High-Dimensional Cointegration and Kuramoto Inspired Systems

Stærk-Østergaard, Jacob; Rahbek, Anders; Ditlevsen, Susanne

Published in:
SIAM Journal on Applied Dynamical Systems

DOI:
10.1137/22M1509771

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
High-Dimensional Cointegration and Kuramoto Inspired Systems

Jacob Stærk-Østergaard, Anders Rahbek, and Susanne Ditlevsen

Abstract. This paper presents a novel estimator for a nonstandard restriction to both symmetry and low rank in the context of high-dimensional cointegrated processes. Furthermore, we discuss rank estimation for high-dimensional cointegrated processes by restricted bootstrapping of the Gaussian innovations. We demonstrate that the classical rank test for cointegrated systems is prone to underestimating the true rank and demonstrate this effect in a 100-dimensional system. We also discuss the implications of this underestimation for such high-dimensional systems in general. Also, we define a linearized Kuramoto system and present a simulation study, where we infer the cointegration rank of the unrestricted \( p \times p \) system and successively the underlying clustered network structure based on a graphical approach and a symmetrized low rank estimator of the couplings derived from a reparametrization of the likelihood under this unusual restriction.

Key words. cointegration, dynamical system, bootstrap, Kuramoto, high-dimensional, stochastic process

MSC codes. 37H10, 62M10, 62P10, 62P20, 62H10, 62H15

DOI. 10.1137/22M1509771

1. Introduction. High-dimensional data is ubiquitous in contemporary statistics, with applications in many fields where laboratory equipment and techniques facilitate recording an ever-increasing number of units simultaneously. In [12] we explored a novel approach to infer network structure in a system of coupled oscillating processes by cointegration analysis. It was demonstrated how both unidirectional, bidirectional, and all-to-all coupling was inferred. However, the studied systems only included four-dimensional processes, which does not agree well with the interest in analysis tools for high-dimensional data series.

In this paper we explore how cointegration analysis performs in high dimensions in terms of both rank and parameter estimation. We assume a high-dimensional system of cointegrated processes with independent and identically distributed Gaussian innovations and a novel structure. To determine the rank of the matrix defining the couplings of the linear high-dimensional system, we use restricted bootstrapping, as presented in [1] for classical cointegrated systems, and adapt it to a high-dimensional setting.

Received by the editors July 15, 2022; accepted for publication (in revised form) by L. DeVille September 12, 2023; published electronically January 18, 2024.

https://doi.org/10.1137/22M1509771

Funding: This work was supported by the Novo Nordisk Foundation (NNF20OC0062958); by Independent Research Fund Denmark — Natural Sciences (9040-00215B) and Danish Council for Independent Research (DSF grant 015-00028B); and by European Union’s Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie grant agreement 956107, “Economic Policy in Complex Environments (EPOC).”

Department of Mathematical Sciences, University of Copenhagen, Copenhagen DK-2100, Denmark (jacobostergaard@gmail.com, susanne@math.ku.dk).

Department of Economics, University of Copenhagen, Copenhagen DK-1353, Denmark (anders.rahbek@gmail.com, anders.rahbek@econ.ku.dk).

236
We furthermore define a linearized version of the classic Kuramoto model [8] which implies a symmetric design of the couplings of the system. Estimation restricted to this nonstandard condition requires development of a novel low rank estimator of the system matrix under simultaneous restriction to both symmetry and low rank. Here, we derive such an estimator by reparametrizing the likelihood function under symmetry restrictions. This estimator diverges from the standard maximum likelihood estimation using reduced rank regression.

A simulation study highlights the performance of this estimator and discusses the implications of underestimating the true rank of the coupling matrix. Finally, using the estimated coupling matrix, we use a graphical method to infer the underlying cluster structure of the linearized Kuramoto system.

In the following, $I_p$ denotes the $p$-dimensional identity matrix and $M'$ denotes the transpose of a matrix $M$. For a real matrix $M \in \mathbb{R}^{p \times r}$, $M_\perp \in \mathbb{R}^{p \times (p-r)}$ denotes any matrix being an orthogonal complement, such that $M'M = 0$, with $M$ and $M_\perp$ both of full rank. The matrix determinant operator is denoted by $|\cdot|$. Finally, time is assumed discrete with positive index. Initial values are either assumed known or explicitly stated.

2. Cointegration. Assume that $y_n = (y_{1n}, \ldots, y_{pn})^\prime \in \mathbb{R}^p$ is a discrete time vector autoregressive process,

$$y_n = Ay_{n-1} + \mu + \varepsilon_n,$$

where $A \in \mathbb{R}^{p \times p}$, $\varepsilon_n$ is a $p$-dimensional Gaussian white noise process with covariance matrix $\Omega$, and $\mu \in \mathbb{R}^p$ is a deterministic term. For simplicity, we assume $\Omega$ diagonal throughout the paper, such that all dependencies between variables are captured in the autoregressive matrix $A$. The characteristic polynomial for (2.1) is the determinant of $I_p - A\zeta$ for $\zeta \in \mathbb{C}$. If the roots of the characteristic polynomial are all outside the unit circle, then the initial values of $y_n$ can be given a distribution such that $y_n$ is stationary; see [7].

If the characteristic polynomial of (2.1) contains the root $\zeta = 1$, then there exists no stationary solution of $y_n$, and we say that the process is integrated. This is usually denoted by $I(d)$, meaning integrated of order $d$. Taking differences $d$ times, the process becomes stationary, often termed $I(0)$. An $I(1)$ process is therefore (nonstationary and) integrated of order 1; however, taking first differences, the process becomes stationary, $I(0)$. Here, in particular (see [7]), $\Pi = A - I_p$ will have reduced rank $r < p$ and can be written as $\Pi = \alpha \beta'$. The process $y_n$ is then integrated of order one, $I(1)$, with $r$ cointegrating relations $\beta'y_n$ under regularity conditions presented in section 2.1.

In this paper we will only deal with $I(1)$ processes, so when we refer to $y_n$ as integrated, we implicitly mean that $y_n$ is integrated of order 1.

2.1. Cointegrated process. Consider first the well-known results from cointegration analysis in [7]. Rewrite (2.1) with $\Pi = A - I_p$ in the vector-error-correction-model form as

$$\Delta y_n = y_n - y_{n-1} = \Pi y_{n-1} + \mu + \varepsilon_n.$$

If $|I - A\zeta| = 0$ implies $|\zeta| > 1$, then $y_n$ has a stationary representation as an $I(0)$ process. In particular, $\Pi$ has full rank $p$ and all linear combinations of $y_n$ are stationary. If the $(p \times p)$-dimensional matrix $\Pi$ has reduced rank $r < p$, then $\Pi = \alpha \beta'$ with $\alpha, \beta$, $(p \times r)$-dimensional...
matrices of rank $r$. Moreover, the process $y_t$ is $I(1)$ with $r$ cointegrating stationary relations $\beta'y_t$, provided that the spectral radius $\rho(I_r + \beta'\alpha) < 1$. This we refer to as the $I(1)$ conditions in the following.

Note that if $r = 0$, the process $y_n$ is $I(1)$ with a linear trend if $\mu \neq 0$, but with no cointegration, while if $r = p$ (and $\rho(I_p + A) < 1$), then $y_n$ is $I(0)$, and $p$ stationary linear combinations exist. Under the reduced rank $r$, the system can be written as

$$\Delta y_n = \alpha' y_{n-1} + \mu + \varepsilon_n,$$

with $\beta$ containing the $r$ cointegration vectors and $\alpha$ the loadings or adjustment coefficients.

Note that the entries of $\alpha$ and $\beta$ are not uniquely identified, since we can use any nonsingular transformation to obtain similar results. Rather, we identify the subspaces $\text{sp}(\alpha), \text{sp}(\beta) \in \mathbb{R}^r$, that is, the subspaces spanned by the columns of $\alpha, \beta$. We can therefore choose the normalization

$$\beta^* = (c' \beta)^{-1}, \quad \text{with } c = (I_r, 0_{p-r \times r})'$$

of $\beta$ in order to identify parameters uniquely, given that the normalization is empirically valid.

2.2. Estimation of parameters. Consider observations $(y_1, \ldots, y_N)$ from the process (2.1) and denote by $H_r$ the hypothesis $H_r : \text{rank}(\Pi) \leq r$ for $r = 0, \ldots, p$. Then the set of hypotheses $H_0, \ldots, H_r$ is nested,

$$H_0 \subseteq H_1 \subseteq \cdots \subseteq H_p,$$

where $H_p$ corresponds to the unrestricted model. The likelihood ratio test (LRT) that compares $H_r$ and $H_p$ is applied sequentially for $r = 0, 1, \ldots, p - 1$ and continued until $H_r$ against $H_p$ cannot be rejected, and thus determine the number of cointegrating relations for $y_t$. The LRT statistic is given by

$$(2.3) \quad -2 \log Q(H_r | H_p) = \sum_{i=r+1}^{p} \hat{\lambda}_i,$$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$ are the solutions to an eigenvalue problem; see [7]. The asymptotic distribution of (2.3) is nonstandard, and previous work on estimation in high-dimensional cointegrated systems discussed the problem of finding spurious cointegration relations in these when using the classic Johansen’s test; see [11]. The authors derive weak convergence towards the so-called Wachter distribution in high-dimensional settings when the limit of the ratio between the dimensionality and number of observations exists.

Here, we opt for a different approach by applying bootstrapping techniques as presented by [1], such that it is possible to infer the critical thresholds of the likelihood ratio test, numerically. Specifically, given observed data $\{y_n\}_{n=1}^N$, bootstrap sequences $\{y_{n}^{(m)}\}_{n=1}^N$ for $m = 1, \ldots, M$ are simulated using the wild bootstrap, where the residuals are scaled by standard Gaussian distributed random variables, and for each sequence the LRT statistic $LRT^{(m)}$ is recomputed. The empirical quantiles of $\{LRT^{(m)}\}_{m=1}^M$ are then used for inference. With $r$ determined, $\hat{\beta}$ is given by the $r$ eigenvectors corresponding to $\hat{\lambda}_i, i = 1, \ldots, r$ and the parameter estimates $\hat{\alpha}, \hat{\mu}, \hat{\Sigma}$ follow by ordinary least squares estimation as described in [7].
3. Linear Kuramoto-type system. The classic Kuramoto model [8] defines a system of \( p \) coupled processes through the differential equations

\[
\frac{d\theta_i}{dt} = \omega_i + \frac{\kappa}{p} \sum_{j=1}^{p} \sin(\theta_j - \theta_i) \quad \text{for } i = 1, \ldots, p,
\]

where \( t \) denotes continuous time. In the following, however, we consider observations of a system in discrete time, inspired by the Kuramoto model. The variables \( \theta_i \) are interpreted as the phases of limit cycle oscillators, with intrinsic frequencies \( \omega_i \), and \( \kappa \) denotes the strength of the coupling between the oscillators. By linearizing the sine function around 0 (phases synchronized) we can write the linearized Kuramoto model as

\[
\frac{d\theta}{dt} = \omega + \Pi \theta,
\]

where \( \theta = (\theta_1, \ldots, \theta_p)' \), \( \omega = (\omega_1, \ldots, \omega_p)' \), and \( \Pi \in \mathbb{R}^{p \times p} \) with the simple structure

\[
\Pi = \kappa \left( \begin{array}{cccc}
1 - p & 1 & \ldots & 1 \\
1 & 1 - p & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & 1 - p
\end{array} \right)
\]

and reduced rank(\( \Pi \)) = \( p - 1 \). Note that this linearization only holds for \( (\theta_j - \theta_i) \in [-\frac{\pi}{2}, \frac{\pi}{2}] \) for all \( i, j = 1, \ldots, p \). In general, replacing the sine function in (3.1) with any function \( f \) such that \( f(0) = 0 \) and \( f'(0) = 1 \) will lead to (3.2), which could therefore also be referred to as a Laplacian model. However, in this paper we refer to a model of this type as a Kuramoto-type system to emphasize the origin of the model as well as the connection with (synchronized) phase coupled oscillators.

Consider now the following network structure inspired by the linearized version of (3.1). Let \( \Pi \) denote a symmetric \( p \times p \) matrix with a block structure

\[
\Pi = \left( \begin{array}{ccc}
\Pi_1 & 0 & \ldots & 0 \\
0 & \Pi_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \Pi_k
\end{array} \right),
\]

where \( \Pi_i \in \mathbb{R}^{p_i \times p_i}, i = 1, \ldots, k, \) are symmetric, reduced rank matrices with rank(\( \Pi_i \)) = \( r_i = p_i - 1, i = 1, \ldots, k, \) of the form given in (3.2), and \( \sum_{i=1}^{k} p_i = p \). Hence, (3.3) implicitly defines a network of clusters. Note that the \( p_i \)'s (and hence the \( r_i \)'s) do not need to be identical. The network structure in (3.3) allows for cointegration since each \( \Pi_i \) has reduced rank, and rank(\( \Pi \)) = \( \sum_{i=1}^{k} \text{rank}(\Pi_i) = \sum_{i=1}^{k} (p_i - 1) \). This implies that rank(\( \Pi \)) = \( p - k \). Interpreting this as a cointegrated system means that each cluster in (3.3) is driven by a stochastic \( I(1) \) trend, and hence the number of stochastic trends equals the number of independent clusters in the system. Since each cluster is \( I(1) \), the system as a whole is \( I(1) \).
The structure of $\Pi$ in (3.3) implies that
\[
\Pi = \text{diag}(\Pi_1, \ldots, \Pi_k)
= \text{diag}(\alpha_1\beta_1', \ldots, \alpha_k\beta_k')
= \text{diag}(\kappa_1 p^{-1}_1 \beta_1 \delta_1 \beta_1', \ldots, \kappa_1 p^{-1}_k \beta_k \delta_k \beta_k'),
\]
where $\text{diag}(\cdot)$ refers to block diagonalization. The last equation implies that $\alpha_i = \kappa_i p^{-1}_i \beta_i \delta_i$, where $\kappa_i$ is a scalar defining the coupling strength of the $i$th cluster and $\delta_i$ is a square $r_i \times r_i$ matrix of full rank. The following lemma states that this construction ensures that $\Pi$ is symmetric.

**Lemma 3.1.** If $\Pi = \Pi'$ is a symmetric $p \times p$ matrix of reduced rank $r < p$, then there exists a $p \times r$ matrix $\beta$ of rank $r$ with $\text{sp}(\beta) = \text{sp}(\Pi)$ and a symmetric $r \times r$ matrix $\delta$ of rank $r$, such that
\[
\Pi = \beta \delta \beta'.
\]

**Proof.** It follows that
\[
\Pi = VAV', V'V = VV' = I_p,
\]
where $V = (v_1, \ldots, v_p)$ are the eigenvectors corresponding to the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$ in $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ of which $p - r$ eigenvalues equal zero; $\lambda_{r+1}, \ldots, \lambda_p = 0$. Let $\Lambda_{(r)} = \text{diag}(\lambda_1, \ldots, \lambda_r)$ denote the nonzero eigenvalues, and $V_{(r)}$ the corresponding eigenvectors; then
\[
\Pi = VAV' = V_{(r)}\Lambda_{(r)} V_{(r)}', V_{(r)} V_{(r)} = V_{(r)} V_{(r)}' = I_r.
\]
With $\beta = V_{(r)} \psi$ and $\delta = \psi^{-1} \Lambda_{(r)} \psi^{-1}$ for a symmetric $r \times r$ matrix $\psi$ of full rank $r$, the result follows.

Adding the restriction that the rows/columns of $\Pi$ must sum to 0, a possible construction of (3.3) is to choose $\delta_i$ and $\beta_i$ such that
\[
\delta_i = \begin{pmatrix}
-r_i & 1 & \ldots & 1 \\
1 & -r_i & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & -r_i
\end{pmatrix}
\text{ is } r_i \times r_i \text{ and } \beta_i = \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-1 & -1 & \ldots & -1
\end{pmatrix}
\text{ is } p_i \times r_i,
\]
implicating that
\[
\beta_i \delta_i \beta_i' = \begin{pmatrix}
-r_i & 1 & \ldots & 1 \\
1 & -r_i & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & -r_i
\end{pmatrix},
\]
which is symmetric, and rows/columns sum to 0 (since it is \( r_i + 1 \times r_i + 1 \)). Furthermore, the only two parameters necessary to describe the interaction within each cluster are

1. \( p_i = r_i + 1 \), the size of the cluster, and
2. \( c_i = \frac{\alpha_i}{\beta_i} \), the strength of coupling.

Each cluster \( i \) corresponds to (3.2). We will refer to a cluster of size \( p_i \) as a \( p_i \)-cluster; hence a 2-cluster consist only of two coupled units. This definition of a cluster does not exclude 1-clusters, since this is just a single independent unit with \( \delta_i = \beta_i = 0 \). The construction (3.6) implies that the full \( \Pi \) matrix can be decomposed into (sparse) \( \beta, \delta \) matrices, where \( \delta \) is a block diagonal matrix, i.e., \( \Pi = \beta \delta \beta' \).

We define a linear Kuramoto-type system as a set of \( k \) independent clusters of sizes \( p_i, i = 1, \ldots, k \), with coupling matrix on the form (3.3), where each cluster has a coupling strength equal to \( c_i \) and a coupling structure equivalent to (3.6).

4. Estimation given structural restrictions. Under \( H_r, \) [7] derives an LRT for linear restrictions on \( \alpha, \beta \). These are formulated as \( \text{sp}(\alpha) \subset \text{sp}(A) \) and \( \text{sp}(\beta) \subset \text{sp}(H) \) for given matrices \( A, H \). However, in this paper we will concentrate on the structural restriction

\[
\Pi \in S^p_r = \{ M \in \mathbb{R}^{p \times p} | M = M', \text{rank}(M) \leq r \},
\]

i.e., \( \Pi \) must lie in \( S^p_r \), the subset of symmetric \( p \times p \) matrices of rank smaller than or equal to \( r \). We write \( S_p \) to denote the set of symmetric \( p \times p \) matrices with no restrictions on the rank. The symmetry condition leads to a nontrivial problem of maximizing the likelihood of (2.2), under the two conditions \( \text{rank}(\Pi) \leq r \) and \( \Pi = \Pi' \).

Consider first the log-likelihood function for (2.2) for a general (nonsymmetric, possibly full rank \( r \leq p \)) \( \Pi \), where we omit constant terms

\[
\log L(\Pi, \Omega) = -\frac{N}{2} \log |\Omega| - \frac{1}{2} \sum_{n=1}^{N} (\Delta y_n - \Pi y_{n-1})' \Omega^{-1} (\Delta y_n - \Pi y_{n-1}).
\]

Defining the \( p \times p \) sufficient statistics matrices \( S_{ij}, i, j = 0, 1 \),

\[
S_{00} = \frac{1}{N} \sum_{n=1}^{N} \Delta y_n (\Delta y_n)', \quad S_{01} = \frac{1}{N} \sum_{n=1}^{N} \Delta y_n y_{n-1}',
\]

\[
S_{10} = \frac{1}{N} \sum_{n=1}^{N} y_{n-1} (\Delta y_n)', \quad S_{11} = \frac{1}{N} \sum_{n=1}^{N} y_{n-1} y_{n-1}',
\]

the log-likelihood (4.2) can be rewritten as

\[
\log L(\Pi, \Omega) = -\frac{N}{2} \log |\Omega| - \frac{1}{2} \text{Tr} \left\{ \sum_{n=1}^{N} (\Delta y_n - \Pi y_{n-1}) (\Delta y_n - \Pi y_{n-1})' \Omega^{-1} \right\}
\]

\[
= -\frac{N}{2} \log |\Omega| - \frac{N}{2} \text{Tr} \left\{ (S_{00} - \Pi S_{10} - S_{01} \Pi' + \Pi S_{11} \Pi') \Omega^{-1} \right\}.
\]
where $\text{Tr}(\cdot)$ denotes the trace operator. The differential in the $\Omega$ direction of this expression is

$$d \log L(d\Omega) = -\frac{N}{2} \text{Tr}(\Omega^{-1}d\Omega) + \frac{N}{2} \text{Tr}(\Omega^{-1}(S_{00} - \Pi S_{10} - S_{01} \Pi' + \Pi S_{11} \Pi') \Omega^{-1}d\Omega)$$

$$= -\frac{N}{2} \text{Tr}\left(\Omega^{-1}(I_p - (S_{00} - \Pi S_{10} - S_{01} \Pi' + \Pi S_{11} \Pi') \Omega^{-1})d\Omega\right),$$

which is zero for all $d\Omega$ with

$$(4.4) \quad \Omega_{\Pi} = \Pi_{\Pi} = S_{00} - \Pi S_{10} - S_{01} \Pi' + \Pi S_{11} \Pi'.$$

Thus, for an arbitrary fixed $\Pi$, the covariance estimator is given by $(4.4)$, which is explicitly dependent on $\Pi$; see also [7]. Inserting $(4.4)$ into the log-likelihood $(4.2)$ yields the profile log-likelihood for $\Pi$. Up to a constant it is

$$\log L(\Pi) = -\frac{N}{2} \log |\Omega_{\Pi}|$$

$$(4.5) \quad = -\frac{N}{2} \log |S_{00} - \Pi S_{10} - S_{01} \Pi' + \Pi S_{11} \Pi'|,$$

implying that the unrestricted maximum likelihood estimator of $(4.5)$ is the ordinary least squares (OLS) estimator

$$(4.6) \quad \hat{\Pi}_{\text{ols}} = S_{01} S_{11}^{-1},$$

which is not symmetric.

Now define $\Pi = \beta \delta \beta'$, with $\beta$ a $p \times r$ matrix of rank $r$, and $\delta$ a (possibly nonsymmetric) $r \times r$ matrix of full rank $r$. The differential of $(4.5)$ in the $\delta$ direction is then

$$d \log L(d\delta) = -\frac{N}{2} \text{Tr}\left(\Omega^{-1}(S_{01} \beta \delta \beta' S_{10} - \beta \delta \beta' S_{11} \beta \delta \beta' S_{01} - \beta \delta \beta' S_{11} \beta \delta \beta' S_{10})\right),$$

which is zero for all $d\delta$ if

$$(4.7) \quad \text{Tr}\left(\Omega^{-1}(S_{01} \beta \delta \beta' S_{10} - \beta \delta \beta' S_{11} \beta \delta \beta' S_{01})\right) = \text{Tr}\left(\Omega^{-1}(\beta \delta \beta' S_{11} \beta \delta \beta' + \beta \delta \beta' S_{11} \beta \delta \beta')\right).$$

So far we have not imposed any symmetry conditions on $\delta$. If we assume that $\delta = \delta'$ and $d\delta = d\delta'$, we obtain the identities

$$\text{Tr}\left(\Omega^{-1}(S_{01} \beta \delta \beta' S_{10})\right) = \text{Tr}(\Omega^{-1} S_{01} \beta \delta \beta' S_{10}) + \text{Tr}(\Omega^{-1} \beta \delta \beta' S_{10} S_{01})$$

$$= \text{Tr}(\beta' \Omega^{-1} S_{10} \beta \delta) + \text{Tr}(\beta' S_{10} \Omega^{-1} \beta \delta)$$

$$= \text{Tr}(\beta' (S_{10} \Omega^{-1} + \Omega^{-1} S_{01}) \beta \delta),$$

and

$$\text{Tr}\left(\Omega^{-1}(\beta \delta \beta' S_{11} \beta \delta \beta' + \beta \delta \beta' S_{11} \beta \delta \beta')\right) = \text{Tr}(\Omega^{-1} \beta \delta \beta' S_{11} \beta \delta \beta' + \Omega^{-1} \beta \delta \beta' S_{11} \beta \delta \beta')$$

$$= 2\text{Tr}(\beta' S_{11} \beta \delta \beta' \Omega^{-1} \beta \delta).$$
Thus, (4.7) becomes
\[ \text{Tr}(\beta'(S_{10} \Omega^{-1} + \Omega^{-1} S_{01}) \beta d\delta) = 2 \text{Tr}(\beta' S_{11} \beta \delta \Theta^{-1} \beta d\delta), \]
which holds for
\[ \delta = \frac{1}{2} (\beta' S_{11} \beta^{-1} \beta'(S_{10} \Omega^{-1}_{\beta,\beta} + \Omega^{-1}_{\beta,\beta} S_{01}) \beta (\beta' \Theta^{-1}_{\beta,\beta} \beta)^{-1}) \]
\[ = \frac{1}{2} (\beta' \Theta^{-1}_{\beta,\beta} \beta^{-1} \beta'(S_{10} \Omega^{-1}_{\beta,\beta} + \Omega^{-1}_{\beta,\beta} S_{01}) \beta (\beta' S_{11} \beta)^{-1} = \delta', \]
where the dependencies of \( \delta \) on \( \beta \), and of \( \Theta \) on \( \beta \) and \( \delta \), are explicitly stated by subindices.

Since \( S_{10} = S_{01} \), (4.8) agrees with the result that the nearest symmetric matrix to any arbitrary quadratic matrix \( M \) is the Hermitian part of \( M, M = \frac{1}{2}(M + M') \), in the sense that \( ||M - \hat{M}||_F \leq ||M - \hat{M}||_F \) for any symmetric \( \hat{M} \), with equality if and only if \( \hat{M} = \hat{M} \), where \( || \cdot ||_F \) denotes the Frobenius norm; see [5].

Consider now
\[ d \log L(d\beta) = \text{Tr}\{\Omega^{-1}(-d \beta \delta \Theta S_{10} - \beta \delta \delta \Theta S_{10} - S_{01} d \beta \delta \delta' - S_{01} \beta \delta d\delta') \]
\[ + d \beta \delta \Theta S_{11} \beta \delta \beta' + \beta \delta \delta' S_{11} \beta \delta \beta' + \beta \delta \beta' S_{11} \beta \delta \delta' + \beta \delta \delta' S_{11} \beta \delta \delta' \}. \]
If \( \beta = 0 \) (implying that \( y_n \) is a \( p \)-dimensional Brownian motion and rank \( r = 0 \)) or the OLS estimator is symmetric, and letting \( \beta \delta \delta' = S_{01} S_{11}^{-1} = S_{11}^{-1} S_{01} \), the score is 0. We postulate that the set of symmetric OLS estimators is a nullset, hence the probability of this is 0. These observations imply that deriving a symmetric (low nonzero rank) estimator of \( \Pi \) through the matrices \( \beta, \delta \) directly from the likelihood function is not feasible. It follows from the fact that the subset of symmetric \( p \times p \) matrices in the set of \( p \times p \) matrices is a null-set with respect to the Lebesgue measure on \( \mathbb{R}^{p \times p} \).

A possibility for estimating \( \Pi \) in (4.5) under the restriction to symmetry is to compute the Hermitian form of the OLS estimator (4.6), which implies that
\[ \tilde{\Pi} = \frac{1}{2} (S_{01} S_{11}^{-1} + S_{11}^{-1} S_{01}) = \frac{1}{2} (\Pi_{\scrS} + \Pi'_{\scrS}), \]
corresponding to the projection of \( \Pi_{\scrS} \) onto \( S_p \). From this, we claim that (4.9) is the optimal choice of an otherwise unrestricted estimator of a symmetric \( \Pi \), since \( \Pi_{\scrS} \) is the optimal unrestricted estimator for \( \Pi \) and (4.9) is the optimal symmetric approximation of this estimator.

The Eckart–Young–Mirsky theorem states that the optimal approximation to a matrix by a lower rank matrix is given by the singular value decomposition (SVD) [4]. Let \( M = U \Lambda V' \) denote the SVD of \( M \) and assume that the entries of the diagonal \( p \times p \) matrix \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) are ordered such that \( \lambda_1 > \lambda_2 > \ldots > \lambda_p \). Partition \( U, V \) as \( U = (U_r, U_{p-r}) \) and \( V = (V_r, V_{p-r}) \) such that \( U_r, V_r \in \mathbb{R}^{p \times r} \) corresponds to the \( r \) largest entries of \( \Lambda \) and \( U_{p-r}, V_{p-r} \in \mathbb{R}^{p \times (p-r)} \) corresponds to the last \( p - r \) entries of \( \Lambda \). Furthermore, let \( \Lambda_r = \text{diag}(\lambda_1, \ldots, \lambda_r) \). Then the optimal low rank approximation of \( M \) is given by
\[ \tilde{M} = U_r \Lambda_r V_r'. \]
The SVD approximation of a low rank estimate of $\Pi$ is
\begin{equation}
\tilde{\Pi}^r_{\text{sym}} = U_r(\tilde{\Pi})\Lambda_r(\tilde{\Pi})V'_r(\tilde{\Pi}),
\end{equation}
where $U_r(\cdot), \Lambda_r(\cdot), V'_r(\cdot)$ refer to the SVD matrices of rank $r$ with respect to the argument.

The log-likelihood is then given by (4.5), with
$$
\Omega(\tilde{\Pi}^r_{\text{sym}}) = S_{00} - \tilde{\Pi}^r_{\text{sym}}S_{10} - S_{01}\tilde{\Pi}^r_{\text{sym}} + \tilde{\Pi}^r_{\text{sym}}S_{11}\tilde{\Pi}^r_{\text{sym}}.
$$

For the general case of a low rank approximation of a $p \times p$ matrix $\Pi$ under some additional restrictions, the problem can be stated as
\begin{equation}
\min_{M} \|\Pi - M\|_F, \text{ subject to } M \in \mathcal{S}_r^p \subset \mathbb{R}^{p \times p},
\end{equation}
where $M$ is a generic matrix restricted to the subspace $\mathcal{S}_r^p$ given in (4.1).

Under a strict rank condition, rank$(M) = r$, (4.12) may not have an optimal solution in the sense that a lower rank matrix might result in a lower Frobenius norm distance [2]. However, the relaxed rank constraint rank$(M) \leq r$ in $\mathcal{S}_r^p$ ensures that the Frobenius norm is minimized, given that the selected subspace is nonempty.

For a general nonempty set, $\mathcal{S}$, a low rank approximation restricted to $\mathcal{S}$ can be solved by numerical methods such as Algorithm 4.1, which is adapted from the Lift-and-Project algorithm in [2] with the order of projection and lifting switched. This is because a low rank approximation using (4.10) of a symmetric matrix is itself symmetric. Hence, in the case where $\mathcal{S} = \mathcal{S}'$ the algorithm requires only 1 iteration to complete. Note that the symmetrized low rank approximation presented above corresponds to this with target $\tilde{\Pi}_{\text{ols}}$. To solve (4.12) in the subspace (4.1), the algorithm can be tried for all allowed ranks, $r_1, \ldots, r_n \leq r$, and the output is the matrix $M_j^{(i)}, j = 1, \ldots, n$, with the lowest Frobenius norm difference compared to the target matrix.

In section 6 we compare the following estimators of $\Pi$:
(i) The unrestricted full rank OLS estimator $\tilde{\Pi}_{\text{ols}} = S_{01}S_{11}^{-1}$.
(ii) The reduced rank (nonsymmetric) Johansen estimator $\tilde{\Pi}_J = \hat{\alpha}\hat{\beta}'$.
(iii) The symmetric Johansen estimator $\tilde{\Pi}_{\text{proj}}$ being the rank $r < p$ approximation of $\frac{1}{2}(\tilde{\Pi}_J + \tilde{\Pi}'_J)$.
(iv) The symmetric low rank OLS estimator $\tilde{\Pi}_{\text{sym}}$ being the rank $r < p$ approximation of $\frac{1}{2}(\tilde{\Pi}_{\text{ols}} + \tilde{\Pi}'_{\text{ols}})$.

Note here that only the estimators (ii)–(iv) have low rank, and only (iii) and (iv) are symmetric.

Algorithm 4.1 Project and Lift.

Initialize $M^{(0)}$ with some target $\tilde{\Pi}$ of arbitrary rank, not necessarily in $\mathcal{S}$. Then for $i = 1, 2, \ldots$ repeat until convergence between $M^i$ and $M^{i-1}$:
1. Project: find the projection $\tilde{M}^{(i)}$ of $M^{(i-1)}$ onto $\mathcal{S}$
2. Lift: compute the low rank approximation $M^{(i)}$ of $\tilde{M}^{(i)}$, such that rank$(M) = r$.  

© 2024 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license
5. **High-dimensional estimation.** There are two crucial steps in obtaining estimates in model (2.2), as the rank of the system as well as \( \Pi \) must be determined. When the dimension, \( p \), of the model (2.2) increases, estimating \( \hat{r} \) close to the true rank becomes progressively difficult as the rank test is prone to underestimating the rank. This is due to the construction of the test sequence, which starts with \( H_r \) against \( H_p \) for \( r = 0 \). If this is rejected, then \( r \) increases by 1 and this continues until \( H_r \) against \( H_p \) cannot be rejected or if \( r + 1 = p \), such that \( \hat{r} = p \). For some true rank \( r \leq p \), this sequence is bound to reach a conclusion. However, since \( H_r \) is evaluated in some distribution then it will eventually fall within a confidence region of this distribution, thus not rejecting \( H_r \). This typically occurs prior to testing at the true rank \( r \), as shown in Figure 1 for \( N = 2000 \) observations of a linear Kuramoto-type system as described above (see parameter settings in section 6). Here the distribution of each test statistic was found by bootstrapping [1].

The figure demonstrates the 95\% confidence bounds (gray) for the distribution of the LRT under \( H_r \), \( r = 78, \ldots, 88 \), using 300 bootstrap samples, when the true rank of the linearized Kuramoto system is 84 (see section 6 for full model specification). The LRT for the observed data (red) approaches the 95\% confidence area from above and is rejected up to, and including, \( H_r : r = 81 \). At \( r = 81 \), the test falls within the 95\% region, and therefore it is not rejected. The conclusion is therefore \( \hat{r} = 81 \) (blue dotted line), when in fact the true rank is \( r = 84 \) (gray dotted line).

![Figure 1](image-url)  
*Figure 1. Example of rank estimation in a 100-dimensional linearized Kuramoto system, without any symmetry restrictions. Bootstrapped likelihood ratio test values of \( H_r \) versus \( H_p \) for \( N = 2000 \). 95\% confidence bounds for the bootstrap test (gray) and the test values from the data (red). The test values enters the 95\% region at \( \hat{r} = 81 \) (blue dotted line), whereas the true value is \( r = 84 \) (gray dotted line). The bootstrap test is based on 300 samples. Increasing to 1000 does not change the rank estimate of \( \hat{r} = 81 \). Increasing the number of observations \( N \) will lead to a more pronounced kink in the red curve at the true rank \( r = 84 \), due to the curve being steeper for values < 84. This implies a more precise estimation of the true rank.*
Inspecting Figure 1 closely, the red curve displays a kink at \( r = 84 \), indicating that at the true value there is a notable change in the LRT trend. For a lower number of observations this is less visible, whereas for a higher number of observations, the kink is more pronounced. The numerical magnitude of the derivative of the LRT curve, up to the true value of \( r \), increases with the number of observations, and thus the test is more likely to find the true value of \( r \), as the trend of the red curve will cut into the gray region at a steeper angle. Indeed, using \( N = 500 \) observations leads to a rank estimate of 0 (indicating a random walk), \( N = 1000 \) observations leads to a rank estimate of 10, and \( N = 5000 \) observations estimates the rank correctly at 84. Thus, for a 100-dimensional linear Kuramoto-type system, a large sample size is needed to infer the rank in an unrestricted model and thus the cointegration relations.

For the data used to produce Figure 1, the conclusion was the same when using 1000 bootstrap samples, confirming that the empirical distribution for the given number of observations is already well determined with 300 bootstrap samples.

It is of interest to assess how well the estimates of (2.2) model the underlying structure of the target \( \Pi \) when the rank is underestimated. Intuitively we are approximating the true subspace of dimension \( r \) with a subspace of dimension rank(\( \hat{\Pi} \)) = \( \hat{r} \). For rank(\( \Pi_1 \)) = \( r_1 \leq \) rank(\( \Pi_2 \)) = \( r_2 \), we have that sp(\( \Pi_1 \)) \subseteq sp(\( \Pi_2 \)) if the \( r_1 \) basis vectors for \( \Pi_1 \) equals the \( r_1 \) numerically largest eigenvalues for \( \Pi_2 \). Thus, \( \Pi_1 \)'s basis vectors are the \( r_1 \) most important basis vectors of the \( \Pi_2 \) space. Therefore, underestimating the rank of \( \Pi \) implies that the basis vectors of the estimation space will be linear combinations of the (most important) basis vectors of the true \( \Pi \)-space.

The quality of the estimation depends on how close the estimation rank is to the true rank. This can be visualized by defining a measure of “closeness” for two subspaces as defined by the matrices \( \Pi \) (true subspace) and \( \hat{\Pi} \) (estimated subspace). For this purpose, one can use a generalized version of the vector angle, using the Frobenius inner product \( \langle U,V \rangle_F = \text{Tr}(U'V) \) for matrices \( U, V \in \mathbb{R}^{p \times p} \). The angle between the matrices \( U \) and \( V \) is defined as

\[
\Theta(U,V) = \arccos \left( \frac{\langle U,V \rangle_F}{\sqrt{\langle U,U \rangle_F \langle V,V \rangle_F}} \right)
\]

(5.1)

For \( U = 0 \) and/or \( V = 0 \) we define \( \Theta(U,V) = \pi/2 \), i.e., the 0 matrix is always orthogonal to itself and any other.

Figure 2 displays the matrix angles between the Johansen estimator (no symmetry restriction) \( \hat{\Pi}_j \) for \( \hat{r} = 70, \ldots, 100 \) against the (unrestricted) full rank OLS estimator \( \hat{\Pi}_{ols} \) for a varying number of observations \( N = 500, 1000, 2000, 5000 \) of the same Kuramoto system simulated in Figure 1 (for parameter settings, see section 6). Shown are median angles, based on 500 replications for each combination of observations \( N \) and rank \( r \). As a reference, the angles between the low rank approximations \( \Pi' \) of the true simulation matrix \( \Pi^{84} \) are presented as well. The vertical dotted line marks rank(\( \Pi \)) = 84.

The reference angles show that as the rank increases, the low rank approximations become better, but when reaching the true rank of \( \Pi \), then \( \Theta(\Pi,\Pi_r) = 0 \) for \( r \geq 84 \). In contrast, the angles between the Johansen estimator and the OLS estimator decrease past the true rank of \( \Pi \) due to the fact that rank(\( \hat{\Pi}_{ols} \)) = 100. However, as the number of observations increases, the kink noticeable for the reference angles \( \Theta(\Pi^{84},\Pi') \) at \( \hat{r} = 84 \) starts to appear, suggesting...
that as the number of observations increases, the OLS estimator is well approximated at the rank of the true simulation matrix $\Pi$.

Generally, as $\hat{r}$ increases, the matrix angles decrease from orthogonality ($\pi/2$) to a minimum. Underestimating the rank will influence the estimation negatively, but the severity is dependent on how far from the true rank $\hat{r}$ is. Overestimation ensures a lower matrix angle, but impacts estimation negatively due to an excessive number of parameters to be estimated. From Figure 2 it is evident that estimating in the vicinity of the true rank is optimal when fitting the model; however, it is not of crucial importance to find the exact true rank, and we can obtain fair estimates of $\Pi$ when the rank is underestimated by a relatively small amount.

6. **Simulation.** Consider a system with $\Pi$ as in (3.2)–(3.3) and $p = 100$ with twelve 8-clusters and four 1-clusters, i.e., a system of 12 coupled clusters each of size 8 and 4 independent processes. The order of the simulated processes is scrambled before any estimation takes place, such that adjacent processes of the observations are not necessarily part of the same cluster. For presentation purposes, the clusters are reordered correctly when visually displaying results. We let each cluster be coupled with varying strengths $c_i = a_i b_i \in [0.5; 2], i = 1, \ldots, 12,$ and assume that $\Omega = I_p$. For the 4 independent processes, the coupling strengths are implicitly set to 0; cf. section 3. The chosen diagonal covariance matrix implies that if two processes do not cointegrate, then they are independent. However, this is not generally true. If the covariance
is nonzero between two processes that do not cointegrate, they still interact through the covariance. A simulation of $N = 2000$ observations with initial condition $y_0 = 0$ is presented in Figure 3, where the coloring is for easy identification of the 8-clusters (blue), the 1-clusters (red), and a pure 100-dimensional random walk for reference.

It is noticeable how each cluster, as a whole, behaves like a random walk. The individual processes within each cluster overlap, appearing as a few blue processes represented by thick lines, visible in Figure 3. The coupling strength is so strong that the internal cluster processes never stray far from the cluster, and as a whole they behave as a stationary process. The random walk reference shows how tightly knit the cointegrated processes are.

**6.1. Estimation of $\Pi$.** The rank estimation of $\Pi$ was performed with bootstrapping. From 300 bootstrap samples, the rank was determined as $\hat{r} = 81$; see Figure 1. Given this rank estimate, we compare 4 estimators for $\Pi$: (i) the unrestricted OLS estimator $\hat{\Pi}_{\text{ols}}$, (ii) the (unrestricted) Johansen estimator $\hat{\Pi}_{\text{J}}$, (iii) the symmetrized Johansen estimator $\hat{\Pi}_{\text{proj}}$, and (iv) the symmetrized low rank OLS estimator $\hat{\Pi}_{\text{sym}}$.

Figure 4 presents the true simulation matrix $\Pi$ with the unrestricted (full rank) OLS estimator $\hat{\Pi}_{\text{ols}}$, correctly ordered. We detail the ordering in section 6.3 below. The decreasing coupling strength of the Kuramoto clusters is evident in both matrices, indicated by the fading colors across the block diagonal. Note the lower right corner, where the 4 1-clusters are represented as 0 entries in $\Pi$. The OLS is capable of reproducing the structure of the true $\Pi$ to some extent.

Figure 5 shows the 3 types of low rank estimator considered, estimators (ii)–(iv) under the inferred rank $\hat{r} = 81$, as well as extreme under/overestimated ranks $\hat{r} = 71, 91$. Evidently,
6.2. Comparing estimators under structural restrictions. We compare the four estimators with respect to two measures: (1) the matrix angle between the estimator and the true matrix $\Pi$, and (2) the standard deviation of the entries of the estimator that lies outside the true block diagonal structure, i.e., the indices that are 0 in the true matrix $\Pi$. The matrix angle measures how well the estimator captures the true structure of the simulation matrix $\Pi$, whereas the standard deviation measure describes the amount of noise in the estimation of entries that should optimally be 0. The individual estimates are also compared to the measures for $\Pi_{\text{ols}}$, which is nonsymmetric and of full rank.

Table 1 presents the results under the inferred rank $\hat{r} = 81$ as well as the underestimated ranks $\hat{r} = 71, 91$. It is clear that the symmetric estimators $\hat{\Pi}_{\text{proj}}$ and $\hat{\Pi}_{\text{sym}}$ are superior in capturing the dynamics as well as lowering the noise in (true) zero entries. Comparing the angles for $\hat{\Pi}_{\text{proj}}$ and $\hat{\Pi}_{\text{sym}}$ across ranks, we see that the decrease from $\hat{r} = 71$ to $\hat{r} = 81$ is much larger than from $\hat{r} = 81$ to $\hat{r} = 91$, suggesting a lower marginal gain. The standard deviation measure does not improve with increasing rank for the symmetric estimators.

Comparing LRT values at $\hat{r} = 81$ for the symmetric estimators, the $\hat{\Pi}_{\text{sym}}$ has a smaller LRT value of 27769 against a value of 28410 for the $\hat{\Pi}_{\text{proj}}$ estimator. Indeed simulating this difference 1000 times, we find that this is generally the case. Figure 6 shows a histogram of the difference of the two LRT values $-2\log Q(\hat{\Pi}_{\text{proj}}, \hat{\Pi}) - (-2\log Q(\hat{\Pi}_{\text{sym}}, \hat{\Pi}))$. Ranking the two similar symmetric low rank estimators, the $\hat{\Pi}_{\text{sym}}$ comes out on top with slightly better performance than $\hat{\Pi}_{\text{proj}}$ based on the measures considered here, and it comes with a higher
Figure 5. Estimating $\Pi$ with the Johansen estimator $\hat{\Pi}_J$, the symmetrized Johansen estimator $\hat{\Pi}_{\text{proj}}$, and the symmetrized low rank OLS estimator $\hat{\Pi}_{\text{sym}}$, under various low rank approximations for $\hat{r} = 71, 81, 91$, when the estimated rank is 81. The middle column corresponds to the estimates under the inferred rank, the left column corresponds to estimates under an (extreme) low rank, and the right column corresponds to estimates under an (extreme) high rank. For $\hat{r} = 71$, estimates are more noisy, but for $\hat{r} = 81, 91$, visual comparisons are similar across rows. Also, $\hat{\Pi}_{\text{proj}}$ and $\hat{\Pi}_{\text{sym}}$ look very similar. All matrices are ordered correctly for easy comparison.

likelihood value. This is of course a result of approximating the unrestricted OLS estimator with a symmetric low rank version as opposed to approximating the Johansen estimator, which is already under a rank condition when being estimated. On the other hand, it is the
Table 1

Measures of estimators of $\Pi$. Matrix angles $\Theta(\hat{\Pi}, \Pi)$ measure how the estimator captures the structure of $\Pi$, and the standard deviation measures the amount of noise in the estimator.

<table>
<thead>
<tr>
<th>Matrix angles $\Theta(\hat{\Pi}, \Pi)$</th>
<th>std. dev. of 0 entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\Pi}_{\text{ols}}$</td>
<td>0.8492</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{J}}$</td>
<td>0.9018</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{proj}}$</td>
<td>0.8249</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{sym}}$</td>
<td>0.8230</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{J}}$</td>
<td>0.8678</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{proj}}$</td>
<td>0.7899</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{sym}}$</td>
<td>0.7891</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{J}}$</td>
<td>0.8514</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{proj}}$</td>
<td>0.7861</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{sym}}$</td>
<td>0.7847</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{J}}$</td>
<td>0.0184</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{proj}}$</td>
<td>0.0149</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{sym}}$</td>
<td>0.0148</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{J}}$</td>
<td>0.0180</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{proj}}$</td>
<td>0.0142</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{sym}}$</td>
<td>0.0142</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{J}}$</td>
<td>0.0176</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{proj}}$</td>
<td>0.0142</td>
</tr>
<tr>
<td>$\hat{\Pi}_{\text{sym}}$</td>
<td>0.0142</td>
</tr>
</tbody>
</table>

Figure 6. Comparing the LRT values of the $\hat{\Pi}_{\text{proj}}$ and the $\hat{\Pi}_{\text{sym}}$ estimators: $-2\log Q(\hat{\Pi}_{\text{proj}}, \hat{\Pi}) - (-2\log Q(\hat{\Pi}_{\text{sym}}, \hat{\Pi}))$, based on 1000 replications. All differences are positive, implying that the LRT values of $\hat{\Pi}_{\text{sym}}$ are lower in all simulations, further implying that this estimator is the superior of the two.

Johansen estimator on which the rank test is based, or in the case where $r = p$ the OLS estimator, as they coincide. Hence, both are necessary for the inference considered here, but in conclusion the approximation of the OLS is superior to determine the underlying structure of the high-dimensional process.

6.3. Recovering clusters from a random ordering. Estimation of $\Pi$ does not rely on the individual processes being ordered according to the cluster structure used in the simulation. However, the observations cannot be assumed to follow this ordering, and we demonstrate how to recover the structure from an estimate of $\Pi$ using the CNM algorithm from [3]. The algorithm is a graphical approach to group nodes in a graph with weighted edges by maximizing a modularity score measuring the community structure in a network, where a measure of
Table 2

Processes in each cluster ordered by coupling strength compared to estimated clusters. Left: the true coupling strengths and constituents of each of the 16 clusters. Right: estimated constituents, using the CNM algorithm. Processes on one side but not the other appear in bold. Evidently as the coupling strength decreases more classification errors occur. Furthermore all four independent processes are misplaced in other clusters. The CNM algorithm uses a modularity score to determine the number of clusters.

<table>
<thead>
<tr>
<th>Coupling</th>
<th>True constituents</th>
<th>Estimated constituents</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00</td>
<td>{15, 33, 54, 60, 62, 90, 93, 94}</td>
<td>{15, 33, 54, 60, 62, 77, 90, 93, 94}</td>
</tr>
<tr>
<td>1.86</td>
<td>{9, 34, 42, 53, 73, 75, 88, 92}</td>
<td>{9, 34, 42, 53, 73, 75, 88, 92}</td>
</tr>
<tr>
<td>1.73</td>
<td>{4, 23, 24, 30, 50, 56, 76, 100}</td>
<td>{4, 23, 24, 30, 50, 56, 76, 100}</td>
</tr>
<tr>
<td>1.59</td>
<td>{12, 13, 46, 48, 57, 68, 72, 80}</td>
<td>{12, 13, 46, 48, 57, 68, 72, 80}</td>
</tr>
<tr>
<td>1.45</td>
<td>{3, 19, 22, 40, 63, 65, 74, 82}</td>
<td>{3, 19, 22, 40, 63, 65, 74, 82}</td>
</tr>
<tr>
<td>1.32</td>
<td>{1, 7, 11, 25, 41, 69, 71, 79}</td>
<td>{1, 7, 11, 25, 41, 69, 71, 79}</td>
</tr>
<tr>
<td>1.18</td>
<td>{5, 21, 39, 43, 47, 49, 59, 95}</td>
<td>{5, 21, 39, 43, 47, 49, 59, 95}</td>
</tr>
<tr>
<td>1.05</td>
<td>{8, 28, 31, 36, 58, 61, 89, 97}</td>
<td>{8, 28, 31, 36, 58, 61, 89, 97}</td>
</tr>
<tr>
<td>0.91</td>
<td>{20, 27, 45, 51, 55, 78, 81, 98}</td>
<td>{14, 17, 20, 27, 51, 55, 78, 81, 85}</td>
</tr>
<tr>
<td>0.77</td>
<td>{6, 26, 37, 38, 52, 64, 83, 99}</td>
<td>{2, 6, 26, 37, 38, 52, 64, 83, 84, 99}</td>
</tr>
<tr>
<td>0.64</td>
<td>{10, 14, 16, 32, 35, 44, 85, 96}</td>
<td>{10, 32, 44}</td>
</tr>
<tr>
<td>0.50</td>
<td>{18, 29, 66, 67, 70, 86, 87, 91}</td>
<td>{16, 18, 29, 35, 45, 66, 67, 70, 86, 87, 91, 96, 98}</td>
</tr>
<tr>
<td>0</td>
<td>{2}</td>
<td>{}</td>
</tr>
<tr>
<td>0</td>
<td>{17}</td>
<td>{}</td>
</tr>
<tr>
<td>0</td>
<td>{77}</td>
<td>{}</td>
</tr>
<tr>
<td>0</td>
<td>{84}</td>
<td>{}</td>
</tr>
</tbody>
</table>

0 corresponds to complete randomness in the network structure. Using the modularity score, the CNM algorithm determines the number of clusters intrinsically, thereby only requiring a matrix of weighted edge connections as input. Here we used \(\text{abs}(\Pi_{\text{sym}})\) as weights, where \(\text{abs}(\cdot)\) denotes the entrywise absolute value, since it is the magnitude of entries that determines the connection strength, the sign being irrelevant.

Table 2 shows the output from the algorithm against the true cluster structure ordered by decreasing coupling strength. There are 12 8-clusters of varying coupling strengths \(c_i, i = 1, \ldots, 12\), and 4 1-clusters in the true structure (the coupling strength is zero for the last four processes). The CNM algorithm identifies the 8-clusters closely for \(c_i > 1\), missing only one process (77), which is one of the four independent 1-clusters. When the coupling strength decreases to \(c_i < 1\), more classification errors appear. In particular, the two clusters with the lowest \(c_i\)’s are mixed up and all independent 1-clusters are misplaced in larger clusters. Figure 7 (middle) displays the \(\Pi\) matrix ordered by the cluster structure identified by the CNM algorithm compared with the true matrix (left side). The misplaced independent 1-cluster in the strongest coupled cluster is visible in the upper left corner. The maximum modularity is estimated at 0.357 by the CNM algorithm, which is indicative of significant community structure according to [3].

Figure 7 (right) shows an estimate of the Kuramoto structure; see section 3. Assuming independence among clusters and using the CNM clusters as input, we estimate the individual cluster’s internal coupling structure separately using the \(\Pi_{\text{sym}}\) estimator for each of the partitions of the full system. This results in a structure similar to that of the middle of Figure 7, but without the noise off the block diagonal, as these entries are set to zero by construc-
7. Discussion. This paper discussed the estimation of a cointegrated system in a high-dimensional setting such as a symmetric linearized Kuramoto system. While previous work has derived asymptotic properties of high-dimensional cointegrated systems (see [11]), we opted instead for bootstrapping for rank estimation, as described in [1]. Bootstrapping is simple to implement and performs well, as demonstrated in a simulation example.

The claim that the sequential testing of the rank is prone to underestimating of $r$ was supported by a visual argument on how the likelihood ratio test behaves when associated with a distribution to evaluate the test statistic. However, it was also demonstrated by examples that this underestimation is not a major cause for concern when the dimension of the system is high and the rank is underestimated to a lesser extent, contrary to low-dimensional systems where determining the precise rank is an essential task. This is also supported by the findings of [6] where the exact distributions of the estimators of the cointegration matrix $\Pi$ under misspecification of the rank are found. Underestimation of the rank introduces a bias; however, it is usually small and compensated for by a smaller variance. Overestimation of the rank does not introduce bias but it inflates the variance. The underestimation of the rank was also empirically found in [9], where lower sample size led to lower rank estimation.

The underestimation of the rank by sequential testing seems contradictory to the findings of [11], which showed that the rank is asymptotically overestimated, leading to wrongly inferred cointegration relations. However, the asymptotic regime investigated in [11] is different from ours. They consider a high-dimensional setting where both $p$ and $N$ go to infinity, keeping the proportion $p/N$ constant. We do not consider an asymptotic regime, but rather finite sample size and large but fixed $p$, say $p \gg 10$.

Determining the rank of a cointegrated system is important in order to find the correct number of stochastic trends that drive the system; however, when $p \to \infty$, the number of
sequential tests can become a problem. Therefore, it might be of value to investigate alternatives in order to minimize the error of selecting a rank very far from the true value; see [10] for a review of penalization methods in cointegration models, including evaluation and comparisons between methods for different error measures.

The task of estimating the matrix $\Pi$ under the nonstandard restriction to symmetry, as well as the usual low rank condition, required a two-step procedure that separated the tasks. We presented an estimator, based on the standard OLS estimator which was sequentially approximated by a symmetric and low rank matrix. This estimator proved superior to the standard low rank Johansen estimator as well as a symmetrized version of this, measured by the matrix angle to the true matrix and the difference in the likelihood ratio test statistic between the two symmetrized estimators. Based on the symmetric low rank estimator, we used the graphical CNM algorithm to recover the underlying cluster structure of the simulated linear Kuramoto system. This proved to be effective for stronger coupling strengths, recovering most of the clusters. However, a few of the weaker coupled clusters were mixed up. The few independent random walks added to the Kuramoto system were all placed in some of the CNM estimated clusters, although all of these should have been placed in separate clusters of the single processes.

The CNM algorithm was selected based on its simplicity and relatively strong performance in recovering the Kuramoto clusters. However, other (graphical) algorithms might provide equally good or better results. The CNM estimates also the number of clusters, hence an algorithm assuming a fixed number might be better suited. From the estimated rank $\hat{r}$ of a $p \times p$ matrix $\Pi$, the number of clusters in a linearized Kuramoto system is $p - \hat{r}$. Thus, when the rank is underestimated, the number of clusters is overestimated. This knowledge could be taken into account as well, such as the maximum number of clusters to search for.

8. Conclusion. We estimated the rank $r$ of a high-dimensional cointegrated system by bootstrapping a distribution for the likelihood ratio test. Although the rank is underestimated, we showed that this is not a major concern in a high-dimensional setting when the underestimation is relatively small. We also derived an estimator that restricts the matrix of the system to be both symmetric and of low rank. This estimator is based on the usual OLS estimator, using a two-step procedure to impose the conditions of symmetry and low rank. We also defined and analyzed a simulation of a 100-dimensional linearized Kuramoto system and used a graphical approach to recover the underlying cluster structure. By combining rank estimation with our new low rank symmetric estimator and the graphical algorithm, we were able to recover most of the structure of the full system, with some of the weaker clusters being mixed up.

REFERENCES

