler [34].

Let $\Lambda = (\lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+d-1})$ be the eigenvalues associated with the eigenvectors from $V$ in (15). The total eigenprojection of $M$ associated with $\Lambda$ is defined as

$$P_0 = \sum_{\lambda \in \Lambda} P_\lambda, \quad P_{\lambda_j} = v_j v_j^\top.$$

The total eigenprojection $\widehat{P}_0$ of $\widehat{M}_S$ is defined analogously by replacing the eigenvalues and eigenvectors in $\Lambda$ and $V$ by their sample counterparts. Consider the hypothesis testing problem

$$H_0 : P_0 Q = Q \quad \text{vs.} \quad H_1 : P_0 Q \neq Q, \tag{16}$$

where $Q$ denotes a $p \times q$ matrix with $\mathrm{rk}\{Q\} = q \leqslant d$. 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 $\mathcal{B}_1$.

By Proposition 1, the estimator $\widehat{M}_S$ satisfies (7) with a limiting covariance matrix $C$ in (10). Using the relation $\mathrm{vec}(M) = D_p \,\mathrm{vech}(M)$, one may also infer that $\sqrt{T}\,\mathrm{vec}(\widehat{M}_S - M) \xrightarrow{\mathrm{d}} \mathcal{N}(0, D_p C D_p^\top)$ holds, which coincides with the required assumptions in Tyler [34]. Furthermore, by Proposition 2, there is a consistent estimator $\widehat{C}$ for $C$ and $C$ is nonsingular. Then, by Theorem 4.1 in Tyler [34],

$$\mathrm{vec}(\sqrt{T}(I_p - \widehat{P}_0)Q) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_Q)$$

under the hypothesis $H_0 : P_0 Q = Q$, where

$$\Sigma_Q = (Q^\top \otimes I_p) R_\Lambda^\top D_p C D_p^\top R_\Lambda (Q \otimes I_p)$$

with $C$ as in (10) 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

$$\widehat{\gamma}_{evd}(Q) = T(\mathrm{vec}(Q))^\top \widehat{\Sigma}_Q^+ \mathrm{vec}(Q),$$

where $\widehat{\Sigma}_Q^+$ denotes the Moore-Penrose inverse of $\widehat{\Sigma}_Q$. The matrix $\widehat{\Sigma}_Q$ is defined by replacing the component matrices of $\Sigma_Q$ by their sample counterparts, i.e., $R_\Lambda$ is written in terms of $\widehat{V}$ and $C$ is replaced by its consistent estimator $\widehat{C}$ given in (12). Then, by Theorem 5.3 in Tyler [34], under $H_0 : P_0 Q = Q$,

$$\widehat{\gamma}_{evd}(Q) \xrightarrow{\mathrm{d}} \chi^2(q(p-d)).$$

As shown in Section 6 in Tyler [34], based on the preceding test, one can define an asymptotic $100(1-\alpha)\%$ confidence region as

$$\{\mathcal{Q} \mid Q \in \mathbb{R}^{p\times q}, \mathrm{rk}\{Q\} = q, \widehat{\gamma}_{evd}(Q) < \chi^2_{1-\alpha}(q(p-d))\}, \tag{17}$$

where $\chi^2_{1-\alpha}(K)$ denotes the $(1-\alpha)$-quantile of a chi-square distribution with $K$ degrees of freedom, and $\mathcal{Q} = \{v \in \mathbb{R}^p | v = Qw \text{ for some } w \in \mathbb{R}^q\}$ 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 (17) can be expressed in terms of the estimated eigenvectors $\widehat{V}$ in (15) as

$$\{\mathcal{Q} \mid \widehat{V}^\top Q = I_d,\ T(\mathrm{vec}(Q-\widehat{V}))^\top \widehat{\Sigma}_{\widehat{V}}^+ \mathrm{vec}(Q-\widehat{V}) < \chi^2_{1-\alpha}(r(p-d))\}. \tag{18}$$

See Section 6 in [34].

## 5. Model generalizations and alternative analysis approaches

In this section, we discuss a number of points related to model assumptions and analysis method. Section 5.1 concerns the case when there is temporal dependence in the error terms $Y_t$ in (1). In Section 5.2, we give an alternative formulation of the cotrending problem, called smooth cotrending.

### *5.1. Model generalizations to temporal dependence*

The model (1) was considered above supposing that the error terms $Y_t$ are i.i.d. over time $t$. However, generalizations to temporal dependent error terms are certainly conceivable. A simple generalization can be achieved by supposing $m$-dependence. Recall that a stationary series $\{Y_t\}_{t\geqslant 1}$ is $m$-dependent if the series $\{Y_t\}_{t\leqslant 1}$ and $\{Y_t\}_{t\geqslant m+2}$ are independent. In this case, the cotrending analysis could be based on the sample autocovariance matrix at lag $m+2$ as an estimator $\widehat{M}$ for $M$, instead of the sample autocovariance matrix at lag 1 as in (8)–(9). The consistency of such estimator $\widehat{M}$ is a simple consequence of the uncorrelatedness of $Y_t$ at different times; see Remark 1.

To have, in particular, a consistent estimator for $M$ when $Y_t$ admits a more general temporal dependence structure, we propose to average the sample autocovariances over different time lags, namely, to consider

$$\widehat{M}_L = \frac{1}{2L}\sum_{l=1}^{L}(\widehat{\Gamma}_X(l) + (\widehat{\Gamma}_X(l))^\top),\ \ \widehat{\Gamma}_X(l) = \frac{1}{T}\sum_{t=1}^{T-l}\Big(X_t - \bar{X}_T\Big)\Big(X_{t+l} - \bar{X}_T\Big)^\top. \tag{19}$$

Note that the estimator $\widehat{M}_L$ is once again nondefinite and symmetric. We use here the notation $\widehat{\Gamma}_X(l)$, which is commonly used for the autocovariance function of a stationary time series at lag $k$, even though in our setting, the time series $\{X_t\}_{t=1,\dots,T}$ is not stationary. The intuition behind the estimator $\widehat{M}_L$ is that for an increasing number of time lags $l$, the temporal dependence averages out. That is,

$$\begin{aligned}\mathrm{E}\,\widehat{M}_L = \mathrm{E}\left(\frac{1}{2L}\sum_{l=1}^{L}(\widehat{\Gamma}_X(l) + \widehat{\Gamma}_X(l)^\top)\right) &\approx M + \mathrm{E}\left(\frac{1}{2L}\sum_{l=1}^{L}(\widehat{\Gamma}_Y(l) + \widehat{\Gamma}_Y(l)^\top)\right)\\ &\approx M + \frac{1}{2L}\sum_{l=1}^{L}\frac{T-l}{T}(\Gamma_Y(l) + \Gamma_Y(l)^\top) \approx M, \quad \text{for} \quad T, L \to \infty.\end{aligned} \tag{20}$$

Here, $\widehat{\Gamma}_Y(l)$ denotes similarly the sample autocovariance function of $Y_t$ at lag $l$ and $\Gamma_Y(l)$ the corresponding population quantity. According to (20), imposing a stronger temporal dependence on the error terms requires that the autocovariances $\Gamma_Y(l)$ of the error terms are absolutely summable

$$\sum_{l\in\mathbb{Z}} \|\Gamma_Y(l)\|_F < \infty. \tag{21}$$

Under mild technical assumptions (including (21) and assuming that $Y_t$ is a linear process), we can show that the limiting covariance matrix can be calculated as

$$\begin{aligned} &T\,\mathrm{E}(\,\mathrm{vech}(\widehat{M}_L - M)(\mathrm{vech}(\widehat{M}_L - M))^\top) \\ &\quad = D_p^+\Big(\Big(\sum_{l\in\mathbb{Z}}\Gamma_Y(l)\otimes\sum_{l\in\mathbb{Z}}\Gamma_Y(l)\Big) + 4(\Gamma_Y(0)\otimes M)\Big)(D_p^+)^\top + o\Big(\frac{1}{T}\Big). \end{aligned} \tag{22}$$

Note the analogy of (22) with (11): In the first summand, the covariance matrix of $Y_t$ got replaced by the series of autocovariace matrices. We omit the proof of (22) and any further discussion as the focus of this work is on the case of independent $Y_t$.

*5.2. Alternative formulation: Smooth cotrending*

The cotrending analysis presented in Section 2–4 relied on the relations (3)–(5), as a pathway to the cotrending dimension $d_1$ and subspace $\mathcal{B}_1$. But there are also other interesting ways to these quantities. For example, assuming smoothness of the function $\mu(u)$, note that the relation (2) is equivalent to

$$B_1^\top \mathcal{M} B_1, \tag{23}$$

where

$$\mathcal{M} = \int_0^1 \mu^{(1)}(u)\mu^{(1)}(u)^\top du \quad \text{with} \quad \mu^{(1)}(u) = \frac{d}{du}\mu(u). \tag{24}$$

The equivalence between (2) and (23) can be proved similarly as Lemma 1, since

$$\frac{d}{du}B_1^\top\mu(u) = B_1^\top\mu^{(1)}(u) = \frac{d}{du}\mu_1 = 0. \tag{25}$$

A possible estimator for $\mathcal{M}$ is

$$\widehat{\mathcal{M}} = \frac{1}{Th^2}\sum_{t=1}^{T-\lfloor Th\rfloor}(X_{t+\lfloor Th\rfloor+1} - X_{t+1})(X_{t+\lfloor Th\rfloor} - X_t)^\top \tag{26}$$

with $h \to 0$ and $Th^2 \to \infty$. For the intuition behind this estimator, observe that

$$\begin{aligned} \mathrm{E}\,\widehat{\mathcal{M}} &= \frac{1}{Th^2}\sum_{t=1}^{T-\lfloor Th\rfloor-1}\mathrm{E}(X_{t+\lfloor Th\rfloor+1} - X_{t+1})(X_{t+\lfloor Th\rfloor} - X_t)^\top \\ &= \frac{1}{Th^2}\sum_{t=1}^{T-\lfloor Th\rfloor-1}\Big(\mu\Big(\frac{t+\lfloor Th\rfloor+1}{T}\Big) - \mu\Big(\frac{t+1}{T}\Big)\Big)\Big(\mu\Big(\frac{t+\lfloor Th\rfloor}{T}\Big) - \mu\Big(\frac{t}{T}\Big)\Big)^\top \approx \int_0^1 \mu^{(1)}(u)\mu^{(1)}(u)^\top du. \end{aligned}$$

One advantage of this smooth cotrending approach is that the resulting limiting covariance matrix can be shown to have a Kronecker product structure. Many of the available matrix rank tests either assume a Kronecker product structure or their test statistics simplify for this structure. However, one downside is the relatively slow convergence rate of $\widehat{M}$ to $M$, which can be shown to be of the order $\sqrt{Th^4}$.

## 6. Connections to other approaches

We discuss here connections of the VM model and the introduced estimation framework to principal component analysis (PCA) (Section 6.1) and cointegration (Section 6.2).

*6.1. Connections to PCA*

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

$$\widehat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X}_T)(X_t - \bar{X}_T)^\top,$$

which is the autocovariance function at lag 0, rather than this function at lag 1 as in (8) and (9). This has the following implications. For the VM model, we get by replacing $\bar{X}_T$ by $\bar{\mu}$ for simplicity,

$$\begin{aligned} \mathrm{E}\,\widehat{\Gamma} &= \mathrm{E}\Big(\frac{1}{T}\sum_{t=1}^{T}(X_t-\bar{\mu})(X_t-\bar{\mu})^\top\Big) = \frac{1}{T}\sum_{t=1}^{T}\Big(\mu\Big(\frac{t}{T}\Big)-\bar{\mu}\Big)\Big(\mu\Big(\frac{t}{T}\Big)-\bar{\mu}\Big)^\top + \frac{1}{T}\sum_{t=1}^{T}\sigma\Big(\frac{t}{T}\Big)\mathrm{E}(Z_tZ_t^\top)\sigma\Big(\frac{t}{T}\Big)^\top \\ &= M + \int_0^1 \sigma^2(u)du + O\Big(\frac{1}{T}\Big), \end{aligned}$$

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

Vice versa, we also note that our estimator $\widehat{M}_S$ would not be of much interest in the many PCA scenarios that work with independent copies of the vectors $X_t$ (e.g., Jolliffe [20]). Indeed, for such vectors, the estimator $\widehat{M}_S$ based on the autocovariances at lag 1 would be zero asymptotically.

*6.2. Connections to cointegration*

In this section, we establish interesting connections of our approach to cointegration (Granger [15], Engle and Granger [13], Johansen [19]). 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 $\widehat{M}_S$ 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 $\ell$th order vector autoregression (VAR($\ell$)) model

$$X_t = \sum_{i=1}^{\ell}\Pi_i X_{t-i} + \varepsilon_t \tag{27}$$

with

$$\mathrm{E}\,\varepsilon_0 = 0,\ \ \mathrm{E}\,\varepsilon_t\varepsilon_t^\top = \Sigma_\varepsilon,\ \ \mathrm{E}\,\varepsilon_t\varepsilon_s^\top = 0,\ \ t \neq s. \tag{28}$$

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

$$\Pi = \Pi_1 + \cdots + \Pi_\ell - I_p,\ \Gamma_i = -(\Pi_{i+1} + \cdots + \Pi_\ell). \tag{29}$$

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 [13] defined the cointegrating rank in terms of the matrix $\Pi$ in (29) as

$$r^* = \mathrm{rk}\{\Pi\}. \tag{30}$$

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 $\beta_i$, $i \in \{1, \dots, r^*\}$, such that $\beta_i^\top X_t$ is stationary. Furthermore, the matrix $\Pi$ in (29) can be written as

$$\Pi = \alpha\beta^\top \tag{31}$$

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 2.** 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 [16], Maddala and Kim [23]).

To analyze our estimator in the context of cointegration, we need some technical assumptions. The so-called Granger representation, introduced in Johansen [19], Theorem 4.1, enables one to separate the cointegrated process into stationary and nonstationary components. Define

$$C(z) = \Pi + \sum_{i=0}^{\ell} \Gamma_i (1-z) z^i,$$

where $\Gamma_0 = -I_p$. Let also $\alpha_\perp, \beta_\perp$ be the orthogonal complements of $\alpha, \beta$ in (31). Then, if $\det(C(z)) = 0$ has roots on or outside the unit circle and if the matrix

$$\alpha_\perp^\top \Big( I_p - \sum_{i=1}^{\ell-1} \Gamma_i \Big) \beta_\perp$$

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

$$X_t = L \sum_{i=1}^{t} \varepsilon_i + \sum_{j=0}^{\infty} \widetilde{L}_j \varepsilon_{t-j} + \widetilde{X}_0, \tag{32}$$

with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ as in (27) and $\widetilde{X}_0$ as an initial value. Furthermore,

$$L = \beta_\perp \Big( \alpha_\perp^\top \Big( I_p - \sum_{i=1}^{\ell-1} \Gamma_i \Big) \beta_\perp \Big)^{-1} \alpha_\perp^\top \tag{33}$$

and the series $\sum_{j=0}^{\infty} \|\widetilde{L}_j\|_F$ is finite, where $\|\cdot\|_F$ denotes the Frobenius norm. The first term in (32) 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 $L = (L_n^\top, 0_{p-n}^\top)^\top$, 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 $I_{p\times p} = (I_{1,n}^\top, I_{2,p-n}^\top)^\top$. The following result investigates the asymptotic behavior of the estimator $\widehat{M}_S$ defined in (8) for a cointegrated system in (27). It is proved in Appendix A.

**Proposition 4.** *Suppose that $X_t$ follows a* VAR($\ell$) *model* (27) *with cointegrating rank $0 \leqslant r^* \leqslant p$. Assume also that $\Sigma_\varepsilon$ in* (28) *is positive definite and* $\mathrm{E}\,\|\varepsilon_0\|^4 < \infty$. *Then, the symmetric estimator $\widehat{M}_S$ in* (8) *satisfies*

$$\Delta_{2,T}^{-\frac{1}{2}} \widehat{M}_S \Delta_{2,T}^{-\frac{1}{2}} \xrightarrow{\mathrm{d}} \begin{pmatrix} L_n \Sigma_\varepsilon^{\frac{1}{2}} Z \Sigma_\varepsilon^{\frac{1}{2}} L_n^\top & 0 \\ 0 & \frac{1}{2}(\Upsilon_{p-n}(1) + \Upsilon_{p-n}^\top(1)) \end{pmatrix}, \tag{34}$$

*where* $\Delta_{2,T} = \mathrm{diag}(TI_n, I_{p-n})$*,* $\Upsilon_{p-n}(1) = I_{2,p-n}\sum_{j=0}^{\infty}\widetilde{L}_j\widetilde{L}_{j+1}^{\top}I_{2,p-n}^{\top}$*, and*

$$Z = \int_0^1 (Z(u)-\bar{Z})(Z(u)-\bar{Z})^{\top}du \quad \textit{with} \quad \bar{Z} = \int_0^1 Z(u)du, \tag{35}$$

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

By construction, $\mathrm{rk}\{L_n\} = \mathrm{rk}\{L\} = p - r^*$. For this reason and since $Z$ in (35) and $\Sigma_\varepsilon^{\frac{1}{2}}$ are positive definite,

$$\mathrm{rk}\{L_n\Sigma_\varepsilon^{\frac{1}{2}}Z\Sigma_\varepsilon^{\frac{1}{2}}L_n^{\top}\} = \mathrm{rk}\{L_n\} = p - r^*.$$

Then,

$$\mathrm{rk}\left\{\begin{pmatrix} L_n\Sigma_\varepsilon^{\frac{1}{2}}Z\Sigma_\varepsilon^{\frac{1}{2}}L_n^{\top} & 0 \\ 0 & \frac{1}{2}(\Upsilon_{p-n}(1)+\Upsilon_{p-n}^{\top}(1)) \end{pmatrix}\right\} \geqslant p - r^*.$$

In general, this suggests that testing for nullity with $\widehat{M}_S$ 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 (34) for $\widehat{M}_S$ reduces to

$$\frac{1}{T}\widehat{M}_S \xrightarrow{\mathrm{d}} L\Sigma_\varepsilon^{\frac{1}{2}}Z\Sigma_\varepsilon^{\frac{1}{2}}L^{\top}.$$

Then, the corresponding rank is given by

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

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($\ell$) model with a random matrix $R \in \mathbb{R}^{p\times p}$ with full rank, since $\mathrm{rk}\{RL\} = \mathrm{rk}\{L\}$. 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 7 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.

The purpose of this section was not to give an estimator for the cointegration rank. We see cotrending as a potential alternative concept to cointegration and rather want to emphasize the similarities between the matrix $M$ in (4) and the limiting process (34) which is characterized by the matrix defined in (35). The matrix in (35) coincides with the matrix $M$ by replacing the vector $\mu$ with a standard Brownian motion $Z$. While $\mu$ in our model determines a deterministic trend, the standard Brownian motion $Z$ determines a stochastic trend since the integral over $Z$ is the limit of the sample mean of a random walk.

## 7. 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 (13) as

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

where $d \in \{1, \ldots, p\}$. 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 (30), we apply the widely used Johansen test ([19]). The corresponding hypothesis testing problem can be written as

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

where $r \in \{0, \ldots, p-1\}$. 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 $d \in \{1, \ldots, p\}$ and $r \in \{0, \ldots, p-1\}$. The probability $\alpha \in (0, 1)$ on the vertical axis is plotted versus estimated $p_l(\alpha) = \mathrm{P}(\widehat{\xi}(l) > q_l(\alpha))$, $l = d, r$ on the horizontal axis. The values $q_l(\alpha)$ are such that $\mathrm{P}(\widehat{\xi}(l) > q_l(\alpha)) = \alpha$. The respective test statistic $\widehat{\xi}(l)$ either coincides with $\widehat{\xi}_{svd}(l)$ in (14) 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 [18], p. 239.

As the first numerical example, we consider $X_t$ from the VM model (1) with $p = 5$, $T = 500$ and

$$\mu(u) = (0, 7, 14, \sin(7u), \sin(7(u + 0.2)))^\top. \tag{36}$$

The errors $Y_t$ are multivariate Gaussian i.i.d. with $\mathrm{E}\, Y_t Y_t^\top = I_5$. The true cotrending dimension for (36) is $d_1 = 3$. Observe from Fig. 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 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 $45^\circ$ 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 (27) reduces to

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \varepsilon_t. \tag{37}$$

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

$$\Pi_1 = \begin{pmatrix} 0.5 & 0.2 & 0 \\ -0.2 & -0.5 & 0.7 \\ 0.3 & 0 & -0.1 \end{pmatrix}, \quad \Pi_2 = \begin{pmatrix} 0.5 & -0.2 & 0 \\ -0.1 & 0.3 & -0.2 \\ 0.7 & 0.1 & -0.5 \end{pmatrix}. \tag{38}$$

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

$$\Pi = -(I_3 - \Pi_1 - \Pi_2) = \begin{pmatrix} 0 & 0 & 0 \\ -0.3 & -1.2 & 0.5 \\ 1 & 0.1 & -1.4 \end{pmatrix} \tag{39}$$

has rank 2. Observe from Fig. 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.

## 8. 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.

Before applying the procedures to the two data sets, we discuss how to fit the varying means model (1), as this will be used to assess its adequacy. The mean functions have to be estimated locally. Therefore, we use a kernel-based estimator

$$\widehat{\mu}(u) = \frac{1}{T} \sum_{t=1}^{T} X_t K_h\Big(u - \frac{t}{T}\Big), \quad K_h(u) = \frac{1}{h} K\Big(\frac{u}{h}\Big), \tag{40}$$

where $K$ is a kernel function and $h$ denotes the bandwidth. A kernel is a symmetric function which integrates to 1. We work throughout with the triangle kernel $K(u) = (1 - |u|)\mathbb{1}_{\{|u|<1\}}$. Furthermore, we choose $h$ by cross-validation. The estimated mean functions $\widehat{\mu}(u)$ are supposed to capture the serial dependence in the time series and, if the model is valid, lead to uncorrelatedness over time in the residuals. The smaller $h$, the more serial dependence is expected to be captured by the estimated mean functions. On the other hand, the larger $h$, the smoother are the estimated functions in

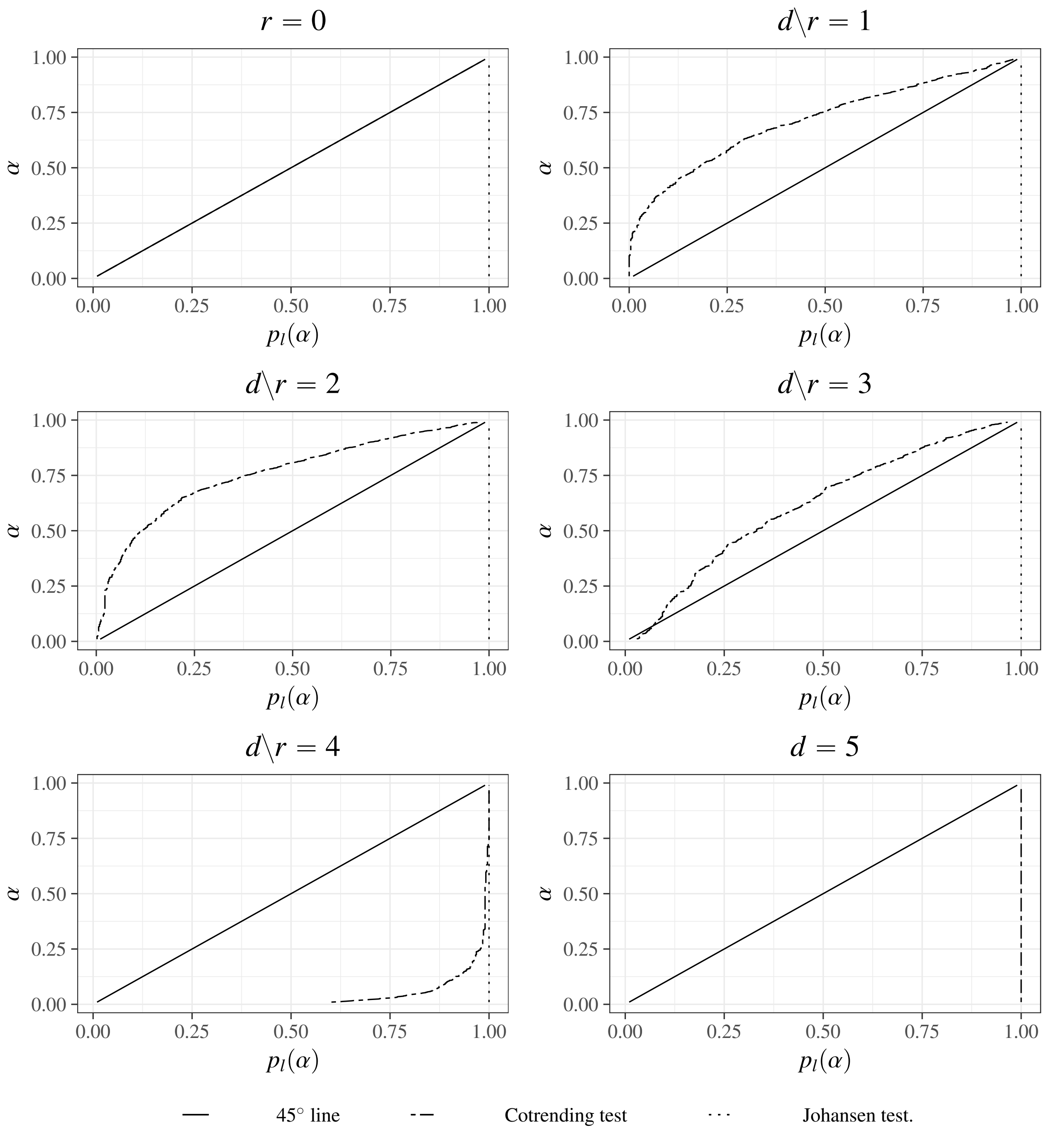


**Fig. 1:** PP-plots for a simulated VM model of a 5-dimensional time series with true cotrending dimension $d_1 = 3$. PP-plots for different values $d \in \{1, \ldots, 5\}$ and $r \in \{0, \ldots, 4\}$ are presented. The probability $\alpha \in (0, 1)$ on the vertical axis is plotted versus estimated $p_l(\alpha) = \mathrm{P}(\widehat{\xi}(l) > q_l(\alpha))$, $l = d, r$ on the horizontal axis. The values $q_l(\alpha)$ are such that $\mathrm{P}(\widehat{\xi}(l) > q_l(\alpha)) = \alpha$. The respective test statistic $\widehat{\xi}(l)$ either coincides with $\widehat{\xi}_{svd}(l)$ in (14) or the Johansen test statistic. The probability $p_l(\alpha)$ is estimated with 500 Monte Carlo replications of the corresponding test statistics.

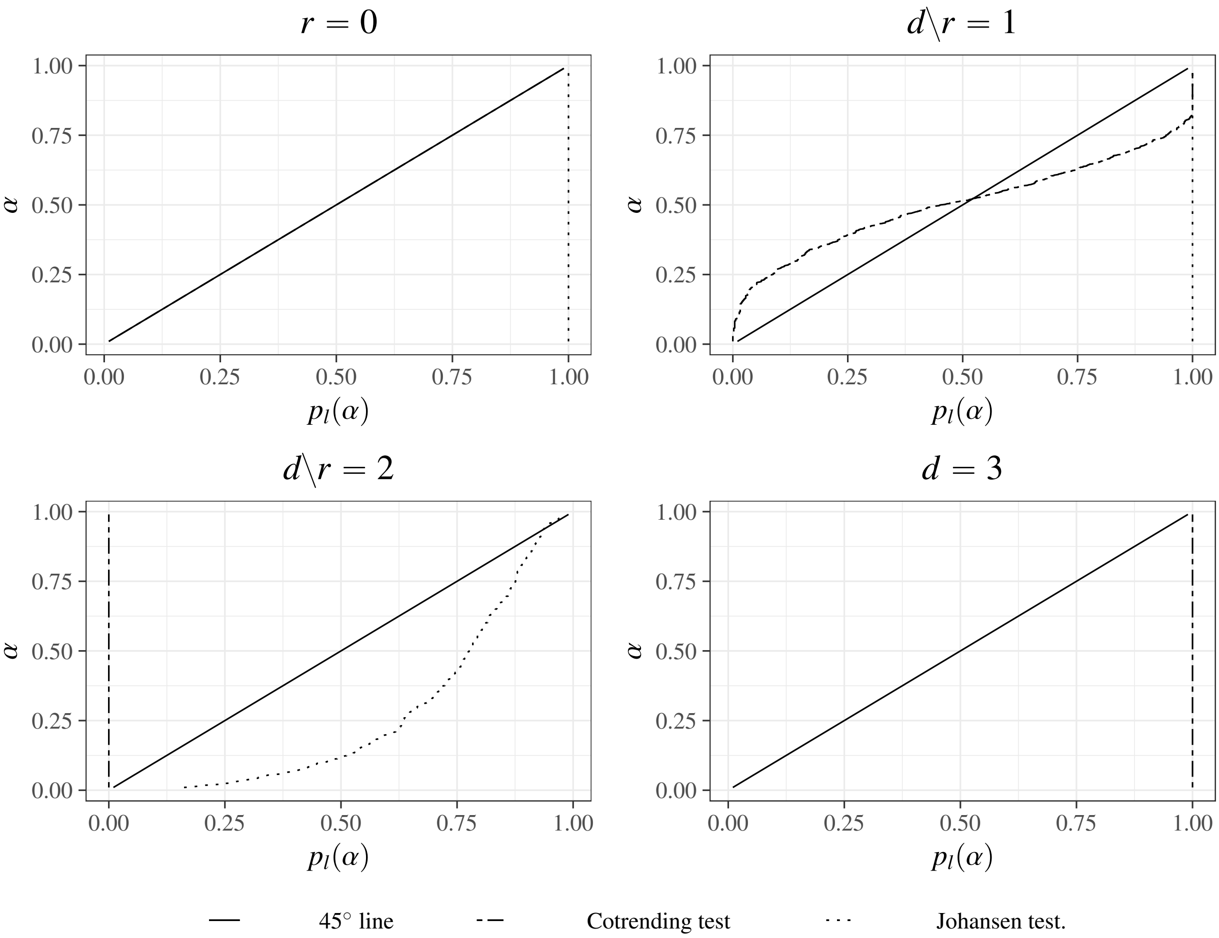


**Fig. 2:** PP-plots of a simulated 3-dimensional VAR(2) model with true cointegrating rank $r^* = 2$. Due to the different hypothesis testing problems, we present 4 plots for different values $d \in \{1, 2, 3\}$ and $r \in \{0, 1, 2\}$. The probability $\alpha \in (0, 1)$ on the vertical axis is plotted versus estimated $p_l(\alpha) = \mathrm{P}(\widehat{\xi}(l) > q_l(\alpha))$, $l = d, r$ on the horizontal axis. The values $q_l(\alpha)$ are such that $\mathrm{P}(\widehat{\xi}(l) > q_l(\alpha)) = \alpha$. The respective test statistic $\widehat{\xi}(l)$ either coincides with $\widehat{\xi}_{svd}(l)$ in (14) or the Johansen test statistic. The probability $p_l(\alpha)$ is estimated with 500 Monte Carlo replications of the corresponding test statistics.

the mean vector. We note that interestingly and perhaps surprisingly, the estimator $\widehat{M}$ in (9) does not involve directly the mean function estimators $\widehat{\mu}(u)$ in (40).

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* [26]. Previous works on cointegrated time series have used this data set; see Holden and Perman [17] and Pfaff [26]. 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 Fig. 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.

To assess the VM model validity, we estimate the mean functions locally by using the proposed estimator (40). It turns out that the bandwidth $h$ estimated by cross-validation results in non-centered residuals. However, the VM model seems to be valid for higher values of $h$ and one can observe that the correlation over time of the residuals is almost not effected by the choice of $h$. For this reason, we picked an $h$ higher than the one estimated by cross-validation. See Fig. 4 for the corresponding residuals and Fig. 5 for the sample autocorrelation and cross-autocorrelation functions which suggest that the residuals are generally uncorrelated over time.

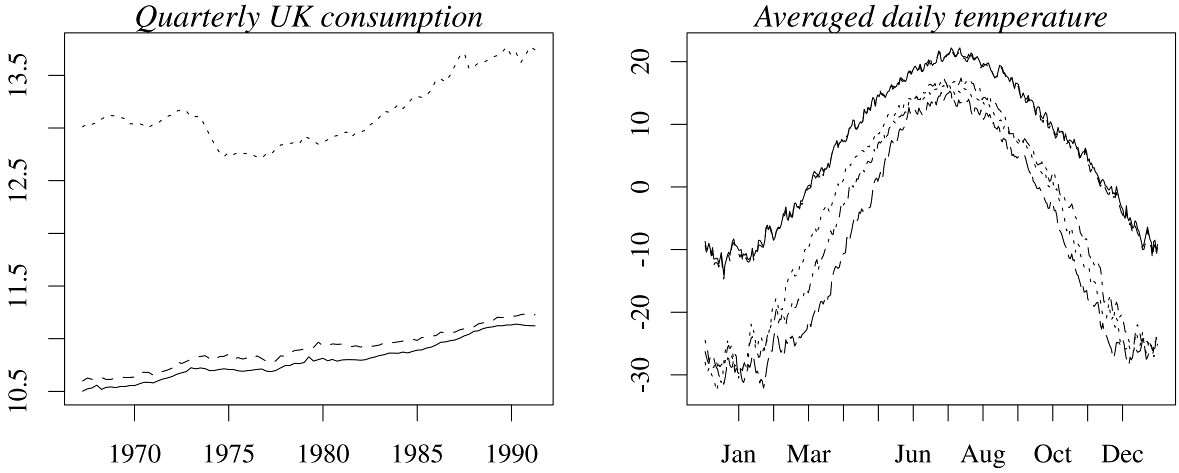


**Fig. 3:** The time plots of the quarterly consumption series in the United Kingdom from 1967 to 1991 (left plot), and the daily temperature series of five different weather stations in Canada averaged over 1960 to 1994 (right plot).

Our testing procedure estimates the cotrending dimension and the cotrending space vector as

$$\widehat{d}_1 = 1 \quad \text{and} \quad \widehat{B}_1 = \begin{pmatrix} 0.7349 & -0.6758 & -0.0571 \end{pmatrix}^\top . \tag{41}$$

(We used a 5% significance level in sequential testing for $\widehat{d}_1$.) 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 $\widehat{B}_1$ in (41) is close to zero, one might suspect that the vector

$$Q = \begin{pmatrix} 0.7349 & -0.6758 & 0 \end{pmatrix}^\top$$

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 (16) is rejected at a 5% significance level. Fig. 6 shows the asymptotic 95% confidence region (18) computed from the estimated vector $\widehat{V} = \widehat{\mathcal{B}}_1$. Observe that the first and second components (represented by the $x$- and $y$-axes, the latter being vertical) 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.

In order to estimate the cointegrating rank, we use the Akaike information criterion to determine the lag order. The selected order is two. Then, the cointegrating rank $r^*$ estimated by the Johansen test, at a 5% significance level, is $\widehat{r}^* = 1$, and coincides with the estimated cotrending dimension. The corresponding cointegrating vector, normalized to the first component of the cotrending vector in (41), is

$$\begin{pmatrix} 0.7349 & -0.6837 & -0.0460 \end{pmatrix}^\top .$$

The second and third components are also similar to those of the cotrending vector in (41).

The second data set concerns daily temperature at five different locations in Canada, namely, Montreal, Ottawa, Dawson, Yellowknife and Inuvik, available from the R package *fda* [28]. The time plot of the five series is given in the right plot of Fig. 3. The data set consists of the daily temperatures averaged over 1960 to 1994. While Montreal and Ottawa are located in the southeast of Canada, the stations in Dawson, Yellowknife and Inuvik are in the northwest of Canada and admit lower temperatures. The whole dataset consists of 35 different weather stations in Canada and is commonly used in functional data analysis.

As for the first data set, we fit a VM model to the data set by estimating the functions in the mean vector. The residuals are generally uncorrelated over time as long as the bandwidth $h$ is sufficiently close to the one estimated by cross-validation. However, we found that for $h$ too small, the residuals exhibit a trend in their mean. For this reason,

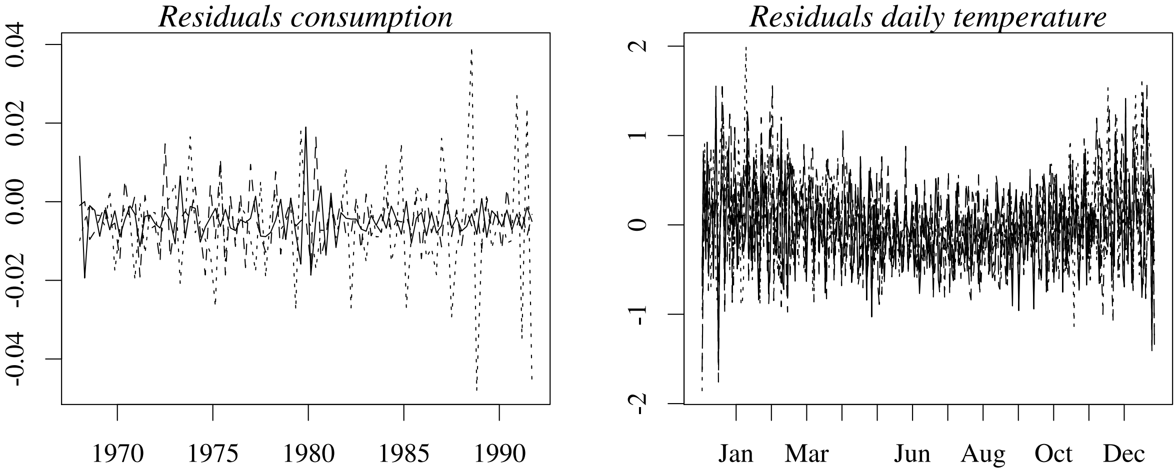


**Fig. 4:** The residuals of the consumption series (left plot), and the residuals of the daily temperature series (right plot). The residuals are obtained by subtracting $\widehat{\mu}(u)$, which are estimated as in (40), from the original time series.

we picked $h$ larger to avoid any trend in the residuals but small enough to observe uncorrelatedness over time in the residuals. We refrain from giving the corresponding plots of the autocorrelation functions due to the higher dimension. Proceeding as for the first data set above, we estimate the cotrending dimension and the cotrending space vectors as

$$\widehat{d}_1 = 2, \quad \widehat{B}_1 = \begin{pmatrix} 0.3063 & 0.2951 & 0.1323 & -0.8452 & 0.2953 \\ 0.6906 & -0.7226 & 0.0302 & 0.0021 & -0.0017 \end{pmatrix}^\top . \tag{42}$$

Replacing the small entries of the matrix $\widehat{B}_1$ in (42) with zero, leads to

$$Q = \begin{pmatrix} 0.3063 & 0.2951 & 0.1323 & -0.8452 & 0.2953 \\ 0.6906 & -0.7226 & 0.0302 & 0 & 0 \end{pmatrix}^\top ,$$

which could naturally be tested to lie in the cotrending subspace $\mathcal{B}_1$. Testing for this through the hypothesis in (16) at a 5% significance level, the null is not rejected. As a result, one gets two cotrending relations, the first involves all five data series, the second only three data series.

To apply the Johansen test, the Akaike information criterion suggests a lag order of three. Then, applying the Johansen test to estimate the cointegrating rank $r^*$ yields $\widehat{r}^* = 3$, which does not coincide with the estimated cotrending dimension. The corresponding cointegrating vectors are estimated as

$$\begin{pmatrix} 1 & -0.9476 & -0.0503 & 0.0179 & 0.0064 \\ 1 & -1.1780 & 0.0884 & 0.0819 & -0.0323 \\ 1 & -0.5618 & 0.2276 & -0.7389 & 0.2178 \end{pmatrix}^\top ,$$

which are normalized one.

**Remark 3.** In Section 5, we discussed possible extensions of the VM model for temporal dependence in the error terms. For comparison, we developed a cotrending test in terms of the estimator (19) and the limiting covariance matrix in (22) and applied it to the two datasets. For the consumption data, the resulting cotrending rank coincides with the result in (41). In the second example of the daily temperatures, the estimated cotrending rank dropped to one.

## 9. 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.

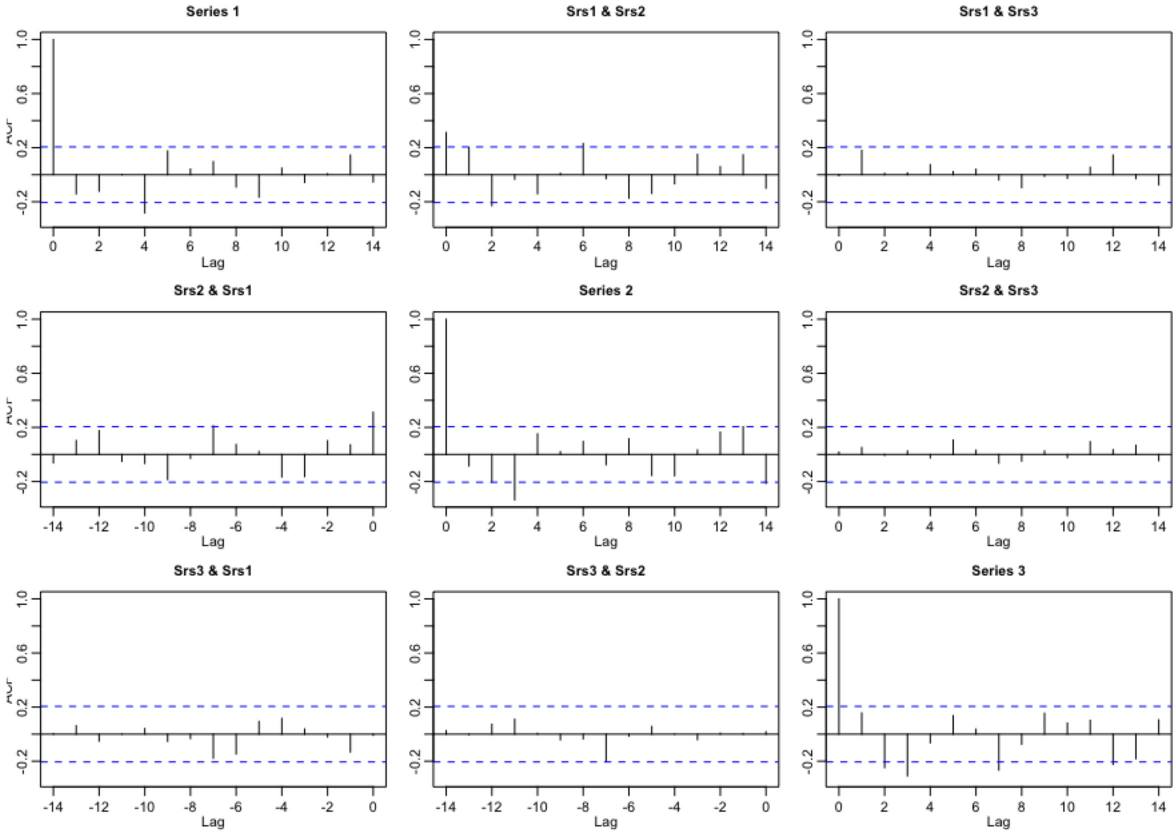


**Fig. 5:** The sample autocorrelation and cross-autocorrelation function of the residuals of the consumption series. The residuals are obtained by subtracting $\widehat{\mu}(u)$, which are estimated as in (40), from the original time series.

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.

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.

**Acknowledgement**

The authors thank two anonymous Reviewers for their comments which helped improving the paper. 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 and, in particular, Vladas Pipiras for their hospitality, and also the DFG (RTG 2131) for financial support. The second author was supported in part by the NSF grant DMS-1712966.

**Supplementary material**

Supplement to "Cotrending: testing for common deterministic trends in varying means model". For the sake of brevity, we moved Appendix B, which contains complementary results, to the supplementary document Supplementary material.

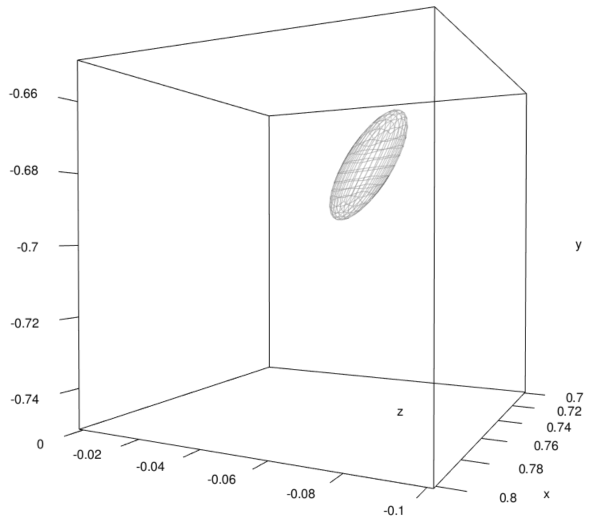


**Fig. 6:** Visualization of the asymptotic 95% confidence region for $\mathcal{B}_1$ associated with the estimated vector $\widehat{\mathcal{B}}_1$ in (41) for the consumption data of the United Kingdom.

## Appendix A. Proofs

**Proof of Proposition 1:** To prove the asymptotic normality of the estimator $\widehat{M}_S$ in (8), we first consider $\widehat{M}$ in (9) and prove its asymptotic normality using a result on possibly nonstationary $m$-dependent random variables in Sen [31].

The estimator $\widehat{M}$ in (9) can be written as

$$\widehat{M} = \frac{1}{T}\sum_{t=1}^{T-1}(X_t - \bar{X}_T)(X_{t+1} - \bar{X}_T)^\top = R_1 - R_2 - R_3 + R_4 \tag{A.1}$$

with

$$R_1 = \frac{1}{T}\sum_{t=1}^{T-1}\left(\mu\Big(\frac{t}{T}\Big) - \bar{\mu}_T\right)\left(\mu\Big(\frac{t+1}{T}\Big) - \bar{\mu}_T\right)^\top, \quad R_2 = \frac{1}{T}\sum_{t=1}^{T-1}(Y_t\bar{Y}_T^\top + \bar{Y}_T Y_{t+1}^\top) - \bar{Y}_T\bar{Y}_T^\top,$$

$$R_3 = \frac{1}{T}\sum_{t=1}^{T-1}\left[\bar{Y}_T\left(\mu\Big(\frac{t+1}{T}\Big) - \bar{\mu}_T\right)^\top + \left(\mu\Big(\frac{t}{T}\Big) - \bar{\mu}_T\right)\bar{Y}_T^\top\right],$$

$$R_4 = \frac{1}{T}\sum_{t=1}^{T-1}\left[Y_tY_{t+1}^\top + Y_t\left(\mu\Big(\frac{t+1}{T}\Big) - \bar{\mu}_T\right)^\top + \left(\mu\Big(\frac{t}{T}\Big) - \bar{\mu}_T\right)Y_{t+1}^\top\right],$$

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

$$R_1 = M + O\Big(\frac{1}{T}\Big) \tag{A.2}$$

and the second term $R_2$ is such that

$$TR_2 = T\bar{Y}_T\bar{Y}_T^\top + o_{\mathrm{p}}(1) = o_{\mathrm{p}}(1). \tag{A.3}$$

The term $R_3$ satisfies

$$\sqrt{T}R_3 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t\frac{1}{T}\sum_{t=1}^{T-1}\left(\mu\Big(\frac{t+1}{T}\Big) - \bar{\mu}_T\right)^\top + \frac{1}{T}\sum_{t=1}^{T-1}\left(\mu\Big(\frac{t}{T}\Big) - \bar{\mu}_T\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t^\top = o_{\mathrm{p}}(1) \tag{A.4}$$

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

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

$$w_t = \operatorname{vec}(\Lambda)^\top \operatorname{vec}(W_t), \qquad W_t = Y_tY_{t+1}^\top + Y_t\Big(\mu\Big(\frac{t+1}{T}\Big)-\bar{\mu}_T\Big)^\top + \Big(\mu\Big(\frac{t}{T}\Big)-\bar{\mu}_T\Big)Y_{t+1}^\top.$$

The multivariate sequence $\{W_t\}$ is 1-dependent and so is the univariate sequence $\{w_t\}$. Lemma 2.2 in Sen [31] requires the moment condition $\mathrm{E}\,|w_t|^{2+\delta} < \infty$ for some $\delta > 0$ and for all $t$. (The moment condition in Sen [31] is stated with $\delta = 1$ but the proof also works for $\delta > 0$.) Let $k = 2+\delta$ and $c$ be a generic constant that depends on $p$ and can change from line to line. Set $W_t = (W_{ij,t})_{i,j\in\{1,\dots,p\}}$ and let a single subscript $i$ refer to the $i$th component of the respective vector. Then,

$$\begin{aligned}\mathrm{E}\,|w_t|^k &= \mathrm{E}\,|\operatorname{tr}(\Lambda^\top W_t)|^k \leqslant \sum_{i,j=1}^{p} p^{2(k-1)}\,\mathrm{E}\,|\lambda_{ji}W_{ij,t}|^k \leqslant c \max_{1\leqslant i,j\leqslant p}|\lambda_{ij}|^k \sum_{i,j=1}^{p}\mathrm{E}\,|W_{ij,t}|^k \\ &\leqslant c3^{k-1}\sum_{i,j=1}^{p}\Big(\mathrm{E}\,|Y_{i,t}Y_{j,t+1}|^k + \mathrm{E}\,\Big|Y_{i,t}\Big(\mu_j\Big(\frac{t+1}{T}\Big)-\bar{\mu}_{j,T}\Big)\Big|^k + \mathrm{E}\,\Big|\Big(\mu_i\Big(\frac{t+1}{T}\Big)-\bar{\mu}_{i,T}\Big)Y_{j,t}\Big|^k\Big),\end{aligned} \tag{A.5}$$

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

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

since $\mathrm{E}\,\|Z_0\|^k < \infty$, and the piecewise continuity of $\mu$ and $\sigma^2$. Combining (A.2), (A.3), (A.4), (A.5) and Lemma 2 below, Lemma 2.2 in Sen [31] gives

$$\sqrt{T}\operatorname{vec}(\widehat{M}-M) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\widetilde{C})$$

with $\widetilde{C}$ as in (A.7). Note that $D_p^+\operatorname{vec}(A) = \operatorname{vech}(A)$ and $D_p^+N_p = D_p^+$ with $N_p = \frac{1}{2}(I_{p^2}+K_p)$, where $K_p$ denotes the so-called commutation matrix, which transforms $\operatorname{vec}(A)$ into $\operatorname{vec}(A^\top)$ for a matrix $A \in \mathbb{R}^{p\times p}$. Furthermore, $N_p\operatorname{vec}(\widehat{M}) = \operatorname{vec}(\widehat{M}_S)$. These observations yield

$$\sqrt{T}\operatorname{vech}(\widehat{M}_S - M) \xrightarrow{\mathrm{d}} \mathcal{N}(0,C)$$

with $C$ as in (10). □

The next auxiliary result was used in the proof of Proposition 1 above.

**Lemma 2.** *Suppose that the assumptions of Proposition 1 hold. Let $\widehat{M}$ be the estimator in* (9) *and $R_4$ be the last term in the decomposition* (A.1). *Then, the covariance matrices of $\widehat{M}-M$ and $R_4$ satisfy*

$$\mathrm{E}(\operatorname{vec}(\widehat{M}-M)(\operatorname{vec}(\widehat{M}-M))^\top) = \mathrm{E}(\operatorname{vec}(R_4)(\operatorname{vec}(R_4))^\top) + o\Big(\frac{1}{T}\Big) = \frac{1}{T}\widetilde{C} + o\Big(\frac{1}{T}\Big) \tag{A.6}$$

*with*

$$\begin{aligned}\widetilde{C} &= \int_0^1 \sigma^2(u)\otimes\sigma^2(u)du + 2\int_0^1(\mu(u)-\bar{\mu})(\mu(u)-\bar{\mu})^\top\otimes\sigma^2(u)duN_p \\ &\quad + 2\int_0^1\sigma^2(u)\otimes(\mu(u)-\bar{\mu})(\mu(u)-\bar{\mu})^\top duN_p.\end{aligned} \tag{A.7}$$

**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

$$R_{41} = \frac{1}{T}\sum_{t=1}^{T}Y_tY_{t+1}^\top,\quad R_{42} = \frac{1}{T}\sum_{t=1}^{T}\Big[Y_t\Big(\mu\Big(\frac{t+1}{T}\Big)-\bar{\mu}_T\Big)^\top + \Big(\mu\Big(\frac{t}{T}\Big)-\bar{\mu}_T\Big)Y_{t+1}^\top\Big].$$

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

For $R_{41}$, we write

$$\begin{aligned}
\mathrm{E}(\operatorname{vec}(R_{41})(\operatorname{vec}(R_{41}))^\top) &= \frac{1}{T^2}\sum_{t,r=1}^{T}\mathrm{E}\left(\operatorname{vec}(Y_tY_{t+1}^\top)(\operatorname{vec}(Y_rY_{r+1}^\top))^\top\right) \\
&= \frac{1}{T^2}\sum_{t,r=1}^{T}\mathrm{E}\left((I_p\otimes Y_t)Y_{t+1}Y_{r+1}^\top(I_p\otimes Y_r)^\top\right) = \frac{1}{T^2}\sum_{t,r=1}^{T}\mathrm{E}\left((I_p\otimes Y_t)(Y_{t+1}Y_{r+1}^\top\otimes 1)(I_p\otimes Y_r)^\top\right) \qquad (A.8)\\
&= \frac{1}{T^2}\sum_{t,r=1}^{T}\mathrm{E}(Y_{t+1}Y_{r+1}^\top\otimes Y_tY_r^\top) = \frac{1}{T}\int_0^1\sigma^2(u)\otimes\sigma^2(u)du + O\Big(\frac{1}{T^2}\Big),
\end{aligned}$$

where the second equality follows by $\operatorname{vec}(AB) = (I_q\otimes A)\operatorname{vec}(B) = (B^\top\otimes I_m)\operatorname{vec}(A)$ for an $m\times n$ matrix $A$ and an $n\times q$ matrix $B$; see Theorem 2 in Magnus and Neudecker [24], 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

$$\begin{aligned}
\mathrm{E}(\operatorname{vec}(R_{42})(\operatorname{vec}(R_{42}))^\top) &= \frac{1}{T}2\int_0^1(\mu(u)-\bar\mu)(\mu(u)-\bar\mu)^\top\otimes\sigma^2(u)duN_p \\
&\quad + \frac{1}{T}2\int_0^1\sigma^2(u)\otimes(\mu(u)-\bar\mu)(\mu(u)-\bar\mu)^\top duN_p + O\Big(\frac{1}{T^2}\Big), \qquad (A.9)
\end{aligned}$$

since for example

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t,r=1}^{T}\mathrm{E}\Big[\operatorname{vec}\Big(Y_t\Big(\mu\Big(\frac{t+1}{T}\Big)-\bar\mu_T\Big)^\top\Big)\Big(\operatorname{vec}\Big(Y_r\Big(\mu\Big(\frac{r+1}{T}\Big)-\bar\mu_T\Big)^\top\Big)\Big)^\top\Big] \\
&= \frac{1}{T^2}\sum_{t,r=1}^{T}\Big(\Big(\mu\Big(\frac{t+1}{T}\Big)-\bar\mu_T\Big)\otimes I_p\Big)\mathrm{E}(Y_tY_r^\top)\Big(\Big(\mu\Big(\frac{r+1}{T}\Big)-\bar\mu_T\Big)^\top\otimes I_p\Big) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\Big(\Big(\mu\Big(\frac{t+1}{T}\Big)-\bar\mu_T\Big)\otimes I_p\Big)\Big(1\otimes\sigma^2\Big(\frac{t}{T}\Big)\Big)\Big(\Big(\mu\Big(\frac{t+1}{T}\Big)-\bar\mu_T\Big)^\top\otimes I_p\Big) \\
&= \frac{1}{T}\int_0^1(\mu(u)-\bar\mu)(\mu(u)-\bar\mu)^\top\otimes\sigma^2(u)du + O\Big(\frac{1}{T^2}\Big),
\end{aligned}$$

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

**Proof of Proposition 2:** To prove the consistency of $\widehat{C}$ in (12), we use Theorem 2 in Andrews [2], 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 $\widehat{C}$ as

$$\frac{1}{T}\sum_{t=1}^{T-3}\Big(\frac{1}{4}(\Delta X_{t+1})^2\otimes(\Delta X_{t+3})^2 + 2((\Delta X_{t+3})^2\otimes(X_t-\bar\mu)(X_{t+1}-\bar\mu)^\top)\Big) = A_1 + A_2 + B_1 + B_2,$$

where setting $\widetilde{M}_t = \mu\big(\frac{t}{T}\big)$,

$$\begin{aligned}
A_1 &= \frac{1}{T}\sum_{t=1}^{T-3}\frac{1}{4}(\Delta\widetilde{M}_{t+1})^2\otimes(\Delta\widetilde{M}_{t+3})^2,\\
A_2 &= \frac{1}{T}\sum_{t=1}^{T-3}\frac{1}{4}\Big[(\Delta\widetilde{M}_{t+1})^2\otimes\Big((\Delta Y_{t+3})^2 + \Delta Y_{t+3}(\Delta\widetilde{M}_{t+3})^\top + \Delta\widetilde{M}_{t+3}(\Delta Y_{t+3})^\top\Big)
\end{aligned}$$

$$+\Big((\Delta Y_{t+1})^2+\Delta Y_{t+1}(\Delta\widetilde{M}_{t+1})^\top+\Delta\widetilde{M}_{t+1}(\Delta Y_{t+1})^\top\Big)\otimes(\Delta\widetilde{M}_{t+3})^2$$
$$+\Big((\Delta Y_{t+1})^2+\Delta Y_{t+1}(\Delta\widetilde{M}_{t+1})^\top+\Delta\widetilde{M}_{t+1}(\Delta Y_{t+1})^\top\Big)\otimes\Big((\Delta Y_{t+3})^2+\Delta Y_{t+3}(\Delta\widetilde{M}_{t+3})^\top+\Delta\widetilde{M}_{t+3}(\Delta Y_{t+3})^\top\Big)\Big]$$
$$=:\frac{1}{T}\sum_{t=1}^{T-3}W_{t,1},$$
$$B_1=\frac{1}{T}\sum_{t=1}^{T-3}2(\Delta\widetilde{M}_{t+3})^2\otimes(\widetilde{M}_t-\bar{\mu})(\widetilde{M}_{t+1}-\bar{\mu})^\top,$$
$$B_2=\frac{1}{T}\sum_{t=1}^{T-3}2\Big[\Big((\Delta Y_{t+3})^2+\Delta Y_{t+3}(\Delta\widetilde{M}_{t+3})^\top+\Delta\widetilde{M}_{t+3}(\Delta Y_{t+3})^\top\Big)\otimes(X_t-\bar{\mu})(X_{t+1}-\bar{\mu})^\top$$
$$+(\Delta\widetilde{M}_{t+3})^2\otimes\Big(Y_tY_{t+1}^\top+Y_t(\widetilde{M}_{t+1}-\bar{\mu})^\top+(\widetilde{M}_t-\bar{\mu})Y_{t+1}^\top\Big)\Big]=:\frac{1}{T}\sum_{t=1}^{T-3}W_{t,2}.$$

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

$$\mathrm{E}\,A_2=\frac{1}{T}\sum_{t=1}^{T-3}\frac{1}{4}\mathrm{E}\Big((\Delta\widetilde{M}_{t+1})^2\otimes(\Delta Y_{t+3})^2+(\Delta Y_{t+1})^2\otimes(\Delta\widetilde{M}_{t+3})^2+(\Delta Y_{t+1})^2\otimes(\Delta Y_{t+3})^2\Big)$$
$$=\frac{1}{T}\sum_{t=1}^{T-3}\frac{1}{4}\mathrm{E}\Big((\Delta Y_{t+1})^2\otimes(\Delta Y_{t+3})^2\Big)+O\Big(\frac{1}{T}\Big)=\int_0^1\sigma^2(u)\otimes\sigma^2(u)du+O\Big(\frac{1}{T}\Big),$$
$$\mathrm{E}\,B_2=\frac{1}{T}\sum_{t=1}^{T-3}2\,\mathrm{E}\Big((\Delta Y_{t+3})^2\otimes(X_t-\bar{\mu})(X_{t+1}-\bar{\mu})^\top\Big)=4\int_0^1\sigma^2(u)\otimes(\mu(u)-\bar{\mu})(\mu(u)-\bar{\mu})^\top du+O\Big(\frac{1}{T}\Big).$$

Then,

$$\frac{1}{T}\sum_{t=1}^{T-3}(W_{t,1}-\mathrm{E}\,W_{t,1})+\frac{1}{T}\sum_{t=1}^{T-3}\mathrm{E}\,W_{t,1}-\int_0^1\sigma^2(u)\otimes\sigma^2(u)du=\frac{1}{T}\sum_{t=1}^{T-3}(W_{t,1}-\mathrm{E}\,W_{t,1})+O\Big(\frac{1}{T}\Big),$$
$$\frac{1}{T}\sum_{t=1}^{T-3}(W_{t,2}-\mathrm{E}\,W_{t,2})+\frac{1}{T}\sum_{t=1}^{T-3}\mathrm{E}\,W_{t,2}-4\int_0^1\sigma^2(u)\otimes(\mu(u)-\bar{\mu})(\mu(u)-\bar{\mu})^\top du=\frac{1}{T}\sum_{t=1}^{T-3}(W_{t,2}-\mathrm{E}\,W_{t,2})+O\Big(\frac{1}{T}\Big),$$

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 $W_{t,1}-\mathrm{E}\,W_{t,1}$ and $W_{t,2}-\mathrm{E}\,W_{t,2}$ componentwise. We write

$$R_{1,t}=(W_{t,1}-\mathrm{E}\,W_{t,1})_{ij},\qquad R_{2,t}=(W_{t,2}-\mathrm{E}\,W_{t,2})_{ij},$$

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

$$\mathrm{E}\,|R_{1,t}|^{2+\delta}<\infty\quad\text{and}\quad\mathrm{E}\,|R_{2,t}|^{2+\delta}<\infty,\quad t\in\{1,\dots,T\}$$

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

**Proof of Proposition 4:** Following the notation in Section 6.2, we decompose $X_t$ given by its Granger representation (32) in accordance to the zero and nonzero rows of the matrix $L$ in (33) into

$$X_t=\begin{pmatrix}X_{1,t}\\X_{2,t}\end{pmatrix}=\begin{pmatrix}L_nZ_{1,t}+I_{1,n}Z_{2,t}+I_{1,n}\widetilde{X}_0\\I_{2,p-n}Z_{2,t}+I_{2,p-n}\widetilde{X}_0,\end{pmatrix},\tag{A.10}$$

where

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

Note that

$$\frac{1}{T^{1/2}} \widetilde{X}_0 = o_{\mathrm{p}}(1), \quad \frac{1}{T} \sum_{t=1}^{T} Z_{2,t} = o_{\mathrm{p}}(1), \quad \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t = o_{\mathrm{p}}(1), \tag{A.11}$$

where the second relation follows by Proposition 6.3.10 in Brockwell and Davis [4]. As in the proof of Proposition 1 we first investigate the convergence result for $\widehat{M}$ in (9). The normalized estimator $\widehat{M}$ can be written as

$$\Delta_{2,T}^{-\frac{1}{2}} \widehat{M} \Delta_{2,T}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{T} R_{11} & \frac{1}{T^{1/2}} R_{12} \\ \frac{1}{T^{1/2}} R_{21} & R_{22} \end{pmatrix},$$

where

$$R_{ij} = \frac{1}{T} \sum_{t=1}^{T-1} (X_{i,t} - \bar{X}_i)(X_{j,t+1} - \bar{X}_j)^\top, \quad \bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{i,t}, \quad i, j \in \{1, 2\}$$

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

$$\begin{aligned} \frac{1}{T} R_{11} &= \frac{1}{T^2} \sum_{t=1}^{T-1} (X_{1,t} - L_n \bar{Z}_1)(X_{1,t+1} - L_n \bar{Z}_1)^\top + o_{\mathrm{p}}(1) \\ &= \frac{1}{T^2} \sum_{t=1}^{T-1} (L_n Z_{1,t} + I_{1,n} Z_{2,t} - L_n \bar{Z}_1)(L_n(Z_{1,t} + \varepsilon_{t+1}) + I_{1,n} Z_{2,t+1} - L_n \bar{Z}_1)^\top + o_{\mathrm{p}}(1) \\ &= \frac{1}{T^2} \sum_{t=1}^{T-1} L_n (Z_{1,t} - \bar{Z}_1)(Z_{1,t} - \bar{Z}_1)^\top L_n^\top + o_{\mathrm{p}}(1), \end{aligned} \tag{A.12}$$

where the first, second and third equalities follow by (A.11) and Lemma B.1, (i), (ii) and (iv), below. Then, (A.12) and Lemma 3.1, (c) in Phillips and Durlauf [27] yield

$$\frac{1}{T} R_{11} = \frac{1}{T^2} \sum_{t=1}^{T} L_n (Z_{1,t} - \bar{Z}_1)(Z_{1,t} - \bar{Z}_1)^\top L_n^\top + o_{\mathrm{p}}(1) \xrightarrow{\mathrm{d}} L_n \Sigma_\varepsilon^{\frac{1}{2}} Z \Sigma_\varepsilon^{\frac{1}{2}} L_n^\top.$$

The matrix $R_{22}$ contains only stationary components. Its convergence

$$R_{22} = \frac{1}{T} \sum_{t=1}^{T-1} I_{2,p-n} (Z_{2,t} - \bar{Z}_2)(Z_{2,t+1} - \bar{Z}_2)^\top I_{2,p-n}^\top + o_{\mathrm{p}}(1) \xrightarrow{\mathrm{p}} \Upsilon_{p-n}(1)$$

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

$$\Upsilon_{p-n}(1) = I_{2,p-n} \operatorname{Cov}(Z_{2,0}, Z_{2,1}) I_{2,p-n}^\top = I_{2,p-n} \sum_{j=0}^{\infty} \widetilde{L}_j \widetilde{L}_{j+1}^\top I_{2,p-n}^\top$$

denotes the autocovariances of order one.

The third term $R_{12}$ satisfies

$$\begin{aligned} \frac{1}{T^{1/2}} R_{12} &= \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} (L_n Z_{1,t} + I_{1,n} Z_{2,t} - L_n \bar{Z}_1)(I_{2,p-n} Z_{2,t+1} - I_{2,p-n} \bar{Z}_2)^\top + o_{\mathrm{p}}(1) \\ &= \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} L_n (Z_{1,t} - \bar{Z}_1)(Z_{2,t+1} - \bar{Z}_2)^\top I_{2,p-n}^\top + o_{\mathrm{p}}(1) \\ &= \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} L_n \Big(Z_{1,t} Z_{2,t+1}^\top - \frac{1}{T^{1/2}} \bar{Z}_1 \bar{Z}_2^\top\Big) I_{2,p-n}^\top + o_{\mathrm{p}}(1), \end{aligned} \tag{A.13}$$

where the first equality follows by (A.11) and the second equality by Lemma B.1, (iv). Finally, $\frac{1}{T^{1/2}}R_{12} = o_{\mathrm{p}}(1)$, since the product of the normalized sample means $\frac{1}{T^{1/2}}\bar{Z}_1\bar{Z}_2 = o_{\mathrm{p}}(1)$ by Lemma B.1, (iii) and (A.11). The remaining term in the last line of (A.13) satisfies $\frac{1}{T^{3/2}}\sum_{t=1}^{T-1} Z_{1,t}Z_{2,t+1}^\top = o_{\mathrm{p}}(1)$ by Lemma B.1 (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 $\widehat{M}_S$ is a consequence of the established convergence of $\widehat{M}$. □

We conclude with the proof of Proposition 3. The proof is based on the so-called Davis-Kahan theorem. For completeness, we give one version of the theorem as stated in Samworth et al. [30, Theorem 2] here.

**Theorem 1.** (Davis-Kahan theorem) *Let $M, \widehat{M}_S \in \mathbb{R}^{p\times p}$ be symmetric, with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ and $\widehat{\lambda}_1 \geqslant \cdots \geqslant \widehat{\lambda}_p$, respectively. Fix $1 \leqslant r \leqslant s \leqslant p$ and assume that $\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\} > 0$, where $\lambda_0 = \infty$ and $\lambda_{p+1} = -\infty$. Let $d := s - r + 1$, and let $V = (v_r, \ldots, v_s) \in \mathbb{R}^{p\times d}$ and $\widehat{V} = (\widehat{v}_r, \ldots, \widehat{v}_s) \in \mathbb{R}^{p\times d}$ have orthonormal columns satisfying $Mv_j = \lambda_j v_j$ and $\widehat{M}_S\widehat{v}_j = \widehat{\lambda}_j\widehat{v}_j$ for $j \in \{r, \ldots, s\}$. Then, there exists an orthogonal matrix $\widehat{O} \in \mathbb{R}^{p\times p}$ such that*

$$\|\widehat{V}\widehat{O} - V\|_F \leqslant 2^{\frac{3}{2}} \frac{\|\widehat{M}_S - M\|_F}{\min\{\lambda_{i-1} - \lambda_i, \lambda_{i+d-1} - \lambda_{i+d}\}}. \tag{A.14}$$

**Proof of Proposition 3:** By Theorem 1, there is an orthogonal matrix $\widehat{O} \in \mathbb{R}^{d\times d}$, such that (A.14) holds. Set $\tau = \varepsilon \min\{\lambda_{i-1} - \lambda_i, \lambda_{i+d-1} - \lambda_{i+d}\}2^{-\frac{3}{2}}$. Then, by applying (A.14) and Chebyshev's inequality, we get for all $\varepsilon > 0$,

$$\mathrm{P}(\|\widehat{V}\widehat{O} - V\|_F \geqslant \varepsilon) \leqslant \mathrm{P}(\|\widehat{M}_S - M\|_F \geqslant \tau) = \mathrm{P}\left(\Big(\sum_{i,j=1}^{p} |e_i^\top(\widehat{M}_S - M)e_j|^2\Big)^{\frac{1}{2}} \geqslant \tau\right) \leqslant \frac{1}{\tau^2}\,\mathrm{E}\sum_{i,j=1}^{p} |e_i^\top(\widehat{M}_S - M)e_j|^2$$

$$= \frac{1}{\tau^2}\sum_{i,j=1}^{p} (\mathrm{vec}(e_ie_j^\top))^\top\Big(\frac{1}{T}N_p\widetilde{C}N_p^\top + o\Big(\frac{1}{T}\Big)\Big)\,\mathrm{vec}(e_ie_j^\top) = \frac{1}{\tau^2}\frac{1}{T}(\mathrm{tr}(N_p^2\widetilde{C}) + p^2o(1)),$$

where $\{e_i\}_{i\in\{1,\ldots,p\}}$ are $p$-dimensional unit vectors and the second to last equality is a consequence of Lemma 2 with $\widetilde{C}$ as in (A.7). □

## References


[1] T. W. Anderson, Estimating linear restrictions on regression coefficients for multivariate normal distributions, The Annals of Mathematical Statistics 22 (1951) 327–351.

[2] D. Andrews, Laws of large numbers for dependent non-identically distributed random variables, Econometric Theory 4 (1988) 458–467.

[3] D. A. J. Blythe, P. von Bünau, F. C. Meinecke, K.-R. Müller, Feature extraction for change-point detection using stationary subspace analysis, IEEE Transactions on Neural Networks and Learning Systems 23 (2012) 631–643.

[4] P. J. Brockwell, R. A. Davis, Time Series: Theory and Methods, Springer Series in Statistics, Springer-Verlag, New York, second edition, 1991.

[5] P. von Bünau, F. C. Meinecke, F. C. Király, K.-R. Müller, Finding stationary subspaces in multivariate time series, Physical Review Letters 103 (2009) 214101.

[6] G. Camba-Mendez, G. Kapetanios, Testing the rank of the Hankel covariance matrix: a statistical approach, IEEE Transactions on Automatic Control 46 (2001) 331–336.

[7] A. Cardinali, G. Nason, Costationarity of locally stationary time series using costat, Journal of Statistical Software 55 (2013).

[8] J. G. Cragg, S. G. Donald, On the asymptotic properties of LDU-based tests of the rank of a matrix, Journal of the American Statistical Association 91 (1996) 1301–1309.

[9] R. Dahlhaus, Locally stationary processes, in: T. S. Rao, S. S. Rao, C. Rao (Eds.), Time Series Analysis: Methods and Applications, volume 30, Handbook of Statistics, Elsevier, 2012, pp. 351–413.

[10] R. Dahlhaus, W. Polonik, Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes, Annals of Statistics 34 (2006) 2790–2824.

[11] S. G. Donald, N. Fortuna, V. Pipiras, On rank estimation in symmetric matrices: the case of indefinite matrix estimators, Econometric Theory 23 (2007) 1217–1232.

[12] M. L. Eaton, D. E. Tyler, On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix, The Annals of Statistics 19 (1991) 260–271.

[13] R. Engle, C. Granger, Co-integration and error correction: Representation, estimation, and testing, Econometrica 55 (1987) 251–76.

[14] L. Gill, A. Lewbel, Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models, Journal of the American Statistical Association 87 (1992) 766–776.

[15] C. Granger, Some properties of time series data and their use in econometric model specification, Journal of Econometrics 16 (1981) 121–130.

[16] F. Hayashi, Econometrics, Princeton University Press, Princeton, 2011.
[17] D. Holden, R. Perman, Unit roots and cointegration for the economist, Unit Roots and Cointegration for the Economist, Palgrave Macmillan UK, London, 1994, pp. 47–112.
[18] S. Johansen, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12 (1988) 231–254.
[19] S. Johansen, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica 59 (1991) 1551–1580.
[20] I. T. Jolliffe, Principal Component Analysis, Springer Verlag, New York, 1986.
[21] F. Kleibergen, R. Paap, Generalized reduced rank tests using the singular value decomposition, Journal of Econometrics 133 (2006) 97–126.
[22] C. Lam, Q. Yao, Factor modeling for high-dimensional time series: Inference for the number of factors, The Annals of Statistics 40 (2012) 694–726.
[23] G. S. Maddala, I.-M. Kim, Unit Roots, Cointegration, and Structural Change, Cambridge University Press, Cambridge, 1999.
[24] J. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley Series in Probability and Statistics, Wiley, New York, second edition, 1999.
[25] H. Ombao, R. von Sachs, W. Guo, SLEX analysis of multivariate nonstationary time series, Journal of the American Statistical Association 100 (2005) 519–531.
[26] B. Pfaff, Analysis of Integrated and Cointegrated Time Series with R, Springer Publishing Company, Incorporated, New York, second edition, 2008.
[27] P. C. B. Phillips, S. N. Durlauf, Multiple time series regression with integrated processes, The Review of Economic Studies 53 (1986) 473–495.
[28] J. O. Ramsay, H. Wickham, S. Graves, G. Hooker, fda: Functional Data Analysis, 2018. R package version 2.4.8.
[29] J.-M. Robin, R. J. Smith, Tests of rank, Econometric Theory 16 (2000) 151–175.
[30] R. J. Samworth, T. Wang, Y. Yu, A useful variant of the Davis-Kahan theorem for statisticians, Biometrika 102 (2014) 315–323.
[31] P. K. Sen, Asymptotic normality of sample quantiles for $m$–dependent processes, The Annals of Mathematical Statistics 39 (1968) 1724–1730.
[32] R. R. Sundararajan, V. Pipiras, M. Pourahmadi, Stationary subspace analysis of nonstationary covariance processes: Eigenstructure description and testing, Bernoulli 27 (2021) 381–418.
[33] R. R. Sundararajan, M. Pourahmadi, Stationary subspace analysis of nonstationary processes, Journal of Time Series Analysis 39 (2018) 338–355.
[34] D. E. Tyler, Asymptotic inference for eigenvectors, The Annals of Statistics 9 (1981) 725–736.