High-dimensional latent Gaussian count time series

Marie Düker (FAU Erlangen)

with R. Lund (UC-Santa Cruz) and V. Pipiras (UNC-Chapel Hill)

University of Athens, December 1st, 2023

Outline

1. Motivation for (vector-valued) count time series models
2. Model
3. Concentration inequalities for autocovariance matrix estimates
4. Sparse estimation for latent VAR processes
5. Conclusions

Examples of correlated counts observed at different time points:

- Number of patients with different but related symptoms

Examples of correlated counts observed at different time points:

- Number of patients with different but related symptoms
- Occurrences of physical phenomena at different locations

Motivation Count time series

Examples of correlated counts observed at different time points:

- Number of patients with different but related symptoms
- Occurrences of physical phenomena at different locations
- Number of occurrences of various crime types

Motivation Count time series

Examples of correlated counts observed at different time points:

- Number of patients with different but related symptoms
- Occurrences of physical phenomena at different locations
- Number of occurrences of various crime types
- Number of trades of different stocks in a portfolio

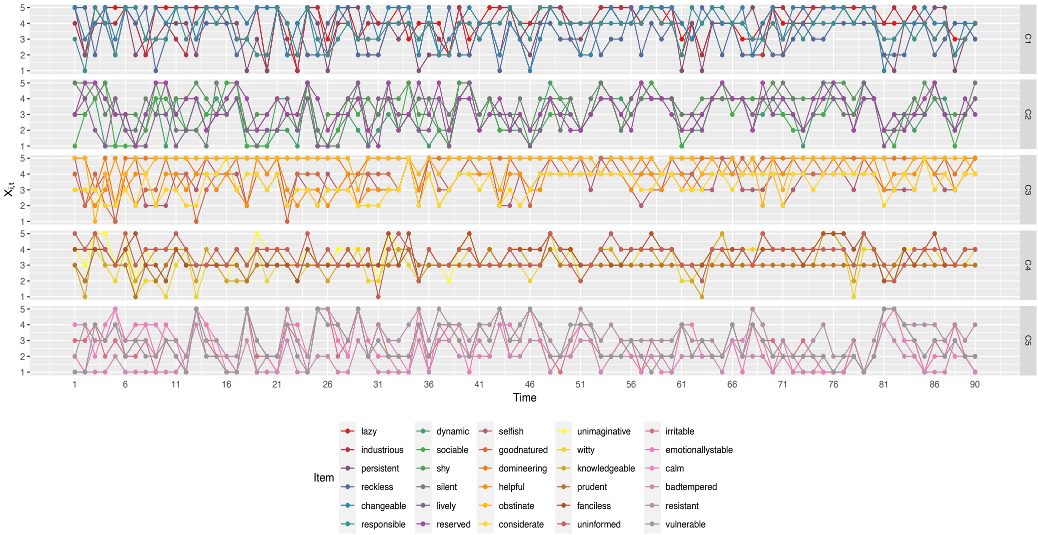
Examples of correlated counts observed at different time points:

- Number of patients with different but related symptoms
- Occurrences of physical phenomena at different locations
- Number of occurrences of various crime types
- Number of trades of different stocks in a portfolio

Literature:

- (Multivariate) INAR, INGARCH

Survey: Discussion of (dis)advantages for several classes of count time series; see Davis et al. (2021).


C1
C2
C3
C4
C5
$X_{i,t}$
Time
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
90
5
4
3
2
1
Item
lazy
industrious
persistent
reckless
changeable
responsible
dynamic
sociable
shy
silent
lively
reserved
selfish
goodnatured
domineering
helpful
obstinate
considerate
unimaginative
witty
knowledgeable
prudent
fanciless
uninformed
irritable
emotionallystable
calm
badtempered
resistant
vulnerable

Model Copula transformation and latent Gaussian variable

- Latent: $Z_t = (Z_{1,t}, \ldots, Z_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary Gaussian series with zero mean and unit variance: $\mathrm{E}[Z_{i,t}] = 0, \mathrm{E}[Z_{i,t}^2] = 1$.

Model Copula transformation and latent Gaussian variable

- Latent: $Z_t = (Z_{1,t}, \dots, Z_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary Gaussian series with zero mean and unit variance: $\mathrm{E}[Z_{i,t}] = 0, \mathrm{E}[Z_{i,t}^2] = 1$.
- Observed: $X_t = (X_{1,t}, \dots, X_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary count time series ($X_{i,t} \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$) with marginal CDF $F_i(x) = \mathrm{P}[X_{i,t} \leq x]$.

- Latent: $Z_t = (Z_{1,t}, \dots, Z_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary Gaussian series with zero mean and unit variance: $\mathrm{E}[Z_{i,t}] = 0, \mathrm{E}[Z_{i,t}^2] = 1$.
- Observed: $X_t = (X_{1,t}, \dots, X_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary count time series ($X_{i,t} \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$) with marginal CDF $F_i(x) = \mathrm{P}[X_{i,t} \le x]$.
- We model $\{X_t\}$ as

$$X_t = (X_{1,t}, \dots, X_{d,t})' = (G_1(Z_{1,t}), \dots, G_d(Z_{d,t}))' = G(Z_t),$$

with

$$G_i(z_i) = F_i^{-1}(\Phi(z_i)), \quad G(z) = (G_1(z_1), \dots, G_d(z_d))', \quad z \in \mathbb{R}^d.$$

- Latent: $Z_t = (Z_{1,t}, \dots, Z_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary Gaussian series with zero mean and unit variance: $\mathrm{E}[Z_{i,t}] = 0, \mathrm{E}[Z_{i,t}^2] = 1$.
- Observed: $X_t = (X_{1,t}, \dots, X_{d,t})'$, $t \in \mathbb{Z}$, is a d-dimensional stationary count time series ($X_{i,t} \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$) with marginal CDF $F_i(x) = \mathrm{P}[X_{i,t} \le x]$.
- We model $\{X_t\}$ as

$$X_t = (X_{1,t}, \dots, X_{d,t})' = (G_1(Z_{1,t}), \dots, G_d(Z_{d,t}))' = G(Z_t),$$

with

$$G_i(z_i) = F_i^{-1}(\Phi(z_i)), \quad G(z) = (G_1(z_1), \dots, G_d(z_d))', \quad z \in \mathbb{R}^d.$$

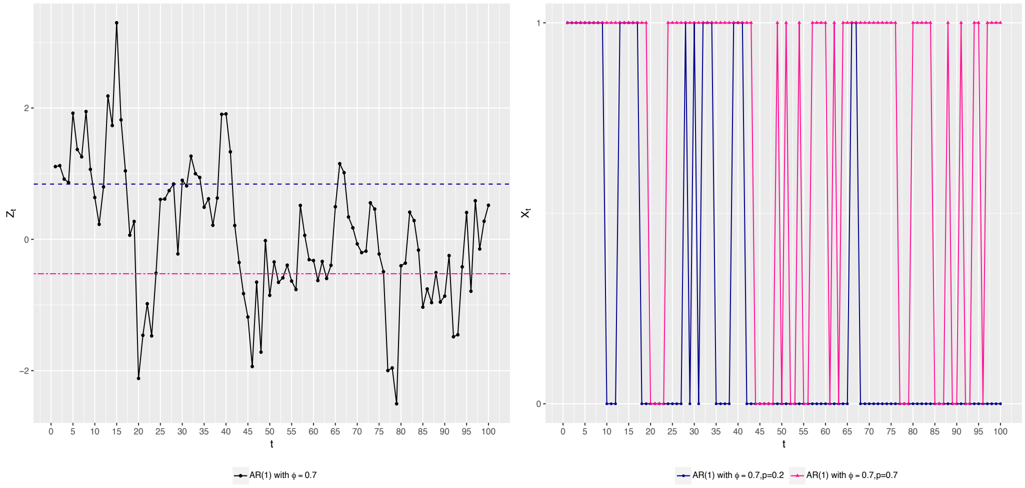
- By construction $\{X_t\}$ has marginal CDF F_i. F_i depends on a parameter vector $\theta_i \in \mathbb{R}^{K_i}$.

Example with $d = 1$ and Bernoulli marginals. $X_t \sim \text{Bernoulli}(p)$, $Z_t = \phi Z_{t-1} + \varepsilon_t$.

$$G(z) = F^{-1}(\Phi(z)) = \begin{cases} 1, & z \geq \Phi^{-1}(1-p), \\ 0, & z < \Phi^{-1}(1-p). \end{cases}$$

Example with $d = 1$ and Bernoulli marginals. $X_t \sim \text{Bernoulli}(p)$, $Z_t = \phi Z_{t-1} + \varepsilon_t$.

$$G(z) = F^{-1}(\Phi(z)) = \begin{cases} 1, & z \geq \Phi^{-1}(1-p), \\ 0, & z < \Phi^{-1}(1-p). \end{cases}$$



Model Autocovariance matrices and their relationships

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i))$.

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i))$.

How can we associate the dependences structure of $\{Z_t\}$ and $\{X_t\}$?

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i))$.

How can we associate the dependences structure of $\{Z_t\}$ and $\{X_t\}$?

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i))$.

How can we associate the dependences structure of $\{Z_t\}$ and $\{X_t\}$?

- Expand G_i using Hermite polynomials

$$G_i(z) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z), \quad c_{i,k} = \mathrm{E}(G_i(Z_{i,0})H_k(Z_{i,0}));$$

with the kth Hermite polynomial defined as

$$H_k(z) = (-1)^k e^{\frac{z^2}{2}} \frac{\partial^k}{\partial z^k} e^{\frac{-z^2}{2}}.$$

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i))$.

How can we associate the dependences structure of $\{Z_t\}$ and $\{X_t\}$?

- Expand G_i using Hermite polynomials

$$G_i(z) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z), \quad c_{i,k} = \mathrm{E}(G_i(Z_{i,0}) H_k(Z_{i,0}));$$

with the kth Hermite polynomial defined as

$$H_k(z) = (-1)^k e^{\frac{z^2}{2}} \frac{\partial^k}{\partial z^k} e^{\frac{-z^2}{2}}.$$

Examples: $H_0(z) = 1, H_1(z) = z, H_2(z) = z^2 - 1, H_3(z) = z^3 - 3z, H_4(z) = z^4 - 6z^2 + 3, \dots$

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i))$.

How can we associate the dependences structure of $\{Z_t\}$ and $\{X_t\}$?

- Expand G_i using Hermite polynomials

$$G_i(z) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z), \quad c_{i,k} = \mathrm{E}(G_i(Z_{i,0}) H_k(Z_{i,0}));$$

with the kth Hermite polynomial defined as

$$H_k(z) = (-1)^k e^{\frac{z^2}{2}} \frac{\partial^k}{\partial z^k} e^{\frac{-z^2}{2}}.$$

Examples: $H_0(z) = 1, H_1(z) = z, H_2(z) = z^2 - 1, H_3(z) = z^3 - 3z, H_4(z) = z^4 - 6z^2 + 3, \ldots$

- Jia et al. (2021):

$$c_{i,k} = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} e^{-Q_{i,n}^2/2} H_{k-1}(Q_{i,n})$$

with $Q_{i,n} = \Phi^{-1}(C_{i,n})$ and $C_{i,n} = \mathrm{P}[X_{i,t} \leq n]$. $C_{i,n}$ depends on parameter vector $\theta_i \in \mathbb{R}^{K_i}$.

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i)) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z)$.

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i)) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z)$.

- The autocovariances of $\{Z_t\}$ (latent) can be associated to $\{X_t\}$ (observed) through

$$\Gamma_X(h) := \mathrm{E}[X_{t+h}X_t'] - \mathrm{E}[X_{t+h}]\,\mathrm{E}[X_t'] = \left(\sum_{k=1}^{\infty} \frac{c_{i,k}c_{j,k}}{k!} R_{Z,ij}(h)^k \right)_{i,j=1,\dots,d}, R_Z(h) = \mathrm{E}[Z_{t+h}Z_t'].$$

Recall that $G_i(z_i) = F_i^{-1}(\Phi(z_i)) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z)$.

- The autocovariances of $\{Z_t\}$ (latent) can be associated to $\{X_t\}$ (observed) through

$$\Gamma_X(h) := \mathrm{E}[X_{t+h}X_t'] - \mathrm{E}[X_{t+h}]\,\mathrm{E}[X_t'] = \left(\sum_{k=1}^{\infty} \frac{c_{i,k}c_{j,k}}{k!} R_{Z,ij}(h)^k\right)_{i,j=1,\dots,d}, R_Z(h) = \mathrm{E}[Z_{t+h}Z_t'].$$

- More compactly

$$\Gamma_X(h) = \ell(\Gamma_Z(h)), \;\; \ell(u) = (\ell_{ij}(u))_{i,j=1,\dots,d} \quad \text{with} \quad \ell_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k}c_{j,k}}{k!} u^k.$$

- $c_{i,k}$ depends on $\theta_i \rightarrow \ell_{ij}$ depends on θ_i, θ_j.

Example

Example with $d = 2$ and Bernoulli marginals.

- $Z_t = (Z_{1,t}, Z_{2,t})'$, $t \in \mathbb{Z}$, stationary Gaussian series with $\mathrm{E}[Z_{i,t}] = 0, \mathrm{E}[Z_{i,t}^2] = 1$ and lag-h autocovariance matrix

$$\Gamma_Z(h) = \begin{pmatrix} \rho_{1,1}(h) & \rho_{1,2}(h) \\ \rho_{2,1}(h) & \rho_{2,2}(h) \end{pmatrix}.$$

- $X_{i,t} \sim \text{Bernoulli}(p)$.
- Then,

$$G_i(z) = F_i^{-1}(\Phi(z)) = \begin{cases} 1, & z \geq \Phi^{-1}(1-p), \\ 0, & z < \Phi^{-1}(1-p). \end{cases}$$

Suppose $p = \frac{1}{2}$ such that $\Phi^{-1}(1-p) = 0$

- Then,

$$\Gamma_X(h) = \frac{1}{2\pi} \begin{pmatrix} \arcsin(\rho_{1,1}(h)) & \arcsin(\rho_{1,2}(h)) \\ \arcsin(\rho_{2,1}(h)) & \arcsin(\rho_{2,2}(h)) \end{pmatrix}.$$

Recall that $\ell(u) = (\ell_{ij}(u))_{i,j=1,\dots,d}$ with $\ell_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k}c_{j,k}}{k!} u^k$.

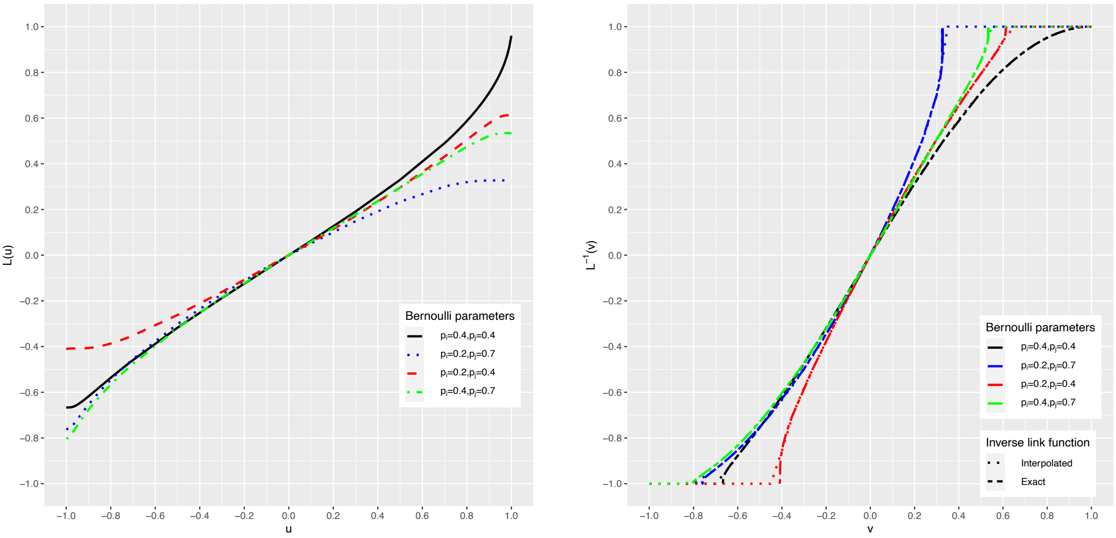
Proposition (Jia et al. (2021))

For $u \in (-1,1)$, the link function ℓ satisfies

$$\ell'_{ij}(u) = \frac{1}{2\pi\sqrt{1-u^2}} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)}(Q^2_{i,n_0} + Q^2_{j,n_1} - 2uQ_{i,n_0}Q_{j,n_1})\right)$$

with $Q_{i,n} = \Phi^{-1}(C_{i,n})$ and $C_{i,n} = \mathrm{P}[X_{i,t} \leq n]$.

There is no explicit representation of ℓ.



L(u)
u
Bernoulli parameters
p_i=0.4,p_j=0.4
p_i=0.2,p_j=0.7
p_i=0.2,p_j=0.4
p_i=0.4,p_j=0.7
$L^{-1}(v)$
v
Bernoulli parameters
p_i=0.4,p_j=0.4
p_i=0.2,p_j=0.7
p_i=0.2,p_j=0.4
p_i=0.4,p_j=0.7
Inverse link function
Interpolated
Exact

Goals

Goals

- Find estimator for autocovariance matrices of latent process.

Goals

- Find estimator for autocovariance matrices of latent process.
- Establish concentration bounds for the differences between the estimated and true latent Gaussian autocovariances, in terms of those for the observed count series and the estimated marginal parameters. (Main result I)

Goals

- Find estimator for autocovariance matrices of latent process.
- Establish concentration bounds for the differences between the estimated and true latent Gaussian autocovariances, in terms of those for the observed count series and the estimated marginal parameters. (Main result I)
- Impose parametric model on latent process (e.g. VAR).

Goals

- Find estimator for autocovariance matrices of latent process.
- Establish concentration bounds for the differences between the estimated and true latent Gaussian autocovariances, in terms of those for the observed count series and the estimated marginal parameters. (Main result I)
- Impose parametric model on latent process (e.g. VAR).
- Utilize Main result I to find high probability bound on sparse estimators for transition matrices of VAR. (Main result II)

Literature

- Univariate case ($d = 1$): Jia et al. (2021) suggests several approaches to estimate parameters in Γ_Z. Kong and Lund (2022) studied seasonal case.

Literature

- Univariate case ($d = 1$): Jia et al. (2021) suggests several approaches to estimate parameters in Γ_Z. Kong and Lund (2022) studied seasonal case.
- Multivariate case ($d = 2$): Livsey et al. (2018) for Poisson marginals.

Literature

- Univariate case ($d = 1$): Jia et al. (2021) suggests several approaches to estimate parameters in Γ_Z. Kong and Lund (2022) studied seasonal case.
- Multivariate case ($d = 2$): Livsey et al. (2018) for Poisson marginals.
- Survey: Discussion of (dis)advantages for several classes of count time series; see Davis et al. (2021).

Literature

- Univariate case ($d = 1$): Jia et al. (2021) suggests several approaches to estimate parameters in Γ_Z. Kong and Lund (2022) studied seasonal case.
- Multivariate case ($d = 2$): Livsey et al. (2018) for Poisson marginals.
- Survey: Discussion of (dis)advantages for several classes of count time series; see Davis et al. (2021).
- Psychometrics: Here t may no longer refer to time but, for example, different individuals; see Lee et al. (1992).

Literature

- Univariate case ($d = 1$): Jia et al. (2021) suggests several approaches to estimate parameters in Γ_Z. Kong and Lund (2022) studied seasonal case.
- Multivariate case ($d = 2$): Livsey et al. (2018) for Poisson marginals.
- Survey: Discussion of (dis)advantages for several classes of count time series; see Davis et al. (2021).
- Psychometrics: Here t may no longer refer to time but, for example, different individuals; see Lee et al. (1992).
- Statistical literature: Studies of Gaussian copula models guaranteeing consistent estimation of latent correlation structures have been done when entries of the copula correlation matrix relate to the entries of the Kendall's tau or Spearman's rho matrices through an explicit link function; see Liu et al. (2012); Mitra and Zhang (2014); Wegkamp and Zhao (2016); Han and Liu (2017); Fan et al. (2017).

Literature

- Univariate case ($d = 1$): Jia et al. (2021) suggests several approaches to estimate parameters in Γ_Z. Kong and Lund (2022) studied seasonal case.
- Multivariate case ($d = 2$): Livsey et al. (2018) for Poisson marginals.
- Survey: Discussion of (dis)advantages for several classes of count time series; see Davis et al. (2021).
- Psychometrics: Here t may no longer refer to time but, for example, different individuals; see Lee et al. (1992).
- Statistical literature: Studies of Gaussian copula models guaranteeing consistent estimation of latent correlation structures have been done when entries of the copula correlation matrix relate to the entries of the Kendall's tau or Spearman's rho matrices through an explicit link function; see Liu et al. (2012); Mitra and Zhang (2014); Wegkamp and Zhao (2016); Han and Liu (2017); Fan et al. (2017).

Our contributions: Theoretical guarantees for consistent estimation of a latent parametric model in a possibly high-dimensional regime.

Recall that $\ell(u) = (\ell_{ij}(u))_{i,j=1,\dots,d}$ with $\ell_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k} c_{j,k}}{k!} u^k$.

- The function ℓ_{ij} depends on the marginal CDF parameters θ_i, θ_j only, so that for all lags h,

$$\Gamma_X(h) = \ell(\Gamma_Z(h)).$$

Recall that $\ell(u) = (\ell_{ij}(u))_{i,j=1,\ldots,d}$ with $\ell_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k} c_{j,k}}{k!} u^k$.

- The function ℓ_{ij} depends on the marginal CDF parameters θ_i, θ_j only, so that for all lags h,

$$\Gamma_X(h) = \ell(\Gamma_Z(h)).$$

- If $\widehat{\theta}_i$ is an estimator of θ_i and $\widehat{\ell}$ one of ℓ, one estimator of $\widehat{\Gamma}_Z$ is

$$\widehat{\Gamma}_Z(h) = \widehat{\ell}^{-1}(\widehat{\Gamma}_X(h)),$$

where $\widehat{\Gamma}_X(h)$ is a standard ACVF estimator of $\Gamma_X(h)$ based on $X_1, \ldots, X_T$.

Recall that $\ell(u) = (\ell_{ij}(u))_{i,j=1,\dots,d}$ with $\ell_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k}c_{j,k}}{k!} u^k$.

- The function ℓ_{ij} depends on the marginal CDF parameters θ_i, θ_j only, so that for all lags h,

$$\Gamma_X(h) = \ell(\Gamma_Z(h)).$$

- If $\widehat{\theta}_i$ is an estimator of θ_i and $\widehat{\ell}$ one of ℓ, one estimator of $\widehat{\Gamma}_Z$ is

$$\widehat{\Gamma}_Z(h) = \widehat{\ell}^{-1}(\widehat{\Gamma}_X(h)),$$

  where $\widehat{\Gamma}_X(h)$ is a standard ACVF estimator of $\Gamma_X(h)$ based on $X_1, \dots, X_T$.
- Assuming that the observations have a zero mean, the autocovariance matrices $\mathbf{\Gamma}_X = (\Gamma_X(r-s))_{r,s=1,\dots,L}$ can be estimated as $\widehat{\mathbf{\Gamma}}_X = \frac{1}{N}\mathcal{X}_X'\mathcal{X}_X$ with $N = T - L$ and

$$\mathcal{X}_X = \begin{pmatrix} X_L' & \dots & X_1' \\ \vdots & \ddots & \vdots \\ X_{T-1}' & \dots & X_{T-L}' \end{pmatrix}.$$

  With a slight abuse of notation, we write both $\mathbf{\Gamma}_X = \ell(\mathbf{\Gamma}_Z)$ and $\Gamma_X(h) = \ell(\Gamma_Z(h))$.

Main result I

Recall that $\widehat{\Gamma}_Z(h) = \widehat{\ell}^{-1}(\widehat{\Gamma}_X(h))$ and set $\mathbf{\Gamma}_X = (\Gamma_X(r-s))_{r,s=1,\dots,L}$

Proposition (D., Lund, Pipiras)

Under mild moment conditions *on* $\{X_t\}$, *we have, for* $\delta, \varepsilon > 0$,

$$\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)\delta\right] \lesssim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta]$$

with $Q(\mathbf{\Gamma}_Z) := Q(\mathbf{\Gamma}_Z, \varepsilon, \delta)$.

The constant $Q(\mathbf{\Gamma}_Z)$ depends on

$$\mu_i^{(k)}(u) = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) |Q_{i,n}|^k \left(\|\nabla_{\theta_i} Q_{i,n}\|_1\right)^b, \quad b \in \{0,1\}.$$

Sparse estimation for latent VAR processes

$$Z_t = \sum_{u=1}^{p} \Phi_u Z_{t-u} + \varepsilon_t, \quad t \in \mathbb{Z},$$

for some $\Phi_u \in \mathbb{R}^{d \times d}$ and white noise series $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ characterized by

$$\mathrm{E}[\varepsilon_t] = 0, \quad \mathrm{E}[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon, \quad \mathrm{E}[\varepsilon_s \varepsilon_t'] = 0 \quad \text{for } s \neq t.$$

The VAR(p) model can be written in a linear models form as

$$\begin{pmatrix} Z'_{p+1} \\ \vdots \\ Z'_T \end{pmatrix} = \begin{pmatrix} Z'_p & \dots & Z'_1 \\ \vdots & \ddots & \vdots \\ Z'_{T-1} & \dots & Z'_{T-p} \end{pmatrix} \begin{pmatrix} \Phi'_1 \\ \vdots \\ \Phi'_p \end{pmatrix} + \begin{pmatrix} \varepsilon'_{p+1} \\ \vdots \\ \varepsilon'_T \end{pmatrix} \quad \text{or} \quad \mathcal{Y}_Z = \mathcal{X}_Z B_0 + \mathcal{E}.$$

A vectorized version:

$$\begin{aligned} \mathrm{vec}(\mathcal{Y}_Z) &= \mathrm{vec}(\mathcal{X}_Z B_0) + \mathrm{vec}(\mathcal{E}) \\ &= (I_d \otimes \mathcal{X}_Z)\,\mathrm{vec}(B_0) + \mathrm{vec}(\mathcal{E}), \\ Y &= Z\beta_0 + E, \end{aligned}$$

Estimation

Recall that $\text{vec}(\mathcal{Y}_Z) = \text{vec}(\mathcal{X}_Z B_0) + \text{vec}(\mathcal{E})$ and $Y = Z\beta_0 + E$

The transition matrices can be estimated through

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^q}{\arg\min} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

- observed Basu and Michailidis (2015)

$$\widehat{\gamma} = \text{vec}(\widehat{\gamma}_Z) = \text{vec}(\mathcal{X}_Z' \mathcal{Y}_Z),$$
$$\widehat{\Gamma} = I_d \otimes \widehat{\mathbf{\Gamma}}_Z = I_d \otimes \mathcal{X}_Z' \mathcal{X}_Z / N.$$

- unobserved D., Lund, Pipiras (2023)

$$\widehat{\gamma} = \text{vec}(\widehat{\gamma}_Z) = \text{vec}(\widehat{\ell}^{-1}(\widehat{\gamma}_X)),$$
$$\widehat{\Gamma} = I_d \otimes \widehat{\mathbf{\Gamma}}_Z = I_d \otimes \widehat{\ell}^{-1}(\widehat{\mathbf{\Gamma}}_X).$$

General path to consistency

Observed Basu and Michailidis (2015)

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^q} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

General path to consistency

Observed Basu and Michailidis (2015)

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^q}{\arg\min} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

Restricted Eigenvalue: A symmetric matrix $\widehat{\Gamma} \in \mathbb{R}^{q \times q}$ satisfies the restricted eigenvalue condition with curvature $\alpha > 0$ and tolerance $\tau > 0$ if

$$x'\widehat{\Gamma}x \geq \alpha \|x\|^2 - \tau \|x\|_1^2 \quad \text{for all} \quad x \in \mathbb{R}^q.$$

We write $\widehat{\Gamma} \sim RE(\alpha, \tau)$ for short.

General path to consistency

Observed Basu and Michailidis (2015)

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^q} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

Restricted Eigenvalue: A symmetric matrix $\widehat{\Gamma} \in \mathbb{R}^{q \times q}$ satisfies the restricted eigenvalue condition with curvature $\alpha > 0$ and tolerance $\tau > 0$ if

$$x'\widehat{\Gamma}x \geq \alpha \|x\|^2 - \tau \|x\|_1^2 \quad \text{for all} \quad x \in \mathbb{R}^q.$$

We write $\widehat{\Gamma} \sim RE(\alpha, \tau)$ for short.
Deviation bound: There exists a deterministic function $Q(\beta_0)$ such that

$$\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \leq Q(\beta_0)\sqrt{\frac{\log(q)}{N}}, \quad N = T - p, \quad q = d^2 p.$$

General path to consistency

Observed Basu and Michailidis (2015)

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^q}{\arg\min} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

Restricted Eigenvalue: A symmetric matrix $\widehat{\Gamma} \in \mathbb{R}^{q \times q}$ satisfies the restricted eigenvalue condition with curvature $\alpha > 0$ and tolerance $\tau > 0$ if

$$x'\widehat{\Gamma}x \geq \alpha \|x\|^2 - \tau \|x\|_1^2 \quad \text{for all} \quad x \in \mathbb{R}^q.$$

We write $\widehat{\Gamma} \sim RE(\alpha, \tau)$ for short.
Deviation bound: There exists a deterministic function $Q(\beta_0)$ such that

$$\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \leq Q(\beta_0)\sqrt{\frac{\log(q)}{N}}, \quad N = T - p, \quad q = d^2 p.$$

Both properties can be reduced to finding bounds on:

$$\mathrm{P}[|v'(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)v| > \delta]$$

Path to consistency for latent process

Unobserved D., Lund, Pipiras (2023)

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^q}{\arg\min} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

Path to consistency for latent process

Unobserved D., Lund, Pipiras (2023)

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^q} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

with

$$\widehat{\gamma} = \text{vec}(\widehat{\gamma}_Z) = \text{vec}(\widehat{\ell}^{-1}(\widehat{\gamma}_X)),$$
$$\widehat{\Gamma} = I_d \otimes \widehat{\boldsymbol{\Gamma}}_Z = I_d \otimes \widehat{\ell}^{-1}(\widehat{\boldsymbol{\Gamma}}_X).$$

Path to consistency for latent process

Unobserved D., Lund, Pipiras (2023)

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^q} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

with

$$\widehat{\gamma} = \text{vec}(\widehat{\gamma}_Z) = \text{vec}(\widehat{\ell}^{-1}(\widehat{\gamma}_X)),$$
$$\widehat{\Gamma} = I_d \otimes \widehat{\mathbf{\Gamma}}_Z = I_d \otimes \widehat{\ell}^{-1}(\widehat{\mathbf{\Gamma}}_X).$$

Restricted Eigenvalue condition and **Deviation bound** can be proven by reducing problem to finding bound on

$$\mathrm{P}[\sup_{v \in \mathcal{K}(2s)} |v'(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)v| > \delta].$$

Path to consistency for latent process

Unobserved D., Lund, Pipiras (2023)

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^q} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big).$$

with

$$\widehat{\gamma} = \text{vec}(\widehat{\gamma}_Z) = \text{vec}(\widehat{\ell}^{-1}(\widehat{\gamma}_X)),$$
$$\widehat{\Gamma} = I_d \otimes \widehat{\mathbf{\Gamma}}_Z = I_d \otimes \widehat{\ell}^{-1}(\widehat{\mathbf{\Gamma}}_X).$$

Restricted Eigenvalue condition and **Deviation bound** can be proven by reducing problem to finding bound on

$$\mathsf{P}[\sup_{v \in \mathcal{K}(2s)} |v'(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)v| > \delta].$$

Apply Main result I:

$$\mathsf{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)\delta\right] \lesssim \mathsf{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta] + \mathsf{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta]$$
$$+ \mathsf{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon] + \mathsf{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta].$$

Main result II

Recall that $\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^q} \left(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \right)$.

There exist finite positive constants c_1 and c_2 such that for any $v \in \mathcal{K}(2s)$,

$$\mathrm{P}[|v'(\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)v| > \delta] \leq c_1 \exp\left(-c_2 \frac{N\delta^2}{s^2}\right), \quad N = T - p.$$

There exist finite positive constants c_1 and c_2 such that

$$\mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \leq c_1 dK \exp\left(-c_2 T\varepsilon^2\right).$$

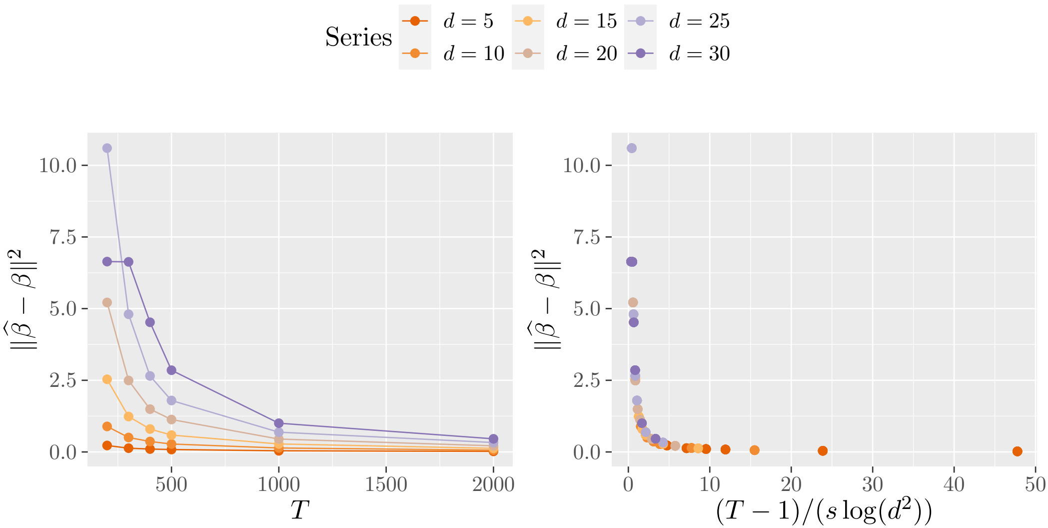
Proposition (D., Lund, Pipiras)

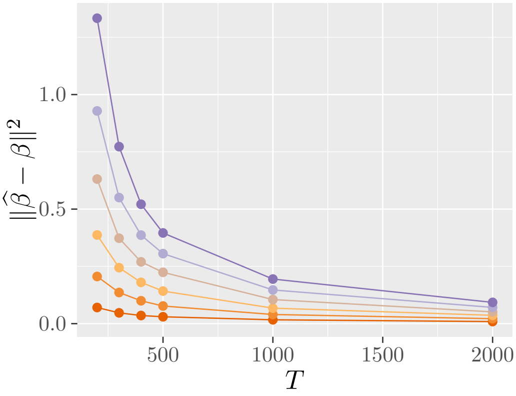
Then, with high probability, for any $\lambda_N \geq 4Q(\beta_0)\sqrt{\frac{\log(q)}{N}}$,

$$\|\widehat{\beta} - \beta_0\| \leq 64s\frac{\lambda_N}{\alpha}.$$

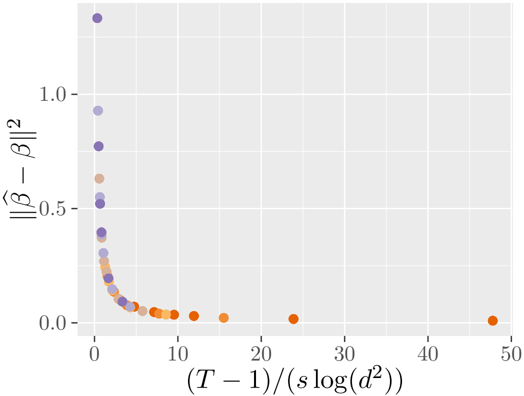
Same convergence rate as in Basu and Michailidis (2015)!

▸ More


Series
$d = 5$
$d = 10$
$d = 15$
$d = 20$
$d = 25$
$d = 30$
10.0
7.5
5.0
2.5
0.0
$\|\widehat{\beta} - \beta\|^2$
500
1000
1500
2000
T
0
10
20
30
40
50
$(T-1)/(s\log(d^2))$



1.0
0.5
0.0
$\|\widehat{\beta} - \beta\|^2$
500
1000
1500
2000
T



1.0
0.5
0.0
$\|\widehat{\beta} - \beta\|^2$
0
10
20
30
40
50
$(T-1)/(s\log(d^2))$

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.
  - Theoretical verifications for consistent model selection.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.
  - Theoretical verifications for consistent model selection.
  - Extensions to non-stationary models, particularly those involving covariates.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.
  - Theoretical verifications for consistent model selection.
  - Extensions to non-stationary models, particularly those involving covariates.
  - Estimation of related model parameters of the latent process like the VAR order.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.
  - Theoretical verifications for consistent model selection.
  - Extensions to non-stationary models, particularly those involving covariates.
  - Estimation of related model parameters of the latent process like the VAR order.
  - Extensions to spatial settings.

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.
  - Theoretical verifications for consistent model selection.
  - Extensions to non-stationary models, particularly those involving covariates.
  - Estimation of related model parameters of the latent process like the VAR order.
  - Extensions to spatial settings.
  - Can we use other distributions for the latent process? For example to achieve tail dependence?

Conclusions

- Summary
  - Possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process.
  - Derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.
  - Results can be used to estimate underlying parametric model.
  - Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR at the same rate as for an observable VAR series.
  - Non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series is allowed to grow with the sample size.
- Related future work
  - Other shrinkage methods, for example, adaptive penalization.
  - Assume that the latent process follows a dynamic factor model.
  - Theoretical verifications for consistent model selection.
  - Extensions to non-stationary models, particularly those involving covariates.
  - Estimation of related model parameters of the latent process like the VAR order.
  - Extensions to spatial settings.
  - Can we use other distributions for the latent process? For example to achieve tail dependence?
  - How to derive consistency for possibly unbounded functions of the latent Gaussian process.

Sumanta Basu and George Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.

Richard A Davis, Konstantinos Fokianos, Scott H Holan, Harry Joe, James Livsey, Robert Lund, Vladas Pipiras, and Nalini Ravishanker. Count time series: A methodological review. *Journal of the American Statistical Association*, pages 1–15, 2021.

Jianqing Fan, Han Liu, Yang Ning, and Hui Zou. High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79 (2):405–421, 2017.

Jianqing Fan, Bai Jiang, and Qiang Sun. Hoeffding's inequality for general Markov chains and its applications to statistical learning. *Journal of Machine Learning Research*, 22(139):1–35, 2021.

Fang Han and Han Liu. Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli*, 23(1):23–57, 2017.

Yisu Jia, Stefanos Kechagias, James Livsey, Robert Lund, and Vladas Pipiras. Latent Gaussian count time series. *Journal of the American Statistical Association*, pages 1–28, 2021.

Younghoon Kim, Zachary F. Fisher, and Vladas Pipiras. Latent Gaussian dynamic factor modeling and forecasting for multivariate count time series. *arXiv preprint*, 2022.

Jiajie Kong and Robert Lund. Seasonal count time series. *Journal of Time Series Analysis*, 2022.

Sik-Yum Lee, Wai-Yin Poon, and Peter M Bentler. Structural equation models with continuous and polytomous variables. *Psychometrika*, 57(1):89–105, 1992.

Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman. High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012.

James Livsey, Robert Lund, Stefanos Kechagias, and Vladas Pipiras. Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics*, 12(1):408–431, 2018.

Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664, 2012.

Ritwik Mitra and Cun-Hui Zhang. Multivariate analysis of nonparametric estimates of large correlation matrices. *arXiv preprint arXiv:1403.6195*, 2014.

Linda K Muthén and Bengt O Muthén. Mplus 7.11. *Los Angeles, CA: Muthén & Muthén*, 1998—2017.

Vladas Pipiras and Murad S. Taqqu. Long-range dependence and self-similarity. *Cambridge Books*, 2017.

Marten Wegkamp and Yue Zhao. Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli*, 22(2):1184–1226, 2016.