Yule–Walker Equations Using a Gini Covariance Matrix for the High-Dimensional Heavy-Tailed PVAR Model

Jin Zou* and Dong Han

School of Mathematics Science, Shanghai Jiao Tong University, Shanghai 200240, China; donghan@sjtu.edu.cn
* Correspondence: mathzj@alumni.sjtu.edu.cn

Abstract: Gini covariance plays a vital role in analyzing the relationship between random variables with heavy-tailed distributions. In this paper, with the existence of a finite second moment, we establish the Gini–Yule–Walker equation to estimate the transition matrix of high-dimensional periodic vector autoregressive (PVAR) processes, the asymptotic results of estimators have been established. We apply this method to study the Granger causality of the heavy-tailed PVAR process, and the results show that the robust transfer matrix estimation induces sign consistency in the value of Granger causality. Effectiveness of the proposed method is verified by both synthetic and real data.

Keywords: high dimensional time series; transition matrix estimation; Gini covariance matrix; U-statistics

1. Introduction

The heavy-tailed distribution is graphically thicker in the tails and sharper in the peaks than the normal distribution. Intuitively, this means that the probability of extreme values is greater for these data than for normally distributed data. The heavy-tailed phenomenon has been encountered empirically in various fields: physics, earth sciences, economics and political science, etc. Periodicity is a wave-like or oscillatory movement around a long-term trend presented in a time series. It is well-known that cyclical causality caused by business and economic activities is usually different from trend movements, not in a single direction of continuous movement, but in alternating fluctuations between ups and downs. When these components, trend and cyclicality, do not evolve independently, traditional differencing filters may not be suitable (see for example, Franses and Paap [1], Bell et al. [2], Aliat and Hamdi [3]). Periodic time series models are used to model periodically stationary time series. Periodic vector autoregressive (PVAR) models extend the classical vector autoregressive (VAR) models by allowing the parameters to vary with the cyclicity.

\[ X_{iT+v} = \sum_{k=1}^{q(v)} a_k(v) X_{iT+v-k} + \epsilon_{iT+v}. \] (1)

For fixed \( v \) and predetermined value \( T \), the random vector \( X_{iT+v} \) represents the realization during the \( v \)th season, with \( v \in \{1, \ldots, T\} \), at year \( i \), \( i \in \{1, \ldots, n\} \). The PVAR model order at season \( v \) is given by \( q(v) \), whereas \( a_k(v), k = 1, \ldots, q(v), \) represent the PVAR model coefficients during season \( v \). The error process \( \epsilon = \{ \epsilon_{iT+v} \} \) corresponds to a periodic white noise, with \( E(\epsilon_{iT+v}) = 0 \) and \( \text{var}(\epsilon_{iT+v}) = \sigma^2(v)I \), where \( 0 \) is a zero vector of order \( d \) and \( I \) is a unit matrix of order \( d \).

This paper seeks to establish a PVAR model to simulate the heavy-tailed time series. Let random vectors \( X_1, \ldots, X_n \) be from a stationary stochastic process \( (X_k)^\infty_{k=-\infty} \), \( X_k = (X_{k1}, \ldots, X_{Kn})^T \), where the transpose (indicated by \( T \) of a row vector is a column vector.
We consider first-order PVAR model,

\[ X_{kT+v} = AX_{kT+v-1} + E_{kT+v}, \tag{2} \]

where \( k \in \{1, \cdots, n\}, v \in \{2, \cdots, T\}, T \) is the period of \( X_k \), \( E_{kT+v} \) denotes the latent heavy-tailed innovation, and \( A \) is the transition matrix. In this paper, we assume that the entries of the heavy-tailed innovation \( E_{kT+v} = (E_{kT+v_1}, \cdots, E_{kT+v_d})^\top \) obey a power-law distribution with \( P(|E_{kT+v}| > x) \sim Cx^{-\alpha_i} \) and \( \alpha_i \leq 2 \), then the second moment is finite and the third moment is infinite. From the Equation (2) we know that, given \( k \), the vector \( X_k \) follows a first-order vector autoregressive (VAR). We call the first-order PVAR process stable if all eigenvalues of transition matrix \( A \) have modulus less than 1, the condition is equivalent to \( \det(I_d - Az) \neq 0 \) for all \( |z| \leq 1 \) (see for example, Section 2.1 in Lütkepohl [4]).

The transition matrix characterizes the temporal dependence for sequence data, and plays a fundamental role in forecasting. Moreover, the zero and nonzero entries of the transition matrix is often closely related to Granger causality. This manuscript focuses on estimating the transition matrix of high-dimensional heavy-tailed PVAR processes.

PVAR models have been extensively studied under the Gaussian assumption, the Gaussian PVAR models assume that the latent innovations are independent identity distribution Gaussian random vectors. Under this model, there are two kinds of methods to estimate the transition matrix under high dimensional setting, one is the Lasso-based estimation procedures, see [5–8], and the other is Dantzig-selector-type estimators, see [9–12]. Under the non-Gaussian VAR process, Qiu et al. [13] proposed a quantile-based dantzig-selector-type estimator of the transition matrix for elliptical VAR process. Wong et al. [14] provided an alternative proof of the consistency of the Lasso for sparse non-Gaussian VAR models. Maleki et al. [15] extended the multivariate setting of autoregressive process, by considering the multivariate scale mixture of skew-normal distributions for VAR innovations.

The statistical second-order information contained in the data is usually expressed by the variance and covariance, most of the literature dealing with time series measure dependence using the variance and covariance. To investigate the validity of the variance estimates, we need the presence of the fourth order moment of the random variables. For the heavy-tailed nature of financial data, the third-order moments of random variables are usually non-existent. Schechtman and Yitzhaki [16] proposed the concept of Gini covariance, which has been used widely to measure the dependence of heavy-tailed distributions. Let \( H \) be the joint distribution of the random variables \( X \) and \( Y \) with the marginal distribution functions \( F_1 \) and \( F_2 \), respectively. The standard Gini covariance is defined as

\[ \text{Gcov}(X, Y) = \text{cov}(X, F_2(Y)) \quad \text{and} \quad \text{Gcov}(Y, X) = \text{cov}(Y, F_1(X)), \tag{3} \]

assuming the random variables with only a finite first moment. The Gini covariance has an advantage when analyzing bivariate data defined by both variate values and ranks of the values. The representation of Gini covariance \( \text{Gcov}(X, Y) \) indicates that it has mixed properties of the variable \( X \) and the rank of the variable \( Y \), and thus complements the usual covariance and rank covariance [16–18]. In terms of balance between efficiency and robustness, Gini covariance plays an important role in measuring association for variables from heavy-tailed distributions [19].

The Yule–Walker equations arise naturally in the problem of linear prediction of any zero-mean weakly stationary process based on a finite number of contiguous observations. The Yule–Walker equations provide a straightforward connection between the autoregressive model parameters and the covariance function of the process. In this paper, relaxing the strong assumption of the existence of higher order moments of the regressors, we use a non-parametric method to estimate the Gini covariance matrix, establish the Gini–Yule–Walker equation to estimate the sparse transition matrix of stationary PVAR processes. The estimator falls into the category of Dantzig-selector-type estimators. With existence
of a finite second moment, we investigate the asymptotic behavior of the estimator in high dimensions.

The paper is organized as follows: In Section 2, we establish the Gini–Yule–Walker equation and estimate the sample Gini covariance matrix. In Section 3, we derive the convergence rate of transfer matrix estimation. In Section 4, we discuss the characterization and estimation of Granger causality under the heavy-tailed PVAR model. In Section 5, both synthetic and real data are used to demonstrate the empirical performance of the proposed methodology.

2. Model

In this section, the notations are set. Then, we establish the Gini–Yule–Walker equation, obtain simple non-parametric estimators for Gini covariance matrix and investigate the convergence rate of the sample Gini covariance matrix.

2.1. Notation

Let \( v = (v_1, ..., v_d)^\top \) be a \( d \)-dimensional real vector, and \( M = [M_{ij}] \in \mathbb{R}^{d_1 \times d_2} \) be a \( d_1 \times d_2 \) matrix. For \( 0 < q < \infty \), we define the vector \( l_q \) norm of \( v \) as \( \|v\|_q = (\sum_{j=1}^d |v_j|^q)^{1/q} \), and the vector \( l_\infty \) norm of \( v \) as \( \|v\|_\infty = \max_{j=1}^d |v_j| \). Let the matrix \( l_1 \) norm of \( M \) as \( \|M\|_1 = \max_{|A_{ij}|} \sum_{i=1}^d |A_{ij}| \), the matrix \( l_\infty \) norm \( \|M\|_\infty = \max_{|B_{ij}|} \sum_{i=1}^d |B_{ij}| \), \( X_k = (X_{1k}, ..., X_{dk})^\top \) and \( Y_k = (Y_{1k}, ..., Y_{dk})^\top \) be two random vectors.

2.2. Gini–Yule–Walker Equation

In this paper, we model the time series vector by a stationary PVAR process under the existence of second moment. For each \( k, 1 \leq k \leq n \), \( X_{1k}, ..., X_{Tk} \in \mathbb{R}^d \) follow a lag-one VAR process, with \( E_{ik} \) independent of \( X_{(t-1)k} \), and \( E(X_{tk}) = 0 \), \( E(E_{ik}) = 0 \).

We define

\[
X_k = (X_{1k}, ..., X_{(T-1)k})^\top,
\]

this VAR process may be performed by concatenating the equation systems to analyze the following equation,

\[
y_k = Y_k \beta + \epsilon_k,
\]

where \( Y_k = X_k \otimes I_{d_t} \) and

\[
y_k = \begin{bmatrix} y_{1k} \\ y_{2k} \\ \vdots \\ y_{dk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_{1} \\ \beta_{2} \\ \vdots \\ \beta_{d} \end{bmatrix}, \quad \beta_i = \begin{bmatrix} A_{i1} \\ A_{i2} \\ \vdots \\ A_{id} \end{bmatrix}, \quad \epsilon_k = \begin{bmatrix} \epsilon_{1k} \\ \epsilon_{2k} \\ \vdots \\ \epsilon_{dk} \end{bmatrix}, \quad \epsilon_{ik} = \begin{bmatrix} E_{2i1} \\ E_{2i2} \\ \vdots \\ E_{Ti1} \end{bmatrix},
\]

the \( \otimes \) is a Kronecker product operator.

Since the matrix \( Y_k \) is a block diagonal matrix, the estimation problem can be decomposed into \( d \) independent sub-problems. Let us consider the \( i \)-th equation of the system.

\[
\begin{bmatrix} X_{1i1} \\ X_{1i2} \\ \vdots \\ X_{1id} \\ X_{2i1} \\ X_{2i2} \\ \vdots \\ X_{2id} \\ \vdots \\ X_{(i-1)i} \end{bmatrix} = \begin{bmatrix} A_{i1} \\ A_{i2} \\ \vdots \\ A_{id} \end{bmatrix}, \quad \epsilon_{i1} = \begin{bmatrix} E_{2i1} \\ E_{2i2} \\ \vdots \\ E_{Ti1} \end{bmatrix},
\]

the above equation system can be considered as a multiple regression

\[
X_{ti} = X_{(t-1)i} A_{i1} + X_{(t-2)i} A_{i2} + \cdots + X_{(i-1)i} A_{id} + \epsilon_{ti}, \quad t \in \{2, 3, \cdots, T\},
\]

this equation system can be abbreviated as

\[
z_{ik} = x_{ik} A_{i1} + x_{2k} A_{i2} + \cdots + x_{dk} A_{id} + \tilde{\epsilon}_{ik},
\]

(5)
with $T - 1$ samples to estimate the $i$-th line of transition matrix $A$.

Let $F(x_{ik})$ be the distribution of $x_{ik}$, we assume the independence between $\xi_{ik}$ and $F(x_{ik})$, for $i = 1, 2, ..., d$. Then we get the Gini covariance matrix equation issue from Equation (5),

$$
\begin{bmatrix}
\text{cov}(z_{ik}, F(x_{ik})) \\
\text{cov}(z_{ik}, F(x_{2k})) \\
\vdots \\
\text{cov}(z_{ik}, F(x_{dk}))
\end{bmatrix} = \begin{bmatrix}
\text{cov}(x_{1k}, F(x_{1k})) & \text{cov}(x_{2k}, F(x_{1k})) & \cdots & \text{cov}(x_{dk}, F(x_{1k})) \\
\text{cov}(x_{1k}, F(x_{2k})) & \text{cov}(x_{2k}, F(x_{2k})) & \cdots & \text{cov}(x_{dk}, F(x_{2k})) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov}(x_{1k}, F(x_{dk})) & \text{cov}(x_{2k}, F(x_{kd})) & \cdots & \text{cov}(x_{dk}, F(x_{dk}))
\end{bmatrix} \begin{bmatrix}
A_{i1} \\
A_{i2} \\
\vdots \\
A_{id}
\end{bmatrix}.
$$

From the above equation, we obtain the so called Gini–Yule–Walker Equation

$$
G = GA^\top,
$$

where $G = \frac{1}{n} \sum_{k=1}^n G_k$, $G = \frac{1}{n} \sum_{k=1}^n \tilde{G}_k$. The entries of $G_k$ are given by $G_{ij,k} = \text{Gcov}(z_{ik}, x_{ik})$, and the entries of $\tilde{G}_k$ are given by $\tilde{G}_{ij,k} = \text{Gcov}(x_{jk}, x_{ik})$.

2.3. Sample Gini Covariance Matrix

We use a $U$-statistic method to estimate the Gini covariance matrix $G_k$ and $\tilde{G}_k$. From Equation (6), the elements of the covariance matrix $G_k$ and $\tilde{G}_k$ can be divided into two categories: Gini covariance $\text{Gcov}(X_k, \tilde{Y}_k)$ and Gini mean difference $\text{Gcov}(Y_k, \tilde{Y}_k)$. For the Gini covariance $\text{Gcov}(\tilde{X}_k, \tilde{Y}_k)$, we have sample space

$$\{(\tilde{X}_{1k}, \tilde{Y}_{1k}), (\tilde{X}_{2k}, \tilde{Y}_{2k}), \ldots, (\tilde{X}_{(T-1)k}, \tilde{Y}_{(T-1)k})\}.$$
For the Gini mean difference $G_{cov}(\tilde{Y}_k, \tilde{Y}_k)$, we have sample space \{\tilde{Y}_{1k}, \tilde{Y}_{2k}, ..., \tilde{Y}_{(T-1)k}\}. Let $\tilde{Y}_{1k}$ and $\tilde{Y}_{2k}$ be two independent random variables with distribution function $F_{\tilde{Y}_k}$, the Gini mean difference $G_{cov}(\tilde{Y}_k, \tilde{Y}_k)$ can be expressed as

$$G_{cov}(\tilde{Y}_k, \tilde{Y}_k) = \frac{1}{2} \left( \int_0^{\infty} 2gF_{\tilde{Y}_k}(g)dF_{\tilde{Y}_k}(g) - \int_0^{\infty} g dF_{\tilde{Y}_k}(g) \right)$$

$$= \frac{1}{2} E(\max(\tilde{Y}_{1k}, \tilde{Y}_{2k}) - \tilde{Y}_{1k}).$$

The estimator of $G_{cov}(\tilde{Y}_k, \tilde{Y}_k)$ based on $U$-statistics is given by

$$U(\tilde{Y}_k, \tilde{Y}_k) = \frac{1}{\left(\frac{1}{2}\right)^T} \sum_{i=1}^{T-1} \sum_{j<i}^{T-1} k(\tilde{Y}_{ik}, \tilde{Y}_{jk}),$$

where $k(\tilde{Y}_{1k}, \tilde{Y}_{2k}) = \frac{1}{4}(2 \max(\tilde{Y}_{1k}, \tilde{Y}_{2k}) - \tilde{Y}_{1k} - \tilde{Y}_{2k})$. After some simplification, we obtain

$$U(\tilde{Y}_k, \tilde{Y}_k) = \frac{1}{2(T-1)(T-2)} \sum_{i=1}^{T-1} (2i - T - 2)\tilde{Y}_{i,(T-1)k},$$

where $\tilde{Y}_{i,(T-1)k}$ is the $i$-th order statistic based on the sample space \{\tilde{Y}_{1k}, ..., \tilde{Y}_{(T-1)k}\}.

2.4. Convergence Rates of the Estimator $U(\tilde{X}_k, \tilde{Y}_k)$ and $U(\tilde{Y}_k, \tilde{Y}_k)$

In this subsection, with the truncation method, we use the Bernstein’s inequality to investigate the convergence rates of the estimator $U(\tilde{X}_k, \tilde{Y}_k)$ and $U(\tilde{Y}_k, \tilde{Y}_k)$. From Equations (8) and (9), we define $\gamma_k = U(\tilde{X}_k, \tilde{Y}_k)$, $\eta_k = U(\tilde{Y}_k, \tilde{Y}_k)$, $1 \leq k \leq n$, and $\Delta = \max\{D(\gamma_k), D(\eta_k)\}$, where $D(\gamma_k)$ is the variance of the variable $\gamma_k$.

For analysis, we require the following three assumptions on the time series and the size of variables $(d, T, n)$:

**Assumption 1.** From Equation (2), suppose that the entries of the heavy-tailed innovation $E_{kT+v} = (E_{kT+v, 1}, \ldots, E_{kT+v, d})^T$ obey a power-law distribution with $P(|E_{kT+v, i}| > x) \sim Cx^{-\alpha_i}$ and $\alpha_i \leq 2$, then the second moment is finite and the third moment is infinite.

**Assumption 2.** Suppose that $\frac{\log n}{n} \to 0$, $\frac{dT}{\sqrt{n}} \leq c$, $c$ is a finite constant, for $T \to \infty$, $n \to \infty$.

**Assumption 3.** Suppose $\frac{d}{n} \to 0$, for $n \to \infty$.

**Lemma 1.** Let \{X_k\}_{k \in \mathbb{Z}} be a stationary PVAR process from Equation (2), and $X_1, \ldots, X_n$ be a sequence of observations from \{X_k\}_{k \in \mathbb{Z}}. Suppose that Assumptions (A1)–(A2) are satisfied. Then, for $T$ and $n$ large enough, with probability no smaller than $1 - \left(\frac{8\Delta \log n}{dT} + \frac{2}{T^2}\right)$, we have

$$\left| \frac{1}{n} \sum_{k=1}^{n} (\gamma_k - E\gamma_k) \right| \leq d \sqrt{\frac{8T \log n}{n}},$$

$$\left| \frac{1}{n} \sum_{k=1}^{n} (\eta_k - E\eta_k) \right| \leq d \sqrt{\frac{8T \log n}{n}}.$$  

**Proof.** Assuming that $m$ and $\lambda$ are constants greater than 0, and

$$\hat{\gamma}_k = (\gamma_k - E\gamma_k)1_{|\gamma_k - E\gamma_k| \leq m},$$
Then, \( \tilde{\gamma}_k \) is a bounded random variable and satisfies the property of independent identical distribution, it follows from the Bernstein’s inequality that

\[
\mathbb{P}(\left| \sum_{k=1}^{n} \tilde{\gamma}_k \right| \geq \lambda) \leq 2 \exp \left( -\frac{\lambda^2}{2\sigma^2 + \frac{2}{3}m\lambda} \right),
\]

(12)

where \( \sigma^2 = nD(\tilde{\gamma}_k) \leq n\Delta \).

Let \( m = d \sqrt{\frac{T_n}{8\log n}} \), \( \delta = d \sqrt{\frac{8T \log n}{n}} \) and \( \lambda = n\delta \), we have

\[
\mathbb{P}(\left| \sum_{k=1}^{n} (\tilde{\gamma}_k - E\tilde{\gamma}_k) \right| \geq \delta) \leq n\mathbb{P}(\left| \gamma_k - E\gamma_k \right| > m) + \mathbb{P}(\left| \sum_{k=1}^{n} \tilde{\gamma}_k \right| \geq n\delta)
\]

\[
\leq \frac{nD(\gamma_k)}{m^2} + 2 \exp \left( -\frac{8d^2T \log n}{2D(\gamma_k) + \frac{2}{3}d^2T} \right)
\]

\[
\leq \frac{8\Delta \log n}{d^2T} + \frac{2}{n^2}.
\]

As \( T \to \infty \) and \( n \to \infty \), assuming \( \frac{\log n}{T} \to 0 \) and \( \frac{dT}{\sqrt{n}} \leq c \), then \( \frac{8\Delta \log n}{d^2T} + \frac{2}{n^2} \to 0 \).

With similar proof methods, we obtain that

\[
\mathbb{P}(\left| \frac{1}{n} \sum_{k=1}^{n} (\eta_k - E\eta_k) \right| \geq \delta) \leq \frac{8\Delta \log n}{d^2T} + \frac{2}{n^2}.
\]

This completes the proof. \( \square \)

From Equations (5) and (6), we define the sample estimation of Gini covariance matrix \( G \) as \( \hat{G} = (\hat{G}_{ij}) \), \( \hat{G}_{ij} = \frac{1}{n} \sum_{k=1}^{n} U(z_{jk}, x_{ik}) \) and the sample estimation of Gini covariance matrix \( \tilde{G} \) as \( \hat{\tilde{G}} = (\hat{\tilde{G}}_{ij}) \), \( \hat{\tilde{G}}_{ij} = \frac{1}{n} \sum_{k=1}^{n} U(x_{jk}, x_{ik}) \), \( 1 \leq i, j \leq d \).

Next, we investigate the convergence rates of the estimator \( \hat{G} \) and \( \hat{\tilde{G}} \) under the \( l_\infty \) norm.

**Lemma 2.** Let \( \{X_k\}_{k \in Z} \) be a stationary PVAR process from Equation (2), and \( X_1, \cdots, X_n \) be a sequence of observations from \( \{X_k\}_{k \in Z} \). Suppose that Assumptions (A1)–(A3) are satisfied. Then, for \( T \) and \( n \) large enough, with probability no smaller than \( 1 - (\frac{8\Delta \log n}{T} + \frac{2d^2}{n^2}) \), we have

\[
\| \hat{G} - G \|_\infty \leq d \sqrt{\frac{8T \log n}{n}},
\]

(13)

\[
\| \hat{\tilde{G}} - \tilde{G} \|_\infty \leq d \sqrt{\frac{8T \log n}{n}}.
\]

(14)
Proof. Let $\delta = d \sqrt{\frac{8T \log n}{n}}$, based on the Lemma 1, we have

$$P\{\|\tilde{G} - G\|_\infty \geq \delta\} \leq \sum_{i,j=1}^{d} P(|\hat{G}_{ij} - G_{ij}| \geq \delta)$$

$$\leq d^2 P(|\hat{G}_{ij} - G_{ij}| \geq \delta)$$

$$= d^2 P\left(\left|\frac{1}{n} \sum_{k=1}^{n} \left(U(z_{jk}, x_{ik}) - G_{i} \text{cov}(z_{jk}, x_{ik})\right)\right| \geq \delta\right)$$

$$\leq d^2 P\left(\left|\frac{1}{n} \sum_{k=1}^{n} (\gamma_k - E \gamma_k)\right| \geq \delta\right)$$

$$\leq \frac{8\Delta \log n}{d^2 T} + \frac{2d^2}{n^2}.$$  \hspace{1cm} (15)

As $T \to \infty$ and $n \to \infty$, assuming $\frac{\log n}{T} \to c$ and $\frac{d}{n} \to 0$, then $\frac{8\Delta \log n}{d^2 T} + \frac{2d^2}{n^2} \to 0$.

Next, we study the convergence of sample Gini covariance matrix $\hat{G}$.

Let $\delta = d \sqrt{\frac{8T \log n}{n}}$, based on the Lemma 1,

$$P\left(\left|\hat{G}_{jj} - G_{jj}\right| \geq \delta\right) = P\left(\left|\frac{1}{n} \sum_{k=1}^{n} \left(U(x_{jk}, x_{jk}) - G_{i} \text{cov}(x_{jk}, x_{ik})\right)\right| \geq \delta\right)$$

$$= P\left(\left|\frac{1}{n} \sum_{k=1}^{n} (\eta_k - E \eta_k)\right| \geq \delta\right)$$

$$\leq \frac{8\Delta \log n}{d^2 T} + \frac{2d^2}{n^2}.$$  \hspace{1cm} (16)

Now, for the off-diagonal entries, we have

$$P\left(\left|\hat{G}_{ij} - G_{ij}\right| \geq \delta\right) = P\left(\left|\frac{1}{n} \sum_{k=1}^{n} \left(U(x_{jk}, x_{ik}) - G_{i} \text{cov}(x_{jk}, x_{ik})\right)\right| \geq \delta\right)$$

$$= P\left(\left|\frac{1}{n} \sum_{k=1}^{n} (\gamma_k - E \gamma_k)\right| \geq \delta\right)$$

$$\leq \frac{8\Delta \log n}{d^2 T} + \frac{2d^2}{n^2}.$$  \hspace{1cm} (16)

Combining Equations (15) and (16), we obtain

$$P\{\|\hat{G} - G\|_\infty \geq \delta\} \leq \sum_{i,j=1}^{d} P(|\hat{G}_{ij} - G_{ij}| \geq \delta)$$

$$\leq d^2 \left(\frac{8\Delta \log n}{d^2 T} + \frac{2d^2}{n^2}\right)$$

$$= \frac{8\Delta \log n}{T} + \frac{2d^2}{n^2}.$$  \hspace{1cm} (17)

This completes the proof. \hspace{1cm} $\Box$

3. Theoretical Properties

Using the Gini covariance matrix Equation (6), we propose to estimate $A$ by

$$\hat{A} = \arg\min_{M \in \mathbb{R}^{d \times d}} \|M_{i,j}\| \text{ s.t. } \|\hat{G} - M^T\|_\infty \leq \lambda.$$  \hspace{1cm} (17)
The above optimization problem can be further decomposed into $d$ subproblems. The $i$-th row of transition matrix $\hat{A}$ by

$$A_{i*} = \arg \min_{v \in \mathbb{R}^{d \times 1}} \|v\|_1 \quad s.t. \quad \|\hat{G}_{i*} - \hat{G}v\|_\infty \leq \lambda. \tag{18}$$

Compared to the lasso-type procedures, the proposed method can be solved in parallel and Equation (17) can be solved efficiently using the parametric simplex method for sparse learning in [21].

Based on Lemma 2, we can further deliver the rates of convergence of $\hat{A}$ under the matrix $l_\infty$ norm and $l_1$ norm. We start with some additional notation. For $\alpha \in [0, 1)$, $\eta > 0$, and $M_d > 0$ that may scale with $d$, we define the matrix class

$$\mathcal{M}(\alpha, \eta, M_d) = \{M \in \mathbb{R}^{d \times d} : \max_{1 \leq j \leq d} \sum_{k=1}^{d} |M_{jk}|^\alpha \leq \eta, \|M\|_1 \leq M_d\}. \tag{19}$$

$\mathcal{M}(\alpha, \eta, M_d)$ requires the transition matrices are sparse in rows. If $\alpha = 0$, then the maximum number of nonzeros in rows of transition matrix is at most $\eta$. $\mathcal{M}(\alpha, \eta, M_d)$ is also investigated in [12].

**Theorem 1.** Let $\{X_k\}_{k \in \mathbb{Z}}$ be a stationary PVAR process from Equation (2). Suppose that Assumptions (A1)–(A3) are satisfied. The transition matrix $A^\top \in \mathcal{M}(\alpha, \eta, M_d)$, if we choose the tuning parameter $\bar{\lambda} = (1 + M_d) d \sqrt{\frac{8 T \log n}{n}}$. Then, for $T$ and $n$ large enough, with probability no small than $1 - (\frac{8 \Delta \log n}{T} + \frac{2d^2}{n^2})$, we have

$$\|A^\top - A^\top\|_\infty \leq 2\bar{\lambda} \|\hat{G}^{-1}\|_1, \tag{20}$$

$$\|A^\top - A^\top\|_1 \leq 4\eta(2\bar{\lambda} \|\hat{G}^{-1}\|_1)^{1-\alpha}. \tag{21}$$

**Proof.** We first show that with large probability, $A$ is feasible to the optimization problem. By the Gini–Yule–Walker equation, we have

$$\|\hat{G} A^\top - \hat{G}\|_\infty = \|\hat{G} G^{-1} G - \hat{G}\|_\infty = \|\hat{G} G^{-1} G - G + G - \hat{G}\|_\infty \leq \|\hat{G} G^{-1} - I\|_\infty + \|G - \hat{G}\|_\infty \leq \|\hat{G} - \hat{G}\|_\infty \|A^\top\|_1 + \|G - \hat{G}\|_\infty \leq \|\hat{G} - \hat{G}\|_\infty M_d + \|G - \hat{G}\|_\infty.$$

The last inequality is due to $A^\top \in \mathcal{M}(\alpha, \eta, M_d)$, by Lemma 2, with probability no small than $1 - (\frac{8 \Delta \log n}{T} + \frac{2d^2}{n^2})$ that,

$$\|\hat{G} A^\top - \hat{G}\|_\infty \leq (1 + M_d) d \sqrt{\frac{8 T \log n}{n}} = \bar{\lambda}.$$

$A$ is feasible in the optimization equation, by checking the Equation (17), we have $\|\hat{A}\| \leq \|A\|$ with probability no smaller than $1 - (\frac{8 \Delta \log n}{T} + \frac{2d^2}{n^2})$. 

Next, we prove Equation (20). Let $A^T = \hat{G}^{-1}G$, we have
\[
\|\hat{A}^T - A^T\|_\infty = \|\hat{A}^T - \hat{G}^{-1}G\|_\infty \\
= \|\hat{G}^{-1}(\hat{G}A^T - G)\|_\infty \\
= \|\hat{G}^{-1}(\hat{G}A^T - \hat{G}A^T + \hat{G} - \hat{G}A^T)\|_\infty \\
\leq \|\hat{G}^{-1}\|_1(\|\hat{G} - \hat{G}\|_\infty \|\hat{A}^T\|_1 + \|\hat{G} - G\|_\infty + \|\hat{G}A^T - \hat{G}1\|_\infty) \\
\leq 2\lambda\|\hat{G}^{-1}\|_1
\]
with probability no smaller than $1 - (\frac{8\lambda \log n}{T} + 2\alpha^2 + \frac{2\alpha^2}{n})$.

Let $\lambda_1 \geq 0$ and $S_j = \{m : |A_{jm}| \geq \lambda_1\}, j = 1, \cdots, d$. Define
\[
\hat{D}(\lambda_1) = \max_{j \leq d} \sum_{i=1}^d \min\{|A_{jm}| \wedge \lambda_1, 1\}.
\]

Then, we have
\[
\|(\hat{A}^T - A^T)_{i\cdot}\|_1 \leq \|\hat{A}_{i,S_j}^T\|_1 + \|A_{j,S_j}^T\|_1 + \|(\hat{A}^T - A^T)_{j,S_j}\|_1 \\
= \|A_{i,S_j}^T\|_1 - \|A_{j,S_j}^T\|_1 + \|A_{j,S_j}^T\|_1 + \|(\hat{A}^T - A^T)_{j,S_j}\|_1 \\
\leq 2\|A_{j,S_j}^T\|_1 + 2\|(\hat{A}^T - A^T)_{j,S_j}\|_1 \\
\leq 2\|A_{j,S_j}^T\|_1 + 2\|\hat{A}^T - A^T\|_\infty \cdot |S_j| \\
\leq 2\hat{D}(\lambda_1)(\lambda_1 + 2\lambda\|\hat{G}^{-1}\|_1).
\]

Suppose $\max_j \sum_{i=1}^d |(A^T)_{ij}|^a \leq \eta$ and setting $\lambda_1 = 2\lambda\|\hat{G}^{-1}\|_1$, we have
\[
\lambda_1S_j = \max_{1 \leq j \leq d} \sum_{i=1}^d \min\{|(A)_{ij}| \wedge \lambda_1\} \leq \lambda_1 \max_{1 \leq j \leq d} \sum_{i=1}^d \min\{|(A)_{ij}|^a / \lambda_1^a, 1\} \leq \lambda_1^{1-a} \eta.
\]

Therefore, we have
\[
\|(\hat{A}^T - A^T)_{i\cdot}\|_1 \leq 4\lambda_1\hat{D}(\lambda_1) \leq 4\eta \lambda_1^{1-a} = 4\eta(2\lambda\|\hat{G}^{-1}\|_1)^{1-a}.
\]

Since the above equation holds for any $j = 1, \cdots, d$, we complete the proof. $\square$

4. Granger Causality

In this section, the practical example is conducted to verify the effectiveness of the proposed methods, moreover, the characterization and estimation of Granger causality under the heavy-tailed PVAR model are discussed. Firstly, we give the definition of Granger causality.

**Definition 1.** (Granger [22]) Let $\{X_t\} \in \mathbb{Z}$ be a stationary process, where $X_t = (X_{t1}, \ldots, X_{td})^T$. For $j \neq k \in \{1, \ldots, d\}$, $\{X_{tk}\}_{t \in \mathbb{Z}}$ Granger causes $\{X_{tj}\}_{t \in \mathbb{Z}}$ if and only if there exists a measurable set $A$ such that
\[
\mathbb{P}\left(X_{(t+1)j} \in A \mid \{X_s\}_{s \leq t}\right) \neq \mathbb{P}\left(X_{(t+1)j} \in A \mid \{X_{s \backslash k}\}_{s \leq t}\right)
\]
for all $t \in \mathbb{Z}$, where $X_{s \backslash k}$ is the subvector obtained by removing $X_{sk}$ from $X_s$.
For a Gaussian VAR process \( \{X_t\}_{t \in \mathbb{Z}} \), we have that \( \{X_t\}_{t \in \mathbb{Z}} \) Granger causes \( \{X_t\}_{t \in \mathbb{Z}} \) if and only if the \((j, k)\) entry of the transition matrix is non-zero [4]. For the heavy-tailed PVAR process, let \( \{X_k\}_{k \in \mathbb{Z}} \) be a stationary PVAR process from Equation (2), we define

\[
\dot{X}_t = \frac{1}{n} \sum_{k=1}^{n} X_{ik}, \quad \dot{E}_t = \frac{1}{n} \sum_{k=1}^{n} E_{ik}, \quad 1 \leq t \leq T.
\]

In the next theorem, we show that a similar property holds for the heavy-tailed PVAR process.

**Theorem 2.** Let \( \{X_k\}_{k \in \mathbb{Z}} \) be a stationary PVAR process from Equation (2). Suppose that Assumptions (A1)–(A3) are satisfied, and \( \text{Gcov} \left( \dot{X}_{ti}, \dot{X}_{sl} \mid \dot{X}_{s \setminus l}, t \leqslant 1 \right) \neq 0 \) for any \( i \in \{1, \ldots, d\} \). Then, for \( j \neq i \in \{1, \ldots, d\} \), we have

1. If \( A_{ji} \neq 0 \), then \( \{\dot{X}_{ti}\}_{t \in \mathbb{Z}} \) Granger causes \( \{\dot{X}_{lj}\}_{l \in \mathbb{Z}} \).
2. If we further assume that \( \dot{E}_{t+1} \) is independent of \( \{\dot{X}_{sl}\}_{s \leq t} \) for any \( t \in \mathbb{Z} \), we have that \( \{\dot{X}_{ti}\}_{t \in \mathbb{Z}} \) Granger causes \( \{\dot{X}_{lj}\}_{l \in \mathbb{Z}} \) if and only if \( A_{ji} \neq 0 \).

**Proof.** In order to prove Issue 1, we only need to prove that \( \{\dot{X}_{ti}\}_{t \in \mathbb{Z}} \) doesn’t Granger cause \( \{\dot{X}_{lj}\}_{l \in \mathbb{Z}} \) implies \( A_{ji} = 0 \). Suppose for some \( t \in \mathbb{Z} \), we have

\[
\mathbb{P}\left( \dot{X}_{(t+1)j} \in A \mid \{\dot{X}_{sl}\}_{s \leq t} \right) = \mathbb{P}\left( \dot{X}_{(t+1)j} \in A \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right)
\]

for any measurable set \( A \). The above equation implies that conditioning on \( \{\dot{X}_{sl}\}_{s \leq t} \), \( \dot{X}_{(t+1)j} \) is independent of \( \{\dot{X}_{s \setminus l}\}_{s \leq t} \). Hence, we have

\[
\text{Gcov}\left( \dot{X}_{(t+1)j}, \dot{X}_{lj} \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right) = 0.
\]

Plugging \( \dot{X}_{(t+1)j} = \sum_{i=1}^{d} A_{ji} \dot{X}_{ti} + \dot{E}_{(t+1)j} \) into the above equation, we have

\[
0 = \text{Gcov}\left( A_{ji} \dot{X}_{ti}, \dot{X}_{lj} \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right) + \text{Gcov}\left( \sum_{i \neq j} A_{ji} \dot{X}_{ti}, \dot{X}_{lj} \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right) + \text{Gcov}\left( \dot{E}_{(t+1)j}, \dot{X}_{lj} \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right).
\]

The second term on the right hand side is 0, since given \( \{\dot{X}_{sl}\}_{s \leq t} \), \( \sum_{i \neq j} A_{ji} \dot{X}_{ti} \) is constant. Since \( \dot{E}_{t+1} \) and \( \{\dot{X}_{sl}\}_{s \leq t} \) are independent for any \( t \in \mathbb{Z} \), we have \( \text{Gcov}\left( \dot{E}_{(t+1)j}, \dot{X}_{sl} \right) = 0 \) for any \( s \leq t \). Using Theorem 2.18 in Fang et al. [23], we have the third term is also 0. Thus, we have \( \text{Gcov}\left( \dot{X}_{ti}, \dot{X}_{lj} \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right) = 0 \) and hence \( A_{ji} = 0 \). This proves Issue 1.

Given Issue 1, to prove Issue 2, it remains to prove that \( A_{ji} = 0 \) implies that \( \{\dot{X}_{ti}\}_{t \in \mathbb{Z}} \) doesn’t Granger cause \( \{\dot{X}_{lj}\}_{l \in \mathbb{Z}} \). Since \( A_{ji} = 0 \), we have

\[
p\left( \dot{X}_{(t+1)j} \mid \{\dot{X}_{si}\}_{s \leq t} \right) = p\left( \dot{X}_{(t+1)j} \mid \{\dot{X}_{sl}\}_{s \leq t} \right) p\left( \{\dot{X}_{si}\}_{s \leq t} \mid \{\dot{X}_{s \setminus l}\}_{s \leq t} \right). \tag{22}
\]
Here $p$ is the conditional probability density function. The last equation is because $E_{t+1} k$ is independent of $\{X_k\}_{s \leq t}$, and the fact that $\sum_{i \neq k} A_{ji} X_{ji}$ is constant given $\{X_s\}_{s \leq t}$. Hence, we have

$$p \left( \tilde{X}_{(t+1)j} \mid \{ \tilde{X}_{s,i} \}_{s \leq t} \right) = p \left( \tilde{X}_{(t+1)j} \mid \{ \tilde{X}_{s,i} \}_{s \leq t} \right) \tag{23}$$

and thus

$$p \left( \tilde{X}_{(t+1)j} \mid \{ \tilde{X}_s \}_{s \leq t} \right) = p \left( \tilde{X}_{(t+1)j} \mid \{ \tilde{X}_{s,i} \}_{s \leq t} \right).$$

\[ \square \]

**Remark 1.** The assumption that $\text{Geo}(\tilde{X}_{ji}, X_{ji} \mid X_{ji} \neq 1) \neq 0$ requires that $\tilde{X}_{ji}$ cannot be perfectly predictable from the past or from the other observed random variables at time $t$. Otherwise, we can simply remove $\{\tilde{X}_{ji}\}_{i \in Z}$ from the process $\{\tilde{X}_i\}_{i \in Z}$ since predicting $\{\tilde{X}_{ji}\}_{i \in Z}$ is trivial. Assuming that $E_{t+1}$ is independent of $\{X_s\}_{s \leq t}$ for any $i \in Z$, the Granger causality relations among the processes $\{\{\tilde{X}_{ji}\}_{i \in Z} : j = 1, \ldots, d\}$ is characterized by the non-zero entries of $A$. To estimate the Granger causality relations, we define $\hat{A} = [\hat{A}_{ji}]$, where

$$\hat{A}_{ji} := \hat{A}_{ji} I (|\hat{A}_{ji}| \geq \gamma)$$

for some threshold parameter $\gamma$. To evaluate the consistency between $\hat{A}$ and $A$ regarding sparsity pattern, we define function $\text{sign}(x) := 1 (x > 0) - 1 (x < 0)$. For a matrix $M$, define $\text{sign}(M) := [\text{sign}(M_{ji})]$. The next theorem gives the rate of $\gamma$ such that $\hat{A}$ recovers the sparsity pattern of $A$ with high probability.

**Theorem 3.** Let $\{X_k\}_{k \in Z}$ be a stationary PVAR process from Equation (2). Suppose that Assumptions 1–3 are satisfied. The transition matrix $A^\top \in M(\alpha, \eta, M_d)$, if we set

$$\gamma = 2\lambda \|(\tilde{C})^{-1}\|_1, \tag{24}$$

then, with probability no smaller than $1 - \left( \frac{8A\log n}{T} + \frac{2\delta^2}{n^2} \right)$, we have $\text{sign}(\hat{A}) = \text{sign}(A)$, provided that

$$\min_{\{(j,i) : A_{ji} > 0\}} |A_{ji}| \geq 2\gamma. \tag{25}$$

**Proof.** The proof is a consequence of Theorem 1. In detail, if $A_{ji} > 0$, by Equation (25), we have $\hat{A}_{ji} \geq 2\gamma$. By Theorem 1, with probability no smaller than $1 - \left( \frac{8A\log n}{T} + \frac{2\delta^2}{n^2} \right)$, we have $|\hat{A}_{ji} - A_{ji}| \leq \gamma$. Thus, we have $\hat{A}_{ji} \geq \gamma$ with probability no smaller than $1 - \left( \frac{8A\log n}{T} + \frac{2\delta^2}{n^2} \right)$.

By the definition of $A$, we have $\hat{A}_{ji} = \hat{A}_{ji} \geq \gamma > 0$.

If $A_{ji} < 0$, by Equation (25), we have $A_{ji} \leq -2\gamma$. Using Theorem 1, we have $\hat{A}_{ji} \leq -\gamma$ with probability no smaller than $1 - \left( \frac{8A\log n}{T} + \frac{2\delta^2}{n^2} \right)$. By the definition of $\tilde{A}$, we have $\hat{A}_{ji} = \hat{A}_{ji} \leq -\gamma < 0$.

If $A_{ji} = 0$, using Theorem 1, we have $\hat{A}_{ji} < \gamma$ with probability no smaller than $1 - \left( \frac{8A\log n}{T} + \frac{2\delta^2}{n^2} \right)$, since $\mathbb{P} (\hat{A}_{ji} = \gamma) = 0$. By the definition of $\hat{A}$, we have $\hat{A}_{ji} = 0$. \[ \square \]

5. Experiments

This section provides some numerical results on synthetic and real data. We consider Lasso and Dantzig selector for comparison.
• **Lasso**: an $l_1$ regularized estimator defined as

$$
\hat{A} = \arg \min_{M \in \mathbb{R}^{d \times d}} \frac{1}{n} \sum_{k=1}^{n} \| Y_k - MX_k \|_F^2 + \rho \sum_{ij} M_{ij},
$$

where $Y_k := (X_{1,k}, \ldots, X_{(T-1),k}) \in \mathbb{R}^{d \times (T-1)}$, $X_k := (X_2,k, \ldots, X_T,k) \in \mathbb{R}^{d \times (T-1)}$.

• **Dantzig selector**: the estimator proposed in [12],

$$
\hat{A} \in \arg \min_{M \in \mathbb{R}^{d \times d}} \sum_{jk} |M_{jk}| \text{ s.t. } \frac{1}{n} \sum_{k=1}^{n} \| S_k M - S_{1k}^T \|_{\text{max}} \leq \lambda,
$$

where $S_k$ and $S_{1k}$ be the marginal and lag one sample covariance matrices of Equation (2).

• **G-Dantzig selector**: the estimator described in Equation (17).

### 5.1. Synthetic Data

In this subsection, we compare the performance of our method with the lasso and Dantzig under synthetic data. The “flare” package in R is used to create the transition matrix according to three patterns: cluster, hub, random. We rescale $A$ such that $\| A \|_2 = 0.8$, generate $\Sigma$ such that $\| \Sigma \|_2 = 2 \| A \|_2$, then calculate the covariance matrix $\Psi$ of the noise vector $E_t$ as $\Psi = \Sigma - A^T \Sigma A$. Let the periodical time series length $n = 2000$, the period length $T = 100$ and the dimension $d = 150$, with $A, \Sigma$ and $\Psi$, we simulate periodical time series according to the model described in Equation (2). Specifically, we consider three models:

• **Model 1**: Data generated from Equation (2), where the errors $E_t \sim N(0, \Psi)$. There is no outliers under Model 1.

• **Model 2**: Data generated from Equation (2), where the errors $E_t \sim t(0, \Psi, 2)$. No outliers appear and the tail of the error distribution is heavier than that of the normal distribution.

• **Model 3**: Data generated from Equation (2) with the same error distribution as Model 1. Then replace $X_i$ by $X_i + (-1)^{U_2(i \leq 1/2)} (20 + 10U_2) I$ for $i = 1, \ldots, T/5$, where $U_1$ and $U_2$ are independently generated from $U[0,1]$ distribution, there are $T/5$ outliers that deviate from the majority of the observations.

The tuning parameter $\lambda$ is chosen by cross validation. We construct 20,000 replicates and compare the three methods described above. Table 1 presents averaged estimation errors under three matrix norms. From this table, we have the following findings: Under the Gaussian model (Model 1), G-Dantzig selector has comparable performance as Dantzig selector and out performs Lasso. Under Model 2 and 3, our methods are more stable than Lasso and Dantzig selector. Thus, we conclude that the G-Dantzig selector is robust to the heavy-tailedness of data and the possible presence of outliers.

Figure 1 plots the prediction errors against sparsity for the three transition matrix estimators. We observe that G-Dantzig selector achieves smaller prediction errors compared to Lasso and Dantzig selector.
Figure 1. Prediction errors in stock prices plotted against the sparsity of the estimated transition matrix.

Table 1. Comparison of estimation errors of three methods under different setups. The standard deviations are shown in parentheses. Here $L_1$, $L_\infty$, $L_F$ are the $l_1$, $l_\infty$ and Frobenius matrix norms respectively.

|               | Lasso        | Dantzig     | Our Method |
|---------------|--------------|-------------|------------|
|               | $L_1$        | $L_\infty$  | $L_F$      |
| Model 1       |              |             |            |
| cluster       | 1.13         | 1.62        | 1.67       |
|               | (0.17)       | (0.16)      | (0.55)     |
| hub           | 1.37         | 2.16        | 2.39       |
|               | (0.34)       | (0.45)      | (0.22)     |
| random        | 1.41         | 2.58        | 2.18       |
|               | (0.43)       | (0.67)      | (0.17)     |
| Model 2       |              |             |            |
| cluster       | 1.25         | 1.76        | 1.79       |
|               | (0.64)       | (0.45)      | (0.37)     |
| hub           | 1.51         | 2.76        | 1.58       |
|               | (0.41)       | (0.56)      | (0.14)     |
| random        | 1.36         | 2.24        | 2.17       |
|               | (0.32)       | (0.17)      | (0.17)     |
| Model 3       |              |             |            |
| cluster       | 1.54         | 2.13        | 1.96       |
|               | (0.29)       | (0.13)      | (0.18)     |
| hub           | 1.52         | 2.67        | 1.41       |
|               | (0.26)       | (0.18)      | (0.07)     |
| random        | 1.78         | 2.56        | 2.21       |
|               | (0.22)       | (0.26)      | (0.54)     |

5.2. Real Data

We further compared the three methods on a real equity data from Yahoo Finance. We collected the 10-minutes time intervals price in Monday of each week for 50 stocks with the highest volatility that were consistently in the S&P 500 index from 1 January 2008 to 31 December 2020. Then, the periodical time series length was $n = 672$, the period length was $T = 36$ and the dimension was $d = 50$. Since we chose the data points on the Monday of each week, the data points can be seen as independent of each other. Estimations of the transition matrix $\hat{A}_s$ were obtained by the Lasso, Dantzig selector and G-Dantzig selector, where $s \in [0, 1]$ is the fraction of non-zero entries of $\hat{A}_s$ and can be controled by the tuning parameters $\lambda$ and $\rho$. We define the prediction error associated with $\hat{A}_s$ to be

$$
\epsilon_s := \frac{1}{n} \sum_{k=1}^{n} \left( \frac{1}{T-1} \sum_{t=2}^{T} \| X_{tk} - \hat{A}_s X_{(t-1)k} \|_2 \right).
$$
6. Conclusions

In this paper, we developed a Gini–Yule–Walker equation for modeling and estimating the heavy-tailedness data and the possible presence of outliers in high dimensions. Our contributions are three-fold. (i) At the model level, we generalized the Gaussian process to time series with the existence of merely second order moments. (ii) Methodologically, we proposed a Gini–Yule–Walker-based estimator of the transition matrix. Experimental results demonstrate that the proposed estimator is robust to heavy-tailedness of data and the possible presence of outliers. (iii) Theoretically, we proved that the adopted method yields a parametric convergence rate in the matrix $l_\infty$ and $l_1$ norm. In this manuscript, we focused on the stationary vector autoregressive model and our method is designed for such stationary process. The stationary requirement is a common assumption in analysis and is adopted by most recent works, for example, see and [14,24]. We notice that there are works in handling time-varying PVAR models, checking for example in [25]. We would like to explore this problem in the future.

Author Contributions: Conceptualization, J.Z. and D.H.; methodology, D.H. and J.Z.; software, J.Z.; validation, J.Z.; formal analysis, J.Z.; investigation, J.Z.; resources, J.Z.; data curation, J.Z.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z.; visualization, J.Z.; supervision, J.Z. and D.H.; project administration, J.Z.; funding acquisition, D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Natural Science Foundations of China [grant number 11531001].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Franses, P.H.; Paap, R. Periodic Time Series Models; Oxford U Press: Oxford, UK, 2004.
2. Bell, W.R.; Holan, S.H.; McElroy, T.S. Economic Time Series: Modeling and Seasonality; Chapman and Hall/CRC Press: Boca Raton, FL, USA, 2012.
3. Aiat, B.; Hamdi, F. On Markov-switching periodic ARMA models. Commun. Stat. Theory Methods. 2018, 47, 344–364. [CrossRef]
4. Lütkepohl, H. New Introduction to Multiple Time Series Analysis; Springer Science & Business Media: Berlin, Germany, 2005.
5. Baek, C.; Davis, R.A.; Pipiras, V. Sparse seasonal and periodic vector autoregressive modeling. Comput. Stat. Data Anal. 2017, 106, 103–126. [CrossRef]
6. Gao, W.; Yang, H.; Yang, L. Change points detection and parameter estimation for multivariate time series. Soft Comput. 2020, 24, 6395–6407. [CrossRef]
7. Basu, S.; Michailidis, G. Regularized estimation in sparse high-dimensional time series models. Ann. Stat. 2015, 43, 1535–1567. [CrossRef]
8. Bai, P.; Safikhani, A.; Michailidis, G. Multiple Change Points Detection in Low Rank and Sparse High Dimensional Vector Autoregressive Models. IEEE Trans. Signal Process. 2020, 68, 3074–3089. [CrossRef]
9. Han, F.; Liu, H. Transition Matrix Estimation in High Dimensional Vector Autoregressive Models. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 172–180.
10. Hong, D.; Gu, Q.; Whitehouse, K. High-dimensional time series clustering via cross-predictability. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 642–651.
11. Chen, X.; Xu, M.; Wu, W.B. Regularized estimation of linear functionals of precision matrices for high-dimensional time series. IEEE Trans. Signal Process. 2016, 64, 6439–6470. [CrossRef]
12. Han, F.; Lu, H.; Liu, H. A Direct Estimation of High Dimensional Stationary Vector Autoregressions. J. Mach. Learn. Res. 2015, 16, 3115–3150.
13. Qiu, H.; Xu, S.; Han, F.; Liu, H.; Caffo, B. Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. In Proceedings of the International Conference on Machine Learning, International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 1843–1851.
14. Wong, K.C.; Li, Z.; Tewari, A. Lasso guarantees for mixing heavy-tailed time series. Ann. Stat. 2020, 48, 1124–1142. [CrossRef]
15. Maleki, M.; Wraith, D.; Mahmoudi, M.R.; Contreras-Reyes, J.E. Asymmetric heavy-tailed vector auto-regressive processes with application to financial data. J. Stat. Comput. Simul. 2020, 90, 324–340. [CrossRef]
16. Schechtman, E.; Yitzhaki, S. A measure of association based on gin’s mean difference. *Commun. Stat. Theory Methods* **1987**, *16*, 207–231. [CrossRef]

17. Schechtman, E.; Yitzhaki, S. On the proper bounds of the Gini correlation. *Econ. Lett.* **1999**, *63*, 133–138. [CrossRef]

18. Schechtman, E.; Yitzhaki, S. A family of correlation coefficients based on the extended Gini index. *J. Econ. Inequal.* **2003**, *1*, 129–146. [CrossRef]

19. Yitzhaki, S.; Schechtman, E. *The Gini Methodology: A Primer on a Statistical Methodology*; Springer Science & Business Media: Berlin, Germany, 2012.

20. Yang, S.S. Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function. *J. Am. Stat. Assoc.* **1981**, *76*, 658–662. [CrossRef]

21. Pang, H.; Liu, H.; Vanderbei, R. The fastclime package for linear programming and large-scale precision matrix estimation in R. *J. Mach. Learn. Res. JMLR* **2014**, *15*, 89–493.

22. Granger, C.W. Testing for causality: A personal viewpoint. *J. Econ. Dyn. Control.* **1980**, *2*, 329–352. [CrossRef]

23. Fang, K.W. *Symmetric Multivariate and Related Distributions*; CRC Press: Boca Raton, FL, USA, 2018.

24. Song, S.; Bickel, P.J. Large vector auto regressions. *arXiv* **2011**, arXiv:1106.3915.

25. Haslbeck, J.M.; Bringmann, L.F.; Waldorp, L.J. A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivar. Behav. Res.* **2020**, *4*, 1–30. [CrossRef] [PubMed]